

Measurement properties of the SF-MPQ-2 Neuropathic Qualities subscale in persons with CRPS: validity, responsiveness and Rasch analysis

Tara L. Packham, PhD^{1,2} Debbie Bean, PhD^{3,4} Malcolm H. Johnson, PhD^{3,4} Joy C. MacDermid, PhD² Sharon Grieve, MSc⁵ Candida S McCabe, PhD⁵ R. Norman Harden, MD⁶

¹ Michael G. DeGroot Institute for Pain Research and Care, Hamilton ON, Canada

² School of Rehabilitation Sciences, McMaster University, Hamilton ON

³ Department of Psychological Medicine, University of Auckland, New Zealand

⁴ The Auckland Regional Pain Service (TARPS), Auckland District Health Board, Auckland, NZ

⁵ Royal United Hospitals NHS Foundation Trust, Bath & University of the West of England, Bristol, UK

⁶ Department of Physical Medicine and Rehabilitation, Northwestern University, Chicago Illinois, USA

Corresponding Author:

Tara L. Packham, packhamt@mcmaster.ca. c/o School of Rehabilitation Sciences, Institute of Applied Health Sciences Rm 403, McMaster University. 1400 Main Street W, Hamilton, ON, Canada. L8S 1C7. 1-905-525-9140; fax 1-905-524-0069.

Acknowledgements: T. Packham is supported by a postdoctoral fellowship from the Michael G. DeGroot Institute for Pain Research and Care. J. MacDermid is supported by the Roth|MacFarlane Chair in Knowledge Translation. Data was collected with support of a grant from the Oakley Mental Health Research Foundation.

Running title: Validity of SF-MPQ2 Neuropathic Subscale

ABSTRACT

Objectives: The purpose of this study was to conduct classical psychometric evaluation and Rasch analysis on the Neuropathic Qualities subscale from the Short Form McGill Pain Questionnaire-2 utilizing scores from persons with complex regional pain syndrome to consider reliability and person separation, validity (including unidimensionality), and responsiveness in this population. **Methods:** Secondary analysis of longitudinal data from persons with acute

complex regional pain syndrome was utilized for analysis of the psychometric properties and fit to the Rasch model of the Neuropathic Qualities subscale. We followed an iterative process of Rasch analysis to evaluate and address data fitting challenges. **Results:** Repeated measures from $n=59$ persons meeting the Budapest criteria were used for analysis. Both item-total correlations and unidimensionality analyses supported theoretical construct validity; all convergent construct validity hypotheses were also supported. Responsiveness was demonstrated comparing baseline and one-year data at $d=0.92$, with a standardized response mean of 0.97. Data were able to fit the Rasch model but all Neuropathic Qualities items had disordered thresholds requiring rescoring. Additionally, local dependency and differential item function were addressed by ‘bundling’, suggesting no further item reduction would be possible. **Conclusions:** This study provided preliminary support for the validity and responsiveness of the Neuropathic Qualities subscale in persons with complex regional pain syndrome. Rasch analysis further endorses use of the Neuropathic Qualities subscale as a ‘stand-alone’ measure for neuropathic features, but with substantial background data transformations. Replication with larger samples is recommended to increase confidence in these findings.

Keywords: complex regional pain syndrome, neuropathic pain qualities, Rasch analysis, outcome measurement

INTRODUCTION

Complex regional pain syndrome (CRPS) is a unique pain presentation that may develop after trauma or surgery, or spontaneously.(1) While some persons may see resolution of their symptoms in the first year,(2) others experience persistent pain and disability.(3,4) The Special Interest Group for CRPS of the International Association for the Study of Pain convened a working group to develop recommendations for outcome measurement for clinical studies in CRPS (COMPACT: Core Outcome Measurement set for complex regional Pain syndrome Clinical studies). The mandate of this group was to select a recommended core set of outcome measures to use in all clinical studies involving CRPS patients with the goal of supporting international collaborations and future meta-analyses to advance the field of CRPS research.(5) One of the consensus recommendations from COMPACT was to prefer the use of assessments that had been tested using item-response theory,(5) which considers the measurement properties invariant across populations, thus facilitating comparisons(6) of pain and disability between persons with CRPS and other diagnostic groups (for example, painful diabetic neuropathy). After consensus was reached, which drew substantially from the PROMIS measures(7,8) to address many of the key constructs, the COMPACT team deliberated on what assessments of pain qualities were concordant with this theoretical lens, or had existing data available that would permit timely Rasch analysis to consider interval-level scoring properties.

Rasch analysis

Rasch analysis is an item-response framework based on probabilistic modelling: simply stated, it considers the odds that any two persons with similar amounts of the trait of interest (for example, neuropathic pain symptoms) would score the same on any given scale item. Another conceptualization is Rasch modelling calculates the observed vs predicted values for each **item** on a scale across all persons in the sample, and the observed vs predicted values for each **person** across items. These predicted values are generated from other scores or known characteristics, also called person factors, and from the item scales themselves. Thus items and person are ‘fit’ to the Rasch model, using **person ability** and **item difficulty**. Accordingly, if someone reports they are able to walk a mile without pain, it would also be predicted they would positively answer an item about if they can easily walk a block without pain. The item about walking a mile is more ‘**difficult**’ than the item about walking a block, and the person who is able to walk a mile has more **ability** than someone who is only able to walk a block. Conversely, if someone reports severe pain, we would also predict they would not positively answer an item about walking a mile without pain. These properties or ‘locations’ of item difficulty and person ability/disability can be visually mapped to illustrate their relationship for the particular assessment being examined.

Developed by Danish mathematician George Rasch,(6,9) the Rasch model: 1) uses a statistical strategy to convert the ordinal scale measurements of individual test items into interval level scaling,(9) 2) identifies systematic bias (termed differential item functioning) across different person-level characteristics such as gender,(10) 3) relies on large sample sizes to validate the item characteristics (11), and 4) assumes the item measurement characteristics are not dependent on the study population (person characteristics) if the model can be applied, and is adequately powered.(12,13) A key theoretical element is the score of any item moves predictably from less

difficult to more difficult (for example from 0-10, where 0 means no pain, and 10 means worst possible pain). The point between any two scores (like 0,1 or 1,2) where the probability is 50/50 a person with a set amount ability would choose 0 vs 1 is called a threshold.(9) The orderly or predictable progression of thresholds is a prerequisite for achieving model fit.

Rasch modelling also generates estimates of item reliability, where the focus is on how well items are differentiated on the variable of interest (analogous to Cronbach's alpha);(9,14) person separation (a similar concept to reliability, where the standard error is used to describe the spread or separation of person abilities);(9) and unidimensionality (which could be considered analogous to content validation, or factorial validity).(15) A scale that fits the Rasch model is assumed to reliably allow meaningful comparisons between individuals with different abilities or attributes (like education, gender, or even diagnosis) but the same amount of whatever trait the scale is intended to measure (for example, pain interference with walking from our previous example).

The Short Form of the McGill Pain Questionnaire-2

One tool considered by COMPACT, for inclusion in the core assessment set for the assessment of pain qualities, was the Neuropathic Qualities subscale (NeQ) of the second version of the short form McGill Pain Questionnaire (SF-MPQ2).(16) The SF-MPQ2 subscale for neuropathic qualities was considered appropriate to address the important domain of pain qualities.(5) While many other measures of neuropathic pain are available, our concerns were 1) many were developed with the intent to discriminate between nociceptive and neuropathic pain features, rather than to measure the severity or impact of neuropathic pain,(17–20) or 2) lacked requisite evidence for test-retest reliability and responsiveness.(21)(22) Further, the brevity of including a brief scale or subscale was desirable, to minimize participant burden related to a core set of measures. To date, psychometric studies of the measurement properties of the total SF-MPQ2 have not included substantial proportions of participants with CRPS: estimates for reliability, validity and responsiveness arise from populations with painful diabetic neuropathy(16), cancer pain,(23) acute low back pain,(24) or a mixture of pain diagnoses.(16,25) While the scaling structure of the SF-MPQ2 has been validated,(16) supporting the legitimacy of each subscale as a unique construct, we were unaware of any published studies looking at the NeQ subscale (see Table 1 for an overview of the SF-MPQ2 items and subscales) as a distinct entity for outcome measurement. If data generated from the NeQ subscale of the SF-MPQ2 demonstrated fit to the Rasch model, this would support its validity as a stand-alone outcome measure for neuropathic pain. When we set out to examine the SF-MPQ2, no Rasch analyses of the measurement characteristics had been published. However, a recent publication reported Rasch analysis of the SF-MPQ2 in participants with knee osteoarthritis(26) and suggested the NeQ subscale could also fit the model. Therefore, an opportunity exists to consider the measurement properties of the NeQ subscale of the SF-MPQ2 using Rasch modeling in a different population, and to compare these findings to estimates of measurement properties using traditional psychometric methods from classical test theory (CTT). This would ultimately assist pain clinicians in the selection of evaluation tools to inform treatment planning for persons with CRPS. The purpose of this study therefore, was to conduct both classical psychometric, and Rasch analysis on the NeQ subscale from the SF-MPQ2 utilizing scores from participants with CRPS to consider its reliability and

person separation, and estimate validity (including unidimensionality), and responsiveness in this population.

Table 1. Overview of the items of the Short Form McGill Pain Questionnaire (version 2).

Continuous pain	Intermittent pain	Neuropathic pain	Affective descriptors
Throbbing pain	Shooting pain	Hot-burning pain	Tiring-exhausting pain
Cramping pain	Stabbing pain	Cold-freezing pain	
Gnawing pain	Sharp pain	Pain caused by light touch	Sickening
Aching pain	Splitting pain	Itching	Fearful
Heavy pain	Electric shock pain	Tingling or 'pins and needles'	Punishing-cruel
Tender	Piercing pain	Numbness	

Items are scored 0-10, with 0= no pain to 10=worst possible pain.

METHODS

Sample and Data Collection

This study represents a secondary analysis of data collected from participants with acute CRPS (n=59) referred by community and hospital-based clinics in Auckland, New Zealand between February 2012 and March 2014.(2,27) The aims of the study were 1) to track symptoms and signs of in persons with CRPS in the first year of the diagnosis (2), 2) to identify predictors of recovery(27), and 3) evaluate predictors of disability and work status in early CRPS.(28) Partner sites were asked to identify eligible subjects as they presented for clinic visits. Adult subjects (≥ 18 years) diagnosed using the IASP 1994 criteria(29,30) with acute (<12 weeks) CRPS type 1 of any limb, with no previous history of CRPS were eligible for the study. However only those who met the IASP Budapest research criteria for CRPS were included in the current analysis. Additional inclusion criteria were English literacy, and ability to provide informed consent. After providing informed consent, participants were assessed on three occasions: 1) at referral (4-116 days after symptom onset), 2) at 6 months after symptom onset, and 3) at 12 months after symptom onset. See Figure 1 for a flow diagram of study recruitment. The original study was approved by the New Zealand Ministry of Health Northern Y ethics committee. Data available for analysis consisted of participant demographics, clinical characteristics, the CRPS Severity Score (CSS),(31) and self-reported questionnaires including the SF-MPQ2,(16) and the Pain Disability Index (PDI).(32,33) The source study also collected data using the short Tampa scale for Kinesiophobia (34), Pain Catastrophizing Scale (35) Depression Anxiety Stress Scale-21 (36), the Bath Body Perception Disturbance Scale for CRPS (37), and work status. See Figure 2 for a complete listing of the variables utilized for our analysis.

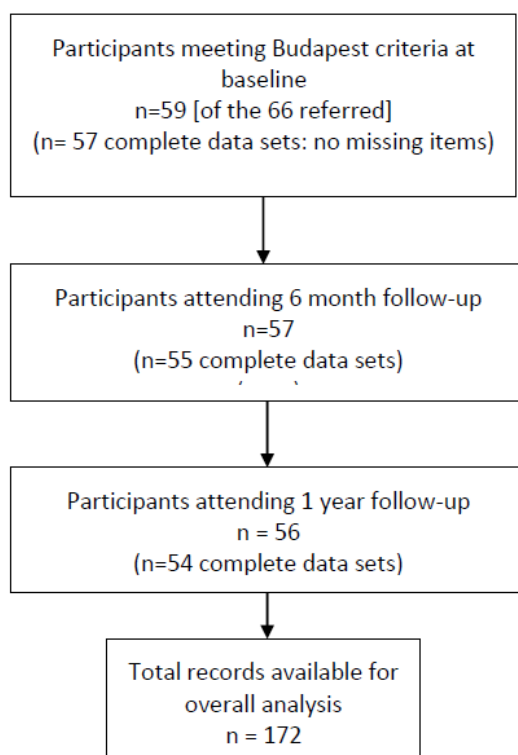


Figure 1. Flow diagram of overall study participation

Description of outcome measures

Short-Form McGill Pain Questionnaire, Version 2 The SF-MPQ2 contains 22 pain descriptors (scored 0 – 10 using the anchors ‘none’ to ‘worst possible’) on 4 subscales representing continuous, intermittent, neuropathic or affective features.(24) The developers reported good internal consistency for each subscale (ranging from 0.73 – 0.87 across several investigations in large samples), and discriminant validity was supported by significant differences in change scores across a clinical trial in those who considered themselves improved compared to those who did not ($p < 0.002$ for all scales).(24)

CRPS Severity Score The CSS is a 17 item condition-specific clinician rating scale for CRPS signs which closely aligns with the Budapest diagnostic criteria(31,38). It addresses sensory, autonomic, vasomotor, motor and trophic features, (rated 1/0 as present/absent), and has demonstrated discriminative validity to distinguish between CRPS and other forms of chronic pain (30). More recently, the scale was modified to 16 items to give equal weighting to signs and symptoms: a multi-site prospective study reported acceptable stability ($ICC = 0.67$) over a 3 month period in persons with chronic (stable) CRPS, and change in CSS scores in persons with acute CRPS were significantly correlated with changes in pain intensity, and Rand-36 pain, fatigue, well-being, social functioning, and physical roles subscale scores.(39)

Pain Disability Index The PDI is a brief 7-item self-reported evaluation of the impact of pain on activities and daily functioning, using 0-5 ratings for each item.(32) It demonstrated good internal consistency ($\alpha = 0.86$) in a large cross-sectional sample, but lower test-retest reliability over a 2 month period ($r = 0.44$, $p < 0.001$).(40) Several studies have supported discriminant validity for distinguishing between persons with high or low levels of disability (32,33,40)

- SF-MPQ2 item and summary scores
- Pain Disability Index total scores
- CRPS Severity Score (total score)
- Pain ratings (NRS)
- Age
- Sex
- Affected limb
- Duration of CRPS symptoms

Figure 2. Variables available for analysis

Data analysis

All data sets from each time point were included in the overall analysis. For analyses requiring only cross-sectional data, the longitudinal data was transformed (or ‘stacked’) by consideration of repeated measures from individual participants as unique data points based on duration of CRPS symptoms. Cases with missing values for any items were excluded from the analysis of Cronbach’s alpha (reliability); however, the number of cases included in each analysis is clearly specified in the reporting. For Rasch modelling, only integers are permitted as data for scaling; therefore a small number of responses indicating a half value were rounded up (i.e. a response of 7.5 would be transformed to 8; this accounted for less than 2% of the study values).

Rasch analysis was conducted using RUMM2030 software (Rumm Laboratory Pty Ltd, Perth, Australia) and following Tennent et al.’s iterative steps for Rasch modelling:(14,41) see Table 2 for a brief overview of this process. For a detailed description of a worked Rasch analysis intended for a clinical audience, the reader is referred to Packham & MacDermid.(41) Using RUMM2030 software constrains the analysis to 9 categorical person-level factors to be used to describe the sample or test population relative to the construct of interest: refer to Figure 2 for the variables used here. These person-level factors are also considered to develop the ‘location’, a Rasch rating of the item severity, which can be understood as how much of the construct of interest is required for a respondent to endorse any level of an individual item. For analysis of reliability, validity and responsiveness, data were imported into STATA13 (StataCorp, College Station, TX) for statistical calculations, including:

- Cronbach’s alpha for the total scale and for the NeQ subscale
- Validity exploration testing the hypotheses that a) self-reported SF-MPQ2 total and SF-MPQ2 NeQ scores will be substantially correlated to pain scores measured by a numeric rating scale and to PDI scores, and b) SF-MPQ2 NeQ scores will be moderately correlated to clinician-generated CSS scores. Strength of the correlations were interpreted employing Landis & Koch’s recommendations where $r=0-.20$ is considered slight, $r=.21-.40$ is fair, $r=.41-.60$ is moderate, $r=.61-.80$ is substantial, and $r>.80$ is considered excellent.(42)
- Responsiveness of the SF-MPQ2 was estimated using the effect size and standardized response means, and compared to the effect sizes and standardized response means of the pain Numeric Rating Scale (NRS), PDI and CSS over the same interval.

Table 2. Overview of the statistical analyses included in Rasch modelling (adapted from Packham & MacDermid) (34)

RASCH ANALYSIS COMPONENT	METHOD OF ANALYSIS	HYPOTHESIS/ EVALUATIVE CRITERIA
Partial credit vs. rating scale parameters	Log-likelihood ratio	If rating scale parameters are confirmed by a non-significant ratio test, then items do not vary from the core rating scale structure, and a rating scale format may be utilized. If the test is significant, a partial credit model is used
Response distribution	Frequency plot of actual responses across all possible items and response options	Each possible category of item/response should be used (this visually maps floor or ceiling effects); ideal to have at least 5 cases for each rating level of every item to ensure adequate analysis power
Thresholds	Graphic representation and plots of probability	Ordered progression of thresholds (the point where there is a 50/50 probability of choosing either response option) from less to more of the latent trait of interest
Item fit	Chi-square, fit residual transformed to a standardized (Z) score	Mean 0, SD up to 1 if data fit the Rasch model; reflects the divergence between expected and actual values looking across all items scored by an individual person
Person fit	Chi-square, fit residual transformed to a Z-score	Mean, SD up to 1 if data fit the Rasch model; reflects the divergence between expected and actual values looking across all persons for a given item
Item trait interaction	Chi-square probability	$p > 0.05$ tests whether the items are working as expected (fitting the model) across the class intervals for the trait
Individual person fit	Chi-square analysis	Values should fall within ± 2.5 if the data fit the Rasch model
Person separation index	Cronbach's alpha	Values over 0.70 allow for group comparison, over 0.85 for individual comparisons of summed scores
Local dependency	Correlation analysis of item residuals	Correlations of less than 0.3 suggest no firm evidence of local dependency
Differential item functioning	Item characteristic curves & ANOVA of item residuals	IC curves plotted by person characteristics (such as gender or age) and $p > 0.05$ for between groups ANOVA reveal no evidence of DIF
Unidimensionality	Principal component factor analysis is used to identify positive and negatively loading items; these repeatedly t-tested against each other to ensure they are not significantly different; the <i>proportion</i> of significant t-tests is reported	The proportion of significant t-tests will be less than 0.05 (or less than 5%) and/or if the 95% confidence interval includes 0.05, unidimensionality is supported

Key: SD = standard deviation; ANOVA = analysis of variance; IC = item characteristic; DIF =

differential item functioning

RESULTS

Demographics and person characteristics

Fifty-nine participants in the Auckland study met the IASP Budapest research criteria for CRPS(30) and were included in this analysis. Table 3 describes the demographics and clinical characteristics of this sample who were initially evaluated an average of nine weeks after the development of CRPS symptoms.(27)

Table 3. Demographics for sample N=59

Characteristic	Mean	SD	Range
Age	48.2 years	13.3	18 - 72
Duration of symptoms at baseline evaluation	67.8 days	24.3	4 - 116
Pain at baseline (NRS/10)	5.5	2.0	1 - 10
CSS at baseline (/17)	12.6	2.0	8 - 17
Total SF-MPQ2 (/10) at baseline	4.1	2.1	0.27 - 10
NeQ subscale score at baseline (n=59)	3.9	2.0	0.3 - 10
NeQ subscale score at 6 months (n=57)	2.3	2.2	0 - 9.2
NeQ subscale score at 1 year (n=56)	1.9	2.2	0 - 9.7
Total PDI at baseline	38.1	14.2	14 - 67
	Frequency	Percent	
Gender	F=43 M=16	72.9 % 27.1 %	
Affected limb	R U/E= 33 L U/E=21 R L/E=3 L L/E=1 Multiple= 1	55.9 % 35.6 % 5.1 % 1.7 % 1.7 %	
Etiology	None/minor=5 Soft tissue=2 Minor surgery=21 Fracture=20 Fracture+surgery=11	8.5 % 3.4 % 35.6 % 33.9 % 18.6 %	

Key: CSS= CRPS Severity Score, NeQ= Neuropathic Qualities, NRS= Numeric Rating Scale, PDI=Pain Disability Index, SD=standard deviation, SF-MPQ2= Short-Form of the McGill Pain Questionnaire (version 2), F=female, M=male, R=right L=left, U/E= upper extremity, L/E=lower extremity

Rasch analysis of the NeQ scale of SF-MPQ2

The total of complete data sets available for this analysis after considering repeated measures data as cross-sectional were $n=156$. The log likelihood ratio was calculated at $p>0.05$, therefore the partial credit model was utilized for analysis(14): this verifies there is not a common scale structure across all items, because the intervals between any two points on the scale may vary both within and across items.(43) The distribution of responses failed to meet the suggested criteria of at least 5 responses per scoring level per item(14), as participants tended to use lower scores, reflecting their recovery over the 3 measurement periods (see Table 4). More than half ($n=34$) failed to meet Linacre's conservative suggestion of at least 10 endorsements per scoring level.(44) Thresholds for all 6 items on the NeQ scale were 'disordered' i.e. did not follow the expected progression for severity of neuropathic pain, and required rescoring by collapsing of scoring categories. The final scales ranged from 4-6 response options (instead of original 11 response options [0-10]). The solutions for combining categories were also not uniform, therefore this would be very difficult to operationalize in practice.

Table 4. Distribution of scoring and endorsements (n=157)

Item/Score	0	1	2	3	4	5	6	7	8	9	10
Burning	<i>72</i>	<i>11</i>	<i>9</i>	<i>10</i>	<i>6</i>	<i>11</i>	<i>8</i>	<i>6</i>	<i>13</i>	<i>7</i>	<i>4</i>
Freezing	<i>95</i>	<i>7</i>	<i>9</i>	<i>12</i>	<i>9</i>	<i>4</i>	<i>8</i>	<i>3</i>	<i>6</i>	<i>2</i>	<i>2</i>
Pain with touch	<i>52</i>	<i>20</i>	<i>12</i>	<i>9</i>	<i>12</i>	<i>6</i>	<i>13</i>	<i>6</i>	<i>19</i>	<i>4</i>	<i>4</i>
Itching	<i>88</i>	<i>9</i>	<i>19</i>	<i>11</i>	<i>6</i>	<i>2</i>	<i>7</i>	<i>5</i>	<i>4</i>	<i>2</i>	<i>4</i>
Tingling	<i>31</i>	<i>18</i>	<i>13</i>	<i>11</i>	<i>13</i>	<i>17</i>	<i>11</i>	<i>9</i>	<i>16</i>	<i>9</i>	<i>9</i>
Numbness	<i>40</i>	<i>11</i>	<i>9</i>	<i>19</i>	<i>14</i>	<i>12</i>	<i>5</i>	<i>11</i>	<i>16</i>	<i>8</i>	<i>12</i>

Key: Italicized numbers represent the number of endorsements (or participant selection of) the different scores

Threshold analysis Figure 3a provides a visual illustration of the disordered thresholds for the 'Tingling' item. The lines represent the probability of an individual choosing any given scoring level (0-10) relative to their predicted level of the construct of interest (or person [dis]ability): neuropathic pain. The figure (Figure 3a) demonstrates a near complete overlap of the probability curves for a score of 0, 1, and 2. In other words, because these curves are located below the average item score which is centred at 0 logits on the x axis, the probability of participants with a low level of neuropathic pain scoring 0, 1, or 2 is unpredictable. Scores using this 0-10 structure on this item do not follow the anticipated pattern of increasing proportionally as participants have more neuropathic pain. However, after collapsing the scoring categories (in this case by combining the scores of 1&2, 3-5, 6&7, 8&9, and retaining a score of 0 or 10 in the original form, yielding a new six level (0-5) scoring structure, the probability curves demonstrate a distinct and orderly progression across the person location values (Figure 3b).

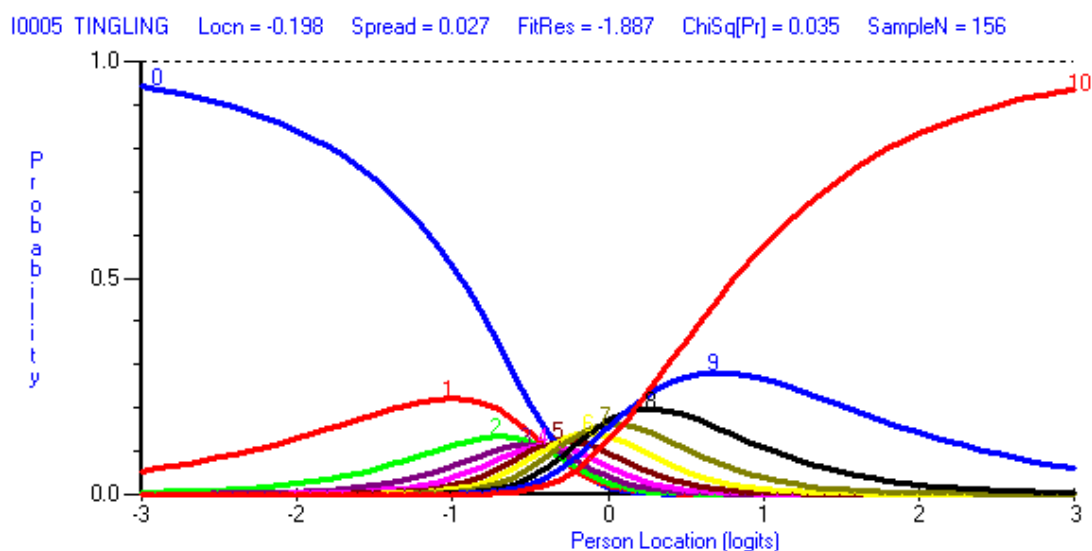


Figure 3A: Probability curve for 'Tingling' before correction of thresholds.

Key: The lines represent the probability of an individual choosing any given scoring level (0-10 in 1A; 0-5 in 1B) relative to their predicted level of Tingling.

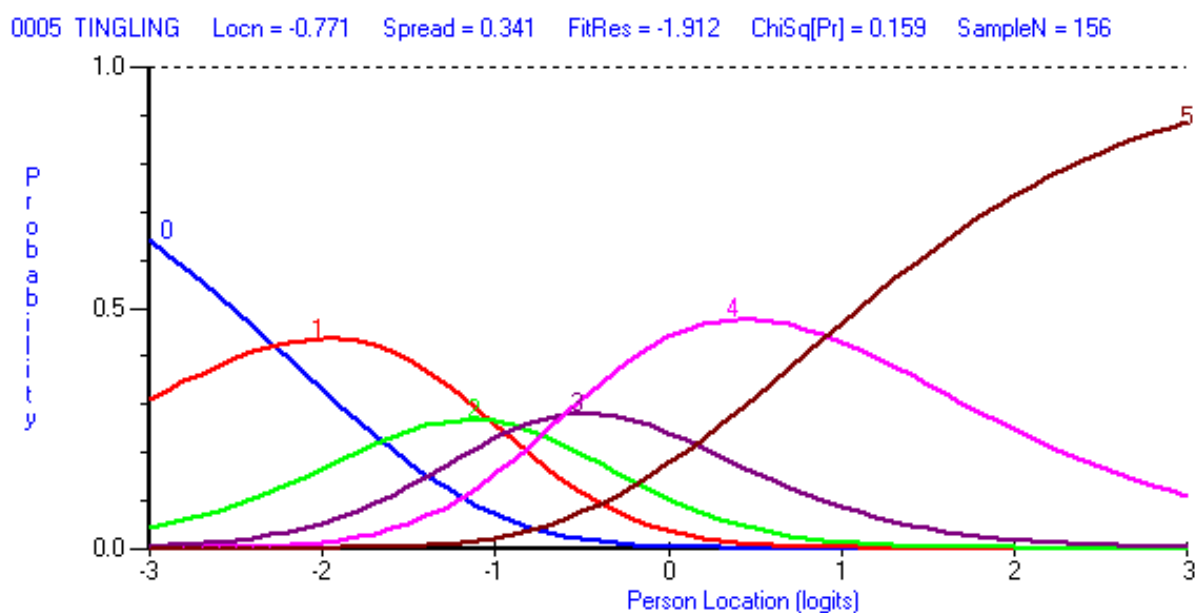


Figure 3B: Probability curves for 'Tingling' after collapsing of categories to correct thresholds

Item and person fit After threshold correction, item fit to the Rasch model was acceptable (mean set to 0 logits, SD=0.50, with fit residuals of $x = -0.28$ logits, SD 1.00); no misfitting items were identified and there was good distribution of items across the difficulty levels, supporting good targeting of the scale.(6,45) Person fit statistics were also acceptable, ($x = -1.17$ logits, SD 1.13), but the negative mean score relative to the average difficulty of the items reflects the generally low scores on the SF-MPQ2 NeQ subscale (starting at 4.1/10 at baseline and decreasing over time). This suggests the participants in the sample were better (i.e. had less neuropathic pain)

than the predicted scale average. Looking at individual person fit, 13 ‘extreme’ cases were identified: 12 had extremely low scores while one reported 10/10 pain on most items at final evaluation. A simple clinical interpretation is the predicted vs. observed scores on the NeQ were not what would be expected given the ratings or ‘person characteristics’ on the other measures: therefore the person’s reported pain scores did not fit with the rest of their clinical picture. Taken together, these statistics can be interpreted as reflecting the recovery of participants across the longitudinal data collection period: this is further supported by the mean scores of the SF-MPQ2 NeQ subscale at the 3 different measurement points recorded in Table 4.

Local dependency and differential item function

Local dependency was found between one item pairing, ‘*Burning*’ and ‘*Numbness*’. This occurs when there is a high correlation between two items representing distinct concepts, violating the assumption items on a scale are independent of each other.(9) Uniform differential item function was indicated between the ‘*Pain with touch*’ item and the CSS, suggesting participants scored the ‘*Pain with touch*’ item differently at different levels of CRPS severity [$F_{(156,5)}=5.07$, $p=0.0002$]. Table 5 shows the average score on the ‘*Pain with touch*’ across the CSS scores. To address both differential item functioning and local dependency, we employed a strategy common to Rasch analysis and bundled the items to create subtests(14) within the NeQ subscale: thereby creating Subtest A = *Burning*, *Pain with light touch* and *Numbness* [to address the local dependency detailed above], and retaining the remaining items *Cold/freezing*, *Tingling*, and *Itching* independently. This strategy resolved both the differential item function and local dependency, while maintaining fit to the model (see Table 6). It is important to note that no DIF was found for duration of CRPS symptoms, which suggests the rating scales operated similarly across time points. [Insert Tables 5 & 6 about here]

Person separation index and Cronbach’s alpha

After correction for the differential item functioning (DIF) and local dependency, the person separation index for the proposed model was 0.78, with the inclusion of all cases. This suggests the scale is able to discriminate between at least 2 groups, but falls below the suggested level of 0.85 for individual level precision.(14) After rescoring of thresholds, internal consistency (Cronbach’s alpha) based on the rescored items is 0.81 for the 6 items of the NeQ subscale of the SF-MPQ2, but drops to 0.70 when the scale is partitioned to address DIF and local dependency.

Unidimensionality

After principal components factor analysis to identify negatively and positively loading items, t-testing estimated these items sets as different in only a very small proportion of the comparisons $p= 0.032$ [95%CI 0.002 -0.066], which clearly meets the standard of $p<0.05$ to support unidimensionality.(46)

Psychometric estimates from classical test theory

Internal consistency Cronbach's alpha for the total SF-MPQ2 scale was estimated at $\alpha = 0.96$, and $\alpha = 0.83$ for the NeQ subscale. The item-total correlation estimate above 0.95 for the entire scale suggests redundancy exists for this population(47) while the estimate for the NeQ subscale would be considered good.(48)

Construct validity Regression analyses to explore *a priori* validity hypotheses confirmed all of the hypotheses in both strength and direction of the relationships, supporting convergent construct validity. The results of the regression analyses are summarized in Table 7.

Responsiveness The NeQ subscale of the SF-MPQ2 appeared to adequately assess change over the interval of baseline to one-year post development of symptoms. Effect sizes and standardized response means for the NeQ subscale, as well as the pain NRS, PDI and CSS as measured over the same interval, are presented in Table 8. [insert Table 8 about here]

DISCUSSION

The Rasch analysis presented here supports use of the NeQ scale for persons with CRPS, albeit with several limitations. This study describes the application of the Rasch model to the NeQ scale of the SF-MPQ2, from a dataset collected longitudinally on 59 participants across their first year of CRPS symptoms. The sample demographics were very characteristic of CRPS, with more women than men, more involvement of the upper extremity than lower, and fractures reported as the most common precipitating event.(49) The participants were each seen 3 times for measurement, and the data were transformed, allowing for the cross-sectional analysis of $n=157$ complete datasets. Rasch analysis relies on large samples to support the probabilistic modelling required.(9) Given that 11 of the 66 possible scoring categories were under-utilized in this sample, we must assume the entire analysis was underpowered because of an uneven distribution of scores. This variability in score distribution can in part be attributed to the recovery of the study participants across the study period. Our finding of disordered thresholds for all six items is also a likely reflection of an underpowered analysis.(44,50) Turner et al. (26) conducted a Rasch analysis of the SF-MP2 and NeQ scale ($n=240$ participants with knee OA completed these questionnaires): they were unable to fit the entire SF-MPQ2 to the Rasch model, but reported good fit of the NeQ subscale and generated a 'Rasch conversion' table for the total NeQ score.(26)

Because our study was a post-hoc analysis of existing data, the variables available to describe the person-level factors were based on constructs intended to inform other investigations.(27,28) Nonetheless, these constructs included a condition-specific measure of symptom severity (the CSS), a global rating of pain (NRS), and pain-related disability (PDI). Additional variables used to build the Rasch model were age, gender, and duration of symptoms. This generated a model of item difficulty that was well targeted across the spectrum of sample characteristics; however, the average level of neuropathic characteristics experienced by participants was lower than the average item difficulty on the NeQ subscale. This is concordant with the recovery expected across the three measurement points given the typical prognosis associated with early diagnosis of CRPS.(3,51,52)

One of the rationales for conduct of a Rasch analysis is the opportunity to consider the influence of systematic bias in how items are scored: in Rasch terms, differential item functioning (DIF). Our analysis identified the potential for DIF in the '*Pain with light touch*' item, but this influence

was mediated by simply considering the item alongside other items as a subtest. Allodynia, which includes pain with touch,(30) is a common feature in CRPS, and has been associated with poor prognosis in other investigations.(4) It is not discordant to see an almost exponential increase in the ‘*Pain with touch*’ score concomitant with an increasing CSS, particularly as allodynia is incorporated in the CSS as both a sign and symptom. However, it is also important to note the average score still remained relatively low (the highest mean score was still less than 3/10); again, this likely reflects the clinical recovery seen across the study interval.

Similarly, correlations (local dependency) were seen between the ‘*Burning*’ and ‘*Numbness*’ items: this was again addressed by combining the items in a subtest. In short, both the differential item function and local dependency suggest we could use the 6 Neuropathic Qualities items of the SF-MPQ2 as a unique scale for the construct of neuropathic pain, but should not use items individually, and may not be able to reduce the number of items on this subscale without alteration of the measurement properties of the remaining items (as well as the whole scale). This is because the estimates of the measurement properties for these individual items were influenced by their local dependency in this evaluation, and therefore should not be assumed to be invariant unless both items are used together.(9) This is in contrast to the previous Rasch analysis which identified a relationship between the ‘*Cold/Freezing*’ and ‘*Numbness*’ items.(26) Furthermore, the person separation index findings suggest a Rasch-corrected NeQ scale based on our study results may be useful for discriminating between 2 groups in a clinical trial, but may not be reliable enough to guide individual treatment decisions.(15) The interval level scores generated by Rasch transformation of the raw data (through the collapsing of scoring categories) also required re-calculation of all six items: while this is easily achieved by a computer or web-based application, it would demand considerable efforts from clinicians or researchers for data gathered using the traditional pen-and-paper methods.

From a classical test theory perspective, we were able to generate support for the internal consistency, construct validity and responsiveness of the NeQ subscale of the SF-MPQ2. The internal consistency findings were concordant with the Rasch item reliability ($\alpha = 0.83$ compared to $\alpha = 0.81$, respectively). While all of the self-reported measures included in this analysis demonstrated responsiveness to the substantial changes experienced by most people in this sample, it is interesting to note the largest effect size was seen for the CSS (see Table 8). This finding adds to the previously reported psychometric properties of the CSS,(31,39) positioning it as an outcome measure and not just a screening tool.

Limitations

The number of data sets included in this study ($n=156$) is above the recommended $n=150$ for item calibration(50) but falls below a ‘rule of thumb’ estimate of at least 250 data sets.(53) The lack of endorsements for all possible responses in our results (Table 4) would suggest underpowering.(14) There may also be bias introduced by our use of multiple data sets from the same participants (albeit from different time points). While there are advanced analysis strategies to check the influence of time-point dependencies in the data, we elected not to apply these to an underpowered analysis. However, the lack of DIF attributable to CRPS duration (our time-dependent person factor) suggested the potential time dependencies were not a strong influence on item measurement properties. Other limitation include that our data were gathered from a single health district in a single country, and the results may not be fully generalizable.

Using a heterogeneous sample of participants with various forms of neuropathic pain in future would also support examination of potential differential item functioning on the basis of diagnosis: something we were unable to accomplish in our sample where all participants included in the analysis met the IASP Budapest criteria for the diagnosis of CRPS. Finally, as this was a secondary analysis of existing data, we were unable to strategically select theoretically-informed person factors for defining item severity and functioning in this population; instead, we were limited to selecting from existing variables.

One final point of consideration is that while we examined construct validity of the NeQ subscale, we did not directly address content validity in this study. We did not directly ask the question if the NeQ accurately represents the pain qualities experienced by persons with CRPS. The substantial correlation ($r=0.63$) seen between NeQ and CSS scores could be considered preliminary evidence for this form of validity; however we would advocate for the need for continued investigations to build evidence for content validity, including face validity, relevance and comprehensiveness.⁽⁵⁴⁾ Further, our sample may not have been the ideal study population, as the improving scores and low score averages reflected a recovery trajectory that may not be seen in other settings.

In conclusion, on the basis of this Rasch analysis, pain clinicians can consider using the NeQ subscale of the SF-MPQ2 as a ‘stand-alone’ outcome measure for the neuropathic features of CRPS if substantial data transformations are made in the background. Replicating this analysis using a larger sample would increase the confidence in these results. From a classical test theory perspective, this examination generated preliminary support for the validity and responsiveness of the NeQ subscale in the CRPS population. However, test-retest reliability for this population group is still unknown: a key consideration given the variability associated with CRPS.⁽²²⁾

REFERENCES

1. Bruehl S. An update on the pathophysiology of complex regional pain syndrome. *Anesthesiol* 2010;113(3):713–25.
2. Bean DJ, Johnson MH, Heiss-Dunlop W, Kydd RR. Extent of recovery in the first 12 months of complex regional pain syndrome type-1: A prospective study. *Eur J Pain* 2016;20(6):884–94.
3. Bean DJ, Johnson MH, Kydd RR. The Outcome of Complex Regional Pain Syndrome Type 1: A Systematic Review. *J Pain* 2014;15(7):1–14.
4. Wertli M, Bachmann LM, Weiner SS, Brunner F. Prognostic factors in complex regional pain syndrome 1: a systematic review. *J Rehabil Med* 2013;45(3):225–31.
5. Grieve S, Perez RS, Birklein F, Brunner F, Bruehl S, Harden RN, Packham T, Gobeil F, Haigh R, Holly J, Terkelsen A, Davies L, Lewis J, Thomassen I, Connett R, Worth T, Vatine JJ, McCabe CS. Recommendations for a first Core Outcome Measurement set for complex regional PAin syndrome Clinical sTudies (COMPACT). *Pain* 2017; 158(6):1083–90.
6. Boone WJ, Staver JR, Yale MS. *Rasch Analysis in the Human Sciences*. Dordrecht, NL: Springer; 2014.
7. Junghaenel DU, Schneider S, Stone AA, Christodoulou C, Broderick JE. Ecological validity and clinical utility of Patient-Reported Outcomes Measurement Information

- System (PROMIS®) instruments for detecting premenstrual symptoms of depression, anger, and fatigue. *J Psychosom Res* 2014;76(4):300–6.
8. Amtmann D, Cook KF, Johnson KL, Cella D. The PROMIS Initiative: Examples of Applications in Rehabilitation. *Arch Phys Med Rehabil* 2011;92(10):s12–9.
 9. Bond T, Fox CM. *Applying the Rasch model: fundamental measurement in the human sciences*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2007.
 10. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: An introduction and illustrative example. *Int J Nurs Stud* 2009;46:380–93.
 11. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther* 2014;36(5):648–62.
 12. Andrich D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res* 2011;11(5):571–85.
 13. Streiner DL. Measure for measure: new developments in measurement and item response theory. *Can J Psychiatry* 2010;55(3):180–6.
 14. Tennant A, Horton M, Pallant JF. *Introductory Rasch Analysis: A workbook*. Leeds, UK: Department of Rehabilitation Medicine, University of Leeds; 2011.
 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*, Dec 2007; 57(8) :1358–1362.
 16. Dworkin RH, Turk DC, Revicki D, Harding G, Coyne KS, Peirce-Sandner S, et al. Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2). *Pain*. 2009;144(1–2):35–42.
 17. Freynhagen R, Baron R, Gockel U, Tölle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. *Curr Med Res Opin*. 2006;22(10):1911–20.
 18. Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, Cunin G, Fermanian J, Ginies P, Grun-Overdyking A, Jafari-Schluep H, Lantéri-Minet M, Laurent B, Mick G, Serrie A, Valade D, Vicaut E. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). *Pain* 2005;114:29–36.
 19. Bennett MI, Smith BH, Torrance N, Potter J. The S-LANSS score for identifying pain of predominantly neuropathic origin: Validation for use in clinical and postal research. *J Pain*. 2005;6(3):149–58.
 20. Packham TL, Cappelleri JC, Sadosky A, Macdermid JC, Brunner F. Measurement properties of painDETECT: Rasch analysis of responses from community-dwelling adult with neuropathic pain. *BMC Neurol*. 2017;17(48):1–9.
 21. Galer BS, Jensen MP. Development and preliminary validation of a pain measure specific to neuropathic pain: the Neuropathic Pain Scale. *Neurology*. 1997;48(2):332–8.
 22. Packham T, MacDermid JC, Henry J, Bain J. A systematic review of psychometric evaluations of outcome assessments for complex regional pain syndrome. *Disabil Rehabil*

- 2012 Jan ;34(13):1059–69.
23. Gauthier LR, Young A, Dworkin RH, Rodin G, Zimmermann C, Warr D, Librach SL, Moore M, Shepherd FA, Pillai Riddell R, Macpherson A, Melzack R, Gagliese L. Validation of the short-form McGill pain questionnaire-2 in younger and older people with cancer pain. *J Pain* 2014;15(7):756–70.
 24. Dworkin RH, Turk DC, Trudeau JJ, Benson C, Biondi DM, Katz NP, Kim M. Validation of the Short-Form McGill Pain Questionnaire-2 (SF-MPQ-2) in Acute Low Back Pain. *J Pain* 2015;16(4):357–66.
 25. Lovejoy TI, Turk DC, Morasco BJ. Evaluation of the psychometric properties of the revised short-form McGill Pain Questionnaire. *J Pain* 2012;13(12):1250–7.
 26. Turner K V., Moreton BM, Walsh DA, Lincoln NB. Reliability and responsiveness of measures of pain in people with osteoarthritis of the knee: a psychometric evaluation. *Disabil Rehabil.* 2017;39(8):822–9.
 27. Bean DJ, Johnson MH, Heiss-Dunlop W, Lee AC, Kydd RR. Do psychological factors influence recovery from Complex Regional Pain Syndrome Type-1? A Prospective Study. *Pain.* 2015;156(11):2310–8.
 28. Bean DJ, Johnson MH, Heiss-Dunlop W, Kydd RR. Factors Associated with Disability and Sick Leave in Early Complex Regional Pain Syndrome Type-1. *Clin J Pain.* 2016 Feb;32(2):130-8.
 29. Galer BS, Bruehl S, Harden RN. IASP diagnostic criteria for complex regional pain syndrome: a preliminary empirical validation study. International Association for the Study of Pain. *Clin J Pain.* 1998 Mar;14(1):48–54.
 30. Merskey H, Bogduk N, (eds). *Classifications of chronic pain: description of pain syndromes and definition of pain terms.*(2nd ed.). Seattle, WA; 1994.
 31. Harden RN, Bruehl S, Perez RSGM, Birklein F, Marinus J, Maihofner C, Lubenow T, Buvanendran A, Mackey S, Graciosa J, Mogilevski M, Ramsden C, Chont M, Vatine JJ. Development of a severity score for CRPS. *Pain* 2010;151(3):870–6.
 32. Tait RC, Pollard CA, Margolis RB, Duckro PN, Krause SJ. The Pain Disability Index: psychometric and validity data. *Arch Phys Med Rehabil.* 1987 Jul;68(7):438–41.
 33. Pollard CA. Preliminary validity study of the pain disability index. *Percept Mot Skills.* 1984;59(3):974.
 34. Woby SR, Roach NK, Urmston M, Watson PJ. Psychometric properties of the TSK-11: A shortened version of the Tampa Scale for Kinesiophobia. *Pain.* 2005 Sep;117(1–2):137–44.
 35. Sullivan M. *The Pain Catastrophizing Scale User Manual.* McGill University, Montreal; 2009.
 36. Lovibond PF, Lovibond SH. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther.* 1995;33(3):335–43.
 37. Lewis J, McCabe C. Correcting the body in mind: body perception disturbances in complex regional pain syndrome (CRPS) and rehabilitation approaches. *Pr Pain Manag.*

- 2010;April:60–6.
38. Harden RN, Bruehl S, Perez RSGM, Birklein F, Marinus J, Maihofner C, Lubenow T, Buvanendran A, Mackey S, Graciosa J, Mogilevski M, Ramsden C, Chont M, Vatine JJ.. Validation of proposed diagnostic criteria (the “Budapest Criteria”) for Complex Regional Pain Syndrome. *Pain*. 2010;150(2):268–74.
 39. Harden RN, Maihofner C, Abousaad E, Vatine J, Kirsling A, Perez RSGM, Kuroda M, Brunner F, Stanton-Hicks MD, Marinus J, van Hilten JJ, Mackey SC, Birklein F, Schlereth T, Mailis-Gagnon A, Graciosa J, Connolly SB, Dayanim D, Massey M, Frank H, Livshitz A, Bruehl SP. A prospective, multisite, international validation of the Complex Regional Pain Syndrome Severity Score. *Pain* 2017;158(8):1430–6.
 40. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: Psychometric properties. *Pain*. 1990;40(2):171–82.
 41. Packham T, Macdermid JC. Measurement properties of the Patient-Rated Wrist and Hand Evaluation: Rasch analysis of responses from a traumatic hand injury population. *J Hand Ther*. 2013;26(3):216–24.
 42. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
 43. Lundgren Nilsson A, Tennant A. Past and present issues in Rasch analysis: The Functional Independence Measure (FIMTM) revisited. *J Rehabil Med*. 2011;43:884–91.
 44. Linacre J. Investigating rating scale category utility. *J Outcome Meas*. 1999;3(2):103–22.
 45. Lai J, Eton D. Clinically meaningful gaps. *Rasch Meas Trans*. 2002;15(4):850.
 46. Smith E. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal components analysis of residuals. *J Appl Meas*. 2002;3(2):205–31.
 47. Streiner DL, Norman GR. *Health Measurement Scales: a practical guide to their development and use*. (4th ed). Oxford, UK: Oxford University Press; 2008.
 48. Nunnally J. *Psychometric theory*. (2nd ed). New York: McGraw-Hill; 1978.
 49. de Mos M, de Bruijn a. GJ, Huygen FJPM, Dieleman JP, Stricker BHC, Sturkenboom MCJM. The incidence of complex regional pain syndrome: A population-based study. *Pain*. 2007;129(1–2):12–20.
 50. Linacre J. Sample size and item calibration stability. *Rasch Meas Trans*. 1994;7:328.
 51. Birklein F, O’Neill D, Schlereth T. Complex regional pain syndrome: An optimistic perspective. *Neurology*. 2015;84(1):89–96.
 52. Zyluk A. The natural history of post-traumatic sympathetic dystrophy. *J Hand Surg - Br*. 1998;23B(1):20–3.
 53. Pallant JF, Tennant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol*. 2007;46(1):1–18.
 54. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HCW. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*.

2010;10(1):22.

