

## How Good Is the Neurophysiology of Pain Questionnaire? A Rasch Analysis of Psychometric Properties

Mark J. Catley,<sup>\*</sup> Neil E. O'Connell,<sup>†</sup> and G. Lorimer Moseley<sup>\*,‡</sup>

<sup>\*</sup>Sansom Institute for Health Research, University of South Australia, Adelaide, Australia.

<sup>†</sup>Centre for Research in Rehabilitation, Brunel University, Uxbridge, United Kingdom.

<sup>‡</sup>Neuroscience Research Australia, Sydney, Australia.

**Abstract:** The Neurophysiology of Pain Questionnaire (NPQ) was devised to assess how an individual conceptualizes the biological mechanisms that underpin his or her pain. Despite its widespread use, its psychometric properties have not been comprehensively interrogated. Rasch analysis was undertaken on NPQ data from a convenience sample of 300 spinal pain patients, and test-retest reliability was assessed in a sample of 45 low back pain patients. The NPQ effectively targeted the ability of the sample and had acceptable internal consistency and test-retest reliability. However, some items functioned erratically for persons of differing abilities or were psychometrically redundant. The NPQ was reanalyzed with 7 questionable items excluded, and superior psychometric properties were observed. These findings suggest that the NPQ could be improved but future prospective studies including qualitative measures are needed. In summary, the NPQ is a useful tool for assessing a patient's conceptualization of the biological mechanisms that underpin his or her pain and for evaluating the effects of cognitive interventions in clinical practice and research. These findings suggest that it has adequate psychometric properties for use with chronic spinal pain patients.

**Perspective:** Rasch analysis was used to analyze the NPQ. Despite several limitations, these results suggest that it is a useful tool with which to assess a patient's conceptualization of the biological mechanisms that underpin his or her pain and to evaluate the effects of cognitive interventions in clinical practice and research.

© 2013 by the American Pain Society

**Key words:** Chronic pain, pain reconceptualization, pain knowledge, pain education, neurophysiology.

Pain is determined not simply by tissue damage and subsequent nociception, but by a suite of influences, which together imply a need to protect body tissue.<sup>17</sup> A critical element of this multifactorial process is the individual's understanding of what is wrong. This is particularly important in chronic pain, because altered sensitivity within spinal and cortical nociceptive networks render tenuous the relationship between tissue injury, nociceptive input, and pain.<sup>34</sup> However, if the patient has no understanding of these changes, then persistence of, or increases in, his or her pain may be seen to reflect continuing, or worsening, in-

jury. A patient's understanding of the biological mechanisms that underpin his or her pain can be expected to modulate the pain itself.

Educational interventions designed to explain pain have exploited this concept, and the data suggest positive effects. In isolation, explaining pain can alter pain-related attitudes and beliefs,<sup>19,22,23</sup> reduce catastrophizing, and increase physical performance.<sup>22</sup> When combined with conventional physical therapy, explaining pain is associated with reductions in pain and disability.<sup>4,28,32</sup> Indeed, systematic review evidence supports the role of explaining pain in chronic pain management.<sup>15,28</sup>

The Neurophysiology of Pain Questionnaire (NPQ)<sup>21</sup> was devised as a method of assessing how an individual conceptualizes his or her pain. Based on current knowledge,<sup>17</sup> it contains items that were originally based on postgraduate pain medicine exam papers, then modified in terms of language through a process of trial and error.<sup>21</sup> The items assess how and why pain is perceived, and the biological mechanisms that underpin pain.

Received July 2, 2012; Revised February 13, 2013; Accepted February 20, 2013.

No funding sources were provided.

There is no conflict of interest in the conduct of this research.

Address reprint requests to G. Lorimer Moseley, University of South Australia, GPO Box 2084, Adelaide 5001, Australia. E-mail: [lorimer.moseley@gmail.com](mailto:lorimer.moseley@gmail.com)

1526-5900/\$36.00

© 2013 by the American Pain Society

<http://dx.doi.org/10.1016/j.jpain.2013.02.008>

## 2 The Journal of Pain

While primarily a test of pain-related knowledge, it has been used in clinical studies to monitor knowledge change with education-based interventions.<sup>19,21,22</sup> It has also been used extensively in clinical practice to identify gaps in patient knowledge, to effectively target education-based interventions, and to evaluate knowledge transfer at a patient-specific level.

Measurement and monitoring of an individual's knowledge requires a tool with adequate psychometric properties. Despite the NPQs widespread use, its psychometric properties have only been partially investigated.<sup>18,20</sup> We used Rasch analysis to comprehensively interrogate NPQ data obtained from a sample of chronic spine pain patients. Rasch analysis is a technique that uses a mathematical model to formally assess outcome scales.<sup>27,31</sup> It is a probabilistic model that proposes that the likelihood of a person's successfully answering a test item is a logistic function of the difference between that person's ability and the difficulty of the item. The Rasch model assumes that all test items assess a single underlying trait and thus form a unidimensional scale.<sup>2</sup> Measurement theory asserts that all items in a scale can be validly summated to provide a total score only if the scale is unidimensional.<sup>16</sup> Rasch analysis also interrogates other issues crucial to measurement including internal consistency, item invariance, category ordering, and item bias.<sup>31</sup> For this reason, Rasch analysis has many benefits over classical test theory approaches.<sup>33</sup> Our aim therefore was to determine the psychometric properties of the NPQ and identify any questions that threatened its validity in a sample of patients with spinal pain.

## Methods

### Sample and Data Collection

Data were previously collected as part of a retrospective clinical audit conducted across several physiotherapy clinics between 2001 and 2010. Patients with ongoing spinal (neck or back) pain for a period greater than 3 months who were eligible to participate in individualized pain education sessions, who had completed the NPQ, and who were 18 years of age or older were included. Complete NPQ and demographic data were available for 553 patients (see Table 1).

Test-retest reliability was assessed on a second convenience sample of low back pain patients who met the aforementioned inclusion criteria. Consecutive patients from 1 of the clinics were assessed on 2 occasions, once prior to education and once after 6 to 8 half-hour sessions of education. On each occasion, the questionnaire was administered twice, with a 2- to 5-day interval between. Permission to access and analyze the deidentified clinical data used for this study was provided by the University of South Australia Human Research Ethics Committee.

### Study Instrument

The NPQ is self-administered and contains 19 items relating to the neurophysiology of pain.<sup>21</sup> Each item has the following response options: true, false, undecided.

How Good Is the NPQ? Assessment With Rasch Analysis

**Table 1. Participant Demographic Characteristics**

	CLINICAL AUDIT DATA (N = 553)	RASCH SAMPLE (N = 300)	TEST-RETEST SAMPLE (N = 45)
Gender, n (%)			
Males	216 (39.1)	120 (40.0)	21 (46.7)
Females	337 (60.9)	180 (60.0)	24 (53.3)
Age (years), mean (SD)	39.9 (13.4)	39.8 (13.7)	41.8 (11.2)
Diagnosis, n (%)			
Chronic back pain	385 (69.6)	209 (69.7)	45 (100)
Chronic neck pain	168 (30.4)	91 (30.3)	—
Duration of pain (weeks), mean (SD)	53.2 (29.8)	52.3 (29.6)	35.8 (30.9)
NPQ score (%), mean (SD)			
Preeducation	26.4 (13.5)	25.8 (13.1)	22.6 (14.7)
Posteducation	63.2 (15.7)	62.6 (15.2)	73.2 (23.5)

The NPQ is scored out of 19 with 1 point awarded for each correct response. A score of 0 is attributed to incorrect responses and those marked as undecided. The NPQ is shown in Table 2.

## Data Analysis

### Data Extraction

For analysis, a dataset of 300 persons was drawn from the wider sample using a random number generator. The NPQ is intended for use in persons with no pain biology knowledge to interrogate their beliefs regarding pain. It is also used to measure change in knowledge and identify gaps in knowledge after pain education. That is, the NPQ is used to assess people of all ability levels. For this reason, we generated a sample that would reflect this range of knowledge by randomly selecting the preeducation NPQ score of 150 persons and the posteducation NPQ scores of the other 150. The distribution of raw scores was assessed for normality using a D'Agostino's K-squared test.<sup>5</sup> Allowing for removal of persons with extreme scores, a sample in excess of 243 persons is required for Rasch analysis to ensure item calibration stability within  $\pm 0.5$  logits with 99% confidence.<sup>12,14</sup>

Descriptive statistics were calculated using PASW Statistics 18 (v.18.0.0; IBM Corporation, Armonk, NY) and the dichotomous Rasch model was analyzed using Winsteps (v.3.73.0; [www.winsteps.com](http://www.winsteps.com)) software.

### Rasch Analysis

The steps involved in Rasch analysis have been extensively described in the literature.<sup>1,2,14</sup> We assessed the following Rasch analysis components.

### Targeting

Targeting refers to the extent to which items target the abilities of the sampled persons. Item difficulties should match the ability range of the questionnaire's intended sample with difficulty thresholds dispersed evenly across the logit scale, rather than in clusters. Targeting is

**Table 2. Patient Version of the Neurophysiology of Pain Test Adapted From Moseley<sup>19</sup>**

ITEM		T	F	U
1	Receptors on nerves work by opening ion channels in the wall of the nerve.	#		
2	When part of your body is injured, special pain receptors convey the pain message to your brain.		#	
3	Pain only occurs when you are injured or at risk of being injured.		#	
4	Special nerves in your spinal cord convey "danger" messages to your brain.	#		
5	Pain is not possible when there are no nerve messages coming from the painful body part.		#	
6	Pain occurs whenever you are injured.		#	
7	The brain sends messages down your spinal cord that can change the message going up your spinal cord.	#		
8	The brain decides when you will experience pain.	#		
9	Nerves adapt by increasing their resting level of excitement.	#		
10	Chronic pain means that an injury hasn't healed properly.		#	
11	The body tells the brain when it is in pain.		#	
12	Nerves can adapt by producing more receptors.	#		
13	Worse injuries always result in worse pain.		#	
14	Nerves adapt by making ion channels stay open longer.	#		
15	Descending neurons are always inhibitory.		#	
16	When you injure yourself, the environment that you are in will not affect the amount of pain you experience, as long as the injury is exactly the same.		#	
17	It is possible to have pain and not know about it.		#	
18	When you are injured, special receptors convey the danger message to your spinal cord.	#		
19	All other things being equal, an identical finger injury will probably hurt the left little finger more than the right little finger in a violinist but not a piano player.	#		

Abbreviations: T, true; F, false; U, undecided.

NOTE. #Denotes the correct answer.

assessed by visual inspection of the distribution of person and item thresholds and comparison of the summary statistics. Items were considered too easy if over 95% of persons answered correctly and too hard if under 5% were answered correctly.

### Unidimensionality

Rasch analysis examines the extent to which items in a scale measure the same underlying construct. Unidimensionality was assessed through analysis of item fit statistics and through principal component analysis of residuals. Item fit statistics indicate how well each item contributes to the measurement of a single underlying construct. That is, they guide the Rasch user to analyze items that do not function in accordance with the Rasch model. Fit statistics are chi-square based and are reported as mean-squares (in logits), with an expected value of 1 logit. Excessively large fit residuals indicate items with large differences between the expected and observed performance of an item.<sup>9</sup> According to Rasch theory, items that perform unexpectedly may bias subgroups within the sample, be dependent on the responses to other items, or assess a construct other than that intended.<sup>2</sup> Smaller fit residuals indicate items that behave too predictably, and are therefore less of a threat to measurement.<sup>2</sup> Sample size appropriate cut-off values of 1.3 and .7 logits were chosen to determine underfit and overfit to the model, respectively.<sup>9</sup>

Both infit and outfit item statistics were examined. Infit item statistics are weighted toward the person performances that are close to the item threshold and therefore pose a greater threat to validity. The outfit statistic is unweighted and is sensitive to the influence of random outlying scores but is considered less of a threat to validity. In addition to fit, the item character-

istic curves for each misfitting item were visually inspected to further identify items that functioned unexpectedly.

Principal component analysis of residuals was conducted via visual inspection of the residual correlation matrix, to identify unexpected patterns in the data. Groups of items sharing the same pattern of unexpectedness may also share a substantive attribute in common, thus constituting a second dimension. Items with substantial positive or negative loadings equivalent to an eigenvalue greater than 2 (the strength of 2 items) were reviewed for possible multidimensionality. Rasch analysis also assumes that items have local independence; that is, the response to one item should be independent of the responses to other items when the underlying construct is controlled for.<sup>29</sup> No association between residuals for individual items provides evidence of local independence, whereas high correlation between residuals provides evidence of local dependence.<sup>35</sup> The standardized residual correlations were examined to identify dependent items. Items with residuals that highly correlate were reviewed as to whether they duplicate each other.

### Person Fit

The true/false nature of the NPQ makes it susceptible to guesses.<sup>8</sup> Rasch software provides fit statistics as a means to identify persons who respond erratically. Identification of persons with acceptable infit statistics (weighted toward the item thresholds close to the person's ability) but excessive outfit statistics (unweighted) is considered to be indicative of lucky guesses or careless mistakes.<sup>2</sup> We examined the response strings of low achieving persons with outfit residuals in excess of 1.5 logits.<sup>26</sup> Misfitting persons were compared to those

#### 4 The Journal of Pain

who fit the model using a chi-square test of significance (for gender and diagnosis) or a 1-way analysis of variance (for age and ability).<sup>11</sup>

#### Internal Consistency

Rasch analysis provides a person separation index (PSI) as an indicator of internal consistency reliability.<sup>31</sup> This is the Rasch equivalent of Cronbach's alpha and can thus be interpreted in a similar way. A PSI of  $<.80$  suggests that an instrument may not be sensitive enough to distinguish between high and low performers.<sup>14</sup>

#### Differential Item Functioning

For the NPQ to be useful, it must be able to assess a wide range of patients. The items should function similarly for all persons of the same level of ability. To ensure that characteristics other than item difficulty do not bias the functioning of the scale, differential item functioning was assessed between 4 subgroups: ability (low ability  $\leq 10$  NPQ score, high ability  $> 10$  NPQ score), diagnosis (chronic neck pain, chronic back pain), gender (male, female), and age (younger than 18–40 years, older than 41–64 years). Differential item functioning was tested using a Mantel-Haenszel chi-square test with significance set at  $\alpha = .01$  for each item.

#### Removal of Poorly Functioning Items and Reassessment

Items were reviewed and considered for removal if they 1) had excessive fit statistics; 2) functioned differently between subgroups; or 3) breached the assumption of local independence.

The remaining items were then interrogated via a second Rasch to assess whether the psychometric properties were superior and acceptable for clinical and research purposes.

#### Reliability

Reliability was assessed on the NPQ data obtained from the second sample of low back pain patients. Test-retest reliability was evaluated via intraclass correlation coefficients (ICCs). D'Agostino's K-squared test was used to assess whether the data were normally distributed prior to analysis. An ICC 2-way mixed model with an absolute agreement type was chosen; values  $\geq .75$  were interpreted as good reliability whereas those  $< .75$  indicated poor-to-moderate reliability.<sup>25</sup>

### Results

#### Participants

Table 1 displays the demographic and clinical characteristics of the samples. Rasch analysis was conducted on 300 randomly selected questionnaires. The raw scores were normally distributed, which suggests that combining pre- and posteducation scores created a sample that reflected a wide range of abilities. Test-retest reliability of the NPQ was assessed on a sample of 45 low back pain patients.

How Good Is the NPQ? Assessment With Rasch Analysis

### Rasch Analysis

#### Targeting

Fig 1A displays the item-person distribution and Table 3 shows the item difficulty thresholds. Item difficulty ranged from  $-3.72$  to  $2.68$  logits and adequately targeted the ability of the sample. Visually, the distribution of the persons reflected that of the item distribution, and the mean person location (mean, SD:  $-.35, 1.64$  logits) was similar to that of the items (mean SD:  $.0, 1.69$  logits). This suggests that the NPQ targeted the sample adequately. Four (1.3%) persons registered an extreme score (zero correct answers), suggesting a minimal floor effect and no ceiling effect. These 4 were excluded from further analysis by the Rasch software, leaving 296 persons for item fit analysis. Individual items were reasonably dispersed across the difficulty range, and person ability was adequately targeted by each of the items. None of the questions were deemed too hard or too easy.

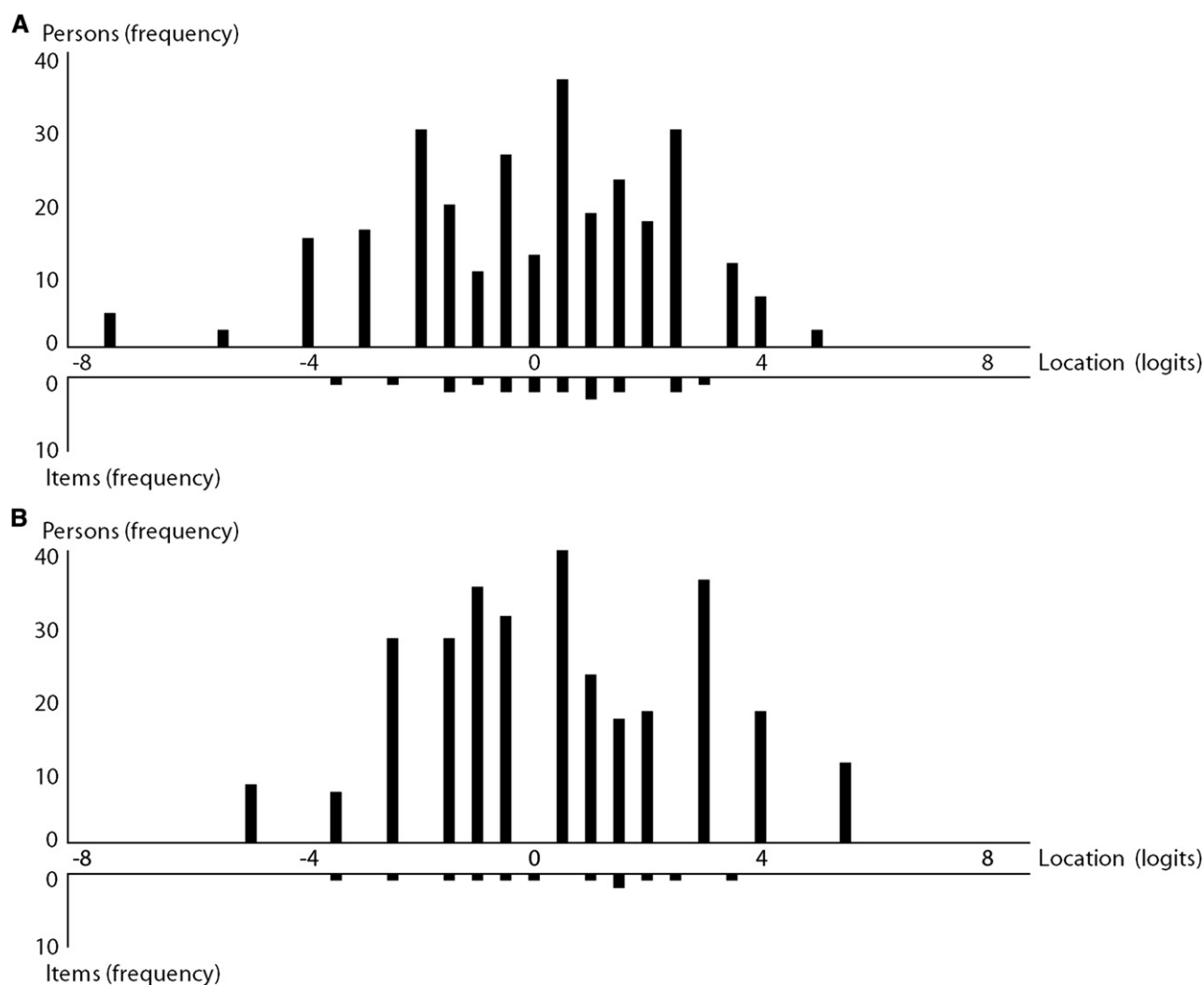
#### Unidimensionality

Table 3 summarizes the fit statistics for the 19 items. Five of the 19 items (26%) demonstrated excessive infit statistics. Three items (Q1, Q7, Q12) showed excessive positive infit, which means they functioned worse for the persons they specifically targeted and may threaten the validity of the scale.<sup>2</sup> Two items (Q5, Q6) showed excessive negative infit, which means they performed too predictably, although they do not reflect a threat to validity. Eleven items demonstrated excessive outfit, which means they were affected by the responses of outlying persons but were less of a threat to validity.<sup>2</sup> Five items (Q1, Q7, Q12, Q14, Q16) showed excessive positive outfit, suggestive of haphazard responses, while 6 items (Q2, Q5, Q6, Q10, Q13, Q17) showed excessive negative outfit, suggesting they were overly predictable.

The item characteristic curves for each of the 11 misfitting items were reviewed. Three of the 11 misfitting items (Q1, Q7, Q12) functioned poorly across the ability range (Fig 2). The remaining 8 misfitting items were deemed acceptable.

Principal component analysis suggested that certain items may be assessing a construct other than pain neurophysiology knowledge. Three items (Q9 "Nerves adapt by increasing their resting level of excitement"; Q12 "Nerves can adapt by producing more receptors"; Q14 "Nerves adapt by making ion channels stay open longer") with negative loadings were separated from the other items toward the bottom of the residual correlation matrix, suggesting the possibility of a second dimension. An eigenvalue of 3.2 supported this possibility.<sup>29</sup> The contents of these 3 items were reviewed and were considered to specifically relate to nerve physiology.

Assessment of local dependence identified strong positive correlations between questions 5 and 6 ( $r = .70$ ), 9 and 14 ( $r = .66$ ), 9 and 12 ( $r = .60$ ), 12 and 14 ( $r = .55$ ), and 8 and 11 ( $r = .55$ ). This suggested that the responses to these items were dependent on knowledge of the others.



**Figure 1.** Item-person threshold map for the (A) neurophysiology of pain questionnaire and (B) revised 12-item neurophysiology of pain questionnaire (reanalyzed with questions 1, 5, 7, 11, 12, 14, and 19 excluded). Note: Persons of lesser ability and items of lesser difficulty are located on the left side of the logit scale (ie, <0 logits); persons of higher ability and items of greater difficulty are located to the right of the logit scale (ie, >0 logits). Item difficulty mean is set at 0 logits by default.

### Person Fit

Analysis of person fit identified 54 persons (18%) with excessive positive outfit. Analysis of these data showed no significant associations between age ( $P = .845$ ), gender ( $P = .851$ ), and diagnosis ( $P = .623$ ), which suggests that they did not differ from the persons who fit the model in regards to sample characteristics. However, misfitting persons were of significantly lower ability (mean difference 1.9 [raw score],  $P = .003$ ) and therefore possibly more likely to guess or use the undecided option. Examination of the misfitting persons' response strings showed typically that low ability persons had achieved success on 1 or 2 difficult items, contributing to the outfit.

### Internal Consistency

The PSI value was .84, which suggests that the NPQ is sensitive enough to distinguish between high and low performers.<sup>14</sup>

### Differential Item Functioning

Fig 3 displays scatterplots for the 4 subgroups. No significant differences in item functioning were observed for gender, age, or diagnosis. Low ability persons found 5 questions (Q4 [ $P = .002$ ], Q5 [ $P < .001$ ], Q6 [ $P < .001$ ], Q10 [ $P < .001$ ], Q17 [ $P < .001$ ]) significantly harder than high ability persons. High ability persons found 4 questions (Q1 [ $P < .001$ ], Q11 [ $P < .001$ ], Q14 [ $P = .005$ ], Q19 [ $P = .005$ ]) significantly harder than low ability persons.

### Removal of Items and Reanalysis

Seven items were considered to function poorly. Questions 1, 11, 14, and 19 were significantly harder for higher ability persons than for lower ability persons. Question 14 also showed local dependence with questions 9 and 12. Question 5 demonstrated local dependence with question 6 and of the two had the poorer fit statistics. Questions 7 and 12 exhibited excessive misfit and



**Table 3. Categorical Order, Item Difficulty Thresholds, and Fit Statistics for the Neurophysiology of Pain Questionnaire**

ITEM	RASCH ANALYSIS OF NPQ (N = 296*)					RASCH ANALYSIS OF REVISED NPQ (N = 296*)				
	MEASURE (LOGITS)	SE	CORRECT RESPONSES	INFIT (MNSQ)	OUTFIT (MNSQ)	MEASURE (LOGITS)	SE	INFIT (MNSQ)	OUTFIT (MNSQ)	
14	2.68	.21	10%	1.18	1.91†	—	—	—	—	
12	2.27	.19	14%	1.40†	1.70†	—	—	—	—	
9	2.23	.19	14%	1.03	1.29	3.40	.22	1.37†	9.90†	
17	1.48	.17	22%	.74	.46†	2.42	.19	.77	.45†	
11	1.45	.16	22%	.87	1.12	—	—	—	—	
2	.93	.15	29%	.91	.67†	1.70	.17	1.30	.99	
7	.72	.15	32%	1.33†	1.31†	—	—	—	—	
15	.54	.15	35%	.91	.75	1.21	.17	1.11	1.11	
18	.50	.15	36%	.81	.86	1.16	.17	.90	1.02	
8	.18	.14	41%	.93	.94	.76	.16	1.12	1.61†	
19	-.05	.14	44%	1.15	1.17	—	—	—	—	
1	-.25	.14	48%	1.80†	2.57†	—	—	—	—	
5	-.75	.14	56%	.59†	.45†	—	—	—	—	
6	-.79	.14	56%	.64†	.49†	-.40	.16	.65†	.45†	
4	-1.12	.15	62%	.87	.72	-.79	.16	.93	.78	
10	-1.54	.15	68%	.71	.52†	-1.27	.16	.64†	.43†	
3	-1.87	.16	73%	.98	.88	-1.65	.17	.99	.85	
16	-2.89	.19	84%	1.06	3.81†	-2.79	.20	1.12	3.49†	
13	-3.72	.23	91%	1.01	.68†	-3.74	.26	1.04	.62†	

Abbreviation: MNSQ, mean square.

\*4 persons with extreme scores were omitted prior to item fit analysis.

†Indicates item misfit.

functioned poorly across the ability range. The NPQ data were reanalyzed with these 7 poorly functioning items removed. Fig 1B displays the item-person distribution and Table 3 displays the item difficulty thresholds and fit statistics of the revised 12-item NPQ. Item ordering remained unchanged and a PSI value of .82 indicated that the revised NPQ remained suitable for individual use. It still effectively targeted the sample with comparable person (mean, SD:  $-.08, 1.95$ ) and item (mean, SD:  $.0, 2.05$ ) distributions. Item difficulty thresholds ranged from  $-3.74$  to  $3.40$  logits. Eleven persons (4%) scored the maximum score and 8 persons (3%) scored zero, suggesting a minimal floor and ceiling effect. Question 9 had excessive positive infit and outfit and Questions 8 and 16 had slightly excessive positive outfit. No significant item bias was observed for gender, diagnosis, and age. Questions 6 [ $P = .009$ ] and 10 [ $P = .002$ ] functioned significantly differently for ability with lower ability persons, finding them .93 and 1.43 logits, respectively, harder than higher ability persons. Importantly, no questions were found significantly harder by higher ability persons than lower ability persons. The removal of questions 12 and 14 meant that the principal component analysis of residuals identified no pattern in the residual correlation matrix, suggesting that the revised version constitutes a unidimensional scale. A second dimension eigenvalue of 1.8 supported this suggestion. Overall, removal of the 7 poorly functioning items resulted in superior psychometric properties.

### Reliability

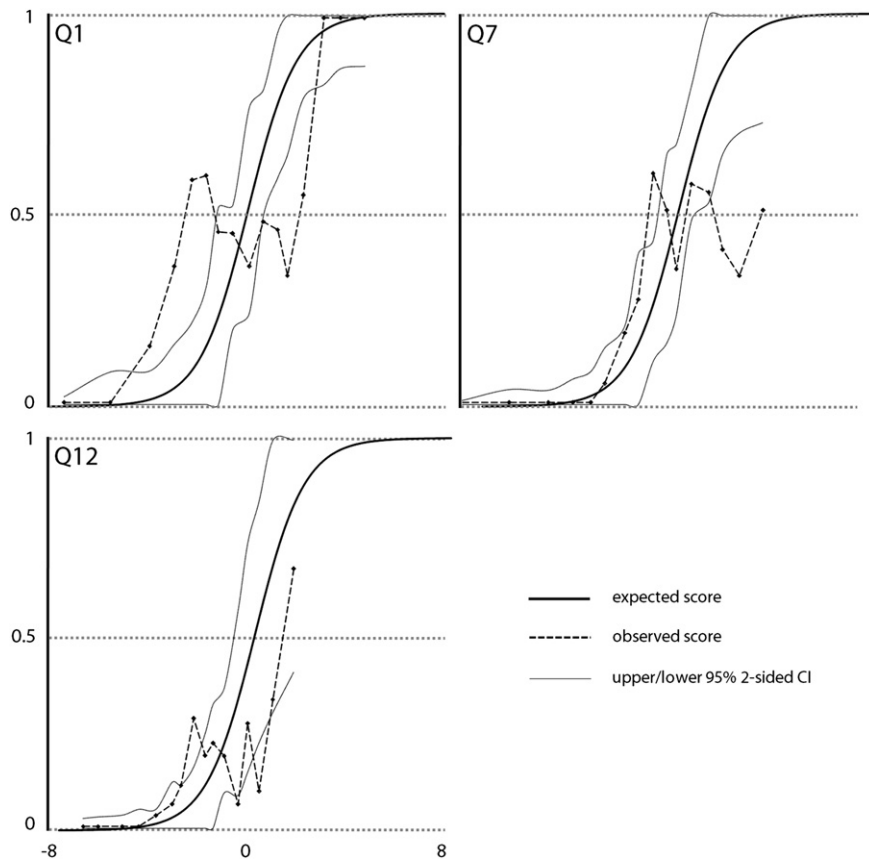
Reliability was assessed in a second sample of low back pain patients. A preeducation ICC of .971 (95% con-

fidence interval [CI]: .925–.987) and a posteducation ICC of .989 (95% CI: .981–.984) suggest that the NPQ has good test-retest reliability.

### Discussion

We used Rasch analysis to evaluate the psychometric properties of the NPQ in a sample of chronic spine pain patients. Our results show that the NPQ has acceptable internal consistency for assessment of individuals, effectively targets the ability of a typical group of chronic pain patients, constitutes a unidimensional scale, and has good test-retest reliability. That only 4 persons scored zero suggests that the NPQ possesses neither floor nor ceiling effects.

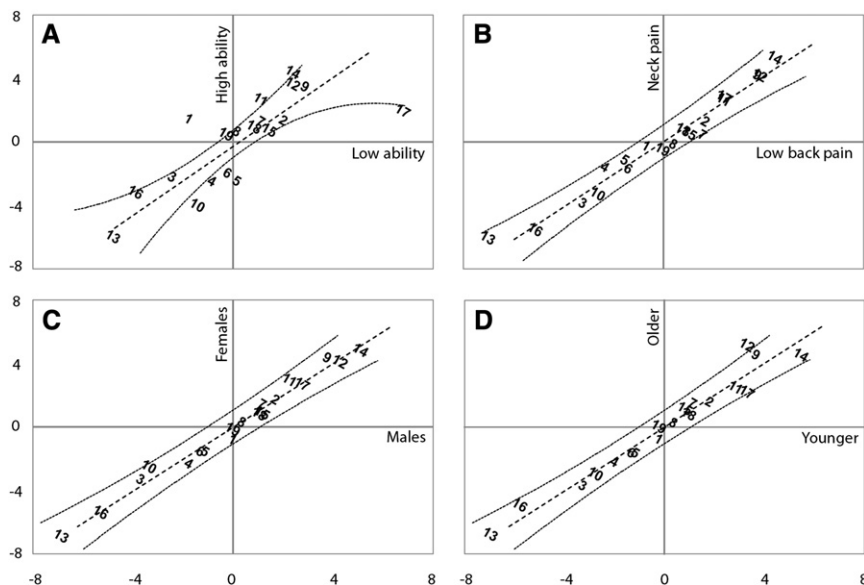
However, the analysis identified several items that functioned poorly, exhibited bias, or were psychometrically redundant. Unfortunately, the retrospective nature of this study meant we were unable to interview the sample; hence, we can only hypothesize why these questions functioned poorly. The Rasch model presumes that all persons are more likely to answer correctly the easier items and that higher ability persons are more likely to answer any item correctly. The fit statistics identify items that do not meet this presumption. Question 1 (“Receptors on nerves work by opening ion channels in the wall of the nerve”) functioned unpredictably for the whole sample, and questions 7 (“The brain sends messages down your spinal cord that can change the message going up your spinal cord”) and 12 (“Nerves can adapt by producing more receptors”) functioned unpredictably for persons with higher ability.<sup>31</sup> That is, responses to these items could not be predicted by



**Figure 2.** Item characteristic curves for items with excessive infit. Items 1, 7, and 12 demonstrated excessive positive infit and were deemed to function poorly.

a person's overall ability, implying that factors other than pain knowledge influence the responses to these questions. The item difficulty in terms of ordering should also function similarly for all persons regardless of their ability. That questions 14 ("Nerves adapt by making ion

channels stay open longer") and 19 ("All other things being equal, an identical finger injury will probably hurt the left little finger more than the right little finger in a violinist but not a piano player") in addition to question 1 were significantly easier for persons of low



**Figure 3.** Scatterplots displaying the relationship between the subgroups for (A) ability, (B) diagnosis, (C) gender, and (D) age. Note: The dashed line represents the Rasch modeled relationship required for invariance and the curved lines represent the upper and lower 95% control.

## 8 The Journal of Pain

ability is problematic. These findings could be explained by guessing. It might be that persons of low ability may systematically guess to the positive, whereas more able persons may consider the question more carefully, thus, ironically, increasing their chance of making a mistake.

It is also concerning that several questions breached the assumption of local dependence. High correlations, greater than .7, mean that items share a large part of their random variance, which suggests that only 1 is needed for measurement.<sup>14</sup> The highest correlation was noted between questions 5 ("Pain is not possible when there are no nerve messages coming from the painful body part") and 6 ("Pain occurs whenever you are injured"), implying that knowledge of 1 preempted knowledge of the other. High correlations were also noted between questions 9, 12, and 14, which all related to nerve function, suggesting that 1 or 2 of these may also be psychometrically redundant.

Reanalysis of the NPQ with the poorly functioning or redundant items excluded demonstrated better fit to the Rasch model. The revised version targeted the sample effectively, possessed minimal floor and ceiling effects, showed reasonable internal consistency, and was less influenced by poor fit statistics and item bias than the original NPQ. Our findings suggest that while 2 items in the revised questionnaire display some degree of misfit, it did not exceed accepted limits,<sup>2,14</sup> and that the removal of questions 12 and 14 meant the unidimensionality of the revised version was no longer in question and the issue of local dependence was resolved. A PSI of .82 suggests that it is suitable to assess pain-related knowledge in individuals and to monitor their change in knowledge. The revised version would be 30% shorter than the original, which would have clear time and recording advantages within both a busy clinical situation and a research context. While a revised, shortened version of the NPQ appears to have superior psychometric qualities, future research is needed to test it in an independent sample.

Should the NPQ include questions relating specifically to nerve function? Intuitively, an understanding of the physiological adaptations that occur in response to injury could provide the mechanistic evidence that will aid some individuals to reconceptualize their pain. These questions are therefore better suited to assess a person's change in knowledge, as we would not expect uninformed persons to answer correctly. Likewise, these questions will not function well for the assessment of pain beliefs and attitudes. It is probably for this reason that these 4 questions caused the most problems. Question 1 functioned erratically and questions 14 ("Nerves adapt by making ion channels stay open longer"), 12 ("Nerves can adapt by producing more receptors"), and 9 ("Nerves adapt by increasing their resting level of excitement") contributed the 3 most difficult items. Together, it seems these concepts are the hardest for laypersons to grasp. Perhaps these concepts are not well covered by the educators or the hardest to teach. We recently reported that pain reconceptualization is improved with the use of stories and metaphors.<sup>10</sup> That stories and metaphors help

How Good Is the NPQ? Assessment With Rasch Analysis entrench abstract concepts is well reported,<sup>3,7</sup> and we propose that those stories and metaphors lend themselves better to the general concepts regarding pain than they do to specific biological mechanisms. Regardless, if these items are to be included, future educators will need to ensure that nerve function is adequately addressed. Interestingly, not all persons struggled with these questions. Questions 9, 12, and 14 together constituted a possible second dimension, and the analysis of person response strings confirmed that a subgroup within the sample had a good understanding of nerve-related physiology, but not of the mechanisms of pain. This finding is perhaps not surprising because basic physiology is widely taught in school and university. It is plausible that a person could understand these principles but not understand the mechanisms of pain. Unfortunately, the retrospective design of this study prevented us from exploring any of these hypotheses in further detail. We suggest that future prospective studies include qualitative measures to further interrogate how the items are being utilized. We also question whether the NPQ would function better as a measurement of knowledge change or of attitudes and beliefs, but not both.

True-false questionnaires are easy to administer and analyze and, despite criticism, have been shown to be of similar difficulty and validity (with regard to assessing ability) to multiple choice tests.<sup>30</sup> The NPQ uses the number-correct scoring method, whereby the total number of correct answers is summed to create an overall score. However, it also includes an undecided option that is scored as zero. Undecided options are usually reserved for formula-based tests (eg, negatively scored) or partial credit models. Unfortunately, we have no method of determining the proportion of "undecided" responses in the current sample. That is, we only had access to dichotomous data (whether a person answered correctly or incorrectly) and were unable to determine the extent to which the undecided option was used. For this reason, we were unable to test whether an alternative scoring system would be superior to that used here and cannot make definitive recommendations regarding the NPQ's format. Interestingly, our results suggest that up to 18% of the sample may have been guessing for several items on the NPQ. This is substantially greater than the 5 to 8% reported as normal by others.<sup>8</sup> It also cannot be definitively explained by use of the "undecided" option, as data from a separate cohort suggested that "undecided" is selected in only 3% of responses (Moseley GL, unpublished data). While we contend that guessing may be a problem, the current scoring method does, however, appear appropriate in that the inclusion of an "undecided" option allows for the identification of gaps in patients' knowledge, while incorrect responses allow for the identification of mistaken beliefs. A true-false format would force recipients to guess, and formula-based tests would be inappropriate because they bias against students with partial knowledge.<sup>24</sup> We can only recommend that future recipients of the NPQ be advised to utilize the undecided option and avoid lucky guesses.



The NPQ has been used to identify knowledge deficits in patients, and the categorical order identified by this study will enable clinicians to analyze a patient's score in depth. For example, unexpected correct answers to difficult questions by low ability persons can identify special knowledge in patients or, conversely, unexpected incorrect answers to easy questions by high ability persons can identify gaps in knowledge.

The results of this study differ slightly from those of Meeus et al.<sup>18</sup> They concluded that the Dutch version of the NPQ had face validity to assess pain-related knowledge in chronic fatigue syndrome patients but only fair internal consistency (Cronbach's  $\alpha = .769$ ) and fair test-retest reliability (ICC = .756). Given that chronic pain is a common comorbidity associated with chronic fatigue,<sup>6</sup> and that the Rasch model's PSI is reportedly more conservative than Cronbach's alpha,<sup>13</sup> one might have expected similar values. However, there are important points that need to be considered. First, the Dutch version of the NPQ was based on a translation of the original English version.<sup>21</sup> However, the vast majority of clinical use and research in English-speaking cohorts uses a modified version. As such, the 2 questionnaires are not only translated, but slightly different. It may be appropriate, then, to label the version used here as the NPQ (modified). Second, the Dutch cohort found the questionnaire more difficult than the English-speaking cohort did. The Dutch were also more likely to select the undecided option, which may reflect that that cohort did not include educated participants. Our study, however, included educated participants so as to obtain a good spectrum and normal distribution of ability. Finally, translation will always introduce the possibility of variability.

Strengths of this study include random sampling from a large clinical dataset, collected from several sites over many years; the results are therefore highly generalizable to chronic spinal pain patients. We also used a novel

sampling technique to create a dataset representative of varying levels of ability, and a comprehensive analytical technique in the Rasch. Finally, we reanalyzed the NPQ with 7 poorly functioning items removed to verify whether a scale with superior psychometric properties for chronic spine pain patients could be produced. Our study also had several limitations. Guessing appeared to be a problem, although our data only revealed whether the recipients answered correctly or not; hence, we were unable to specifically determine the extent to which participants were guessing or had utilized the "undecided" option. Also, as the data were analyzed well after they were collected, none of the participants could be contacted. Subsequently, assumptions were made regarding whether or not they were guessing. Communication with participants might have helped us determine whether semantic issues contributed to other items functioning poorly, although the influence of recall bias would be difficult to manage. Also, while we have reported good test-retest reliability, we contend that the interval between assessments was between 2 and 5 days and it is plausible that participants simply remembered their previous answers.

In summary, the NPQ is a useful tool with which to assess a patient's conceptualization of the biological mechanisms that underpin his or her pain and to evaluate the effects of cognitive interventions in clinical practice and research. These findings suggest that it has adequate psychometric properties for use with chronic spinal pain patients. Nonetheless, 7 of the 19 items were flawed, and reanalysis with these items excluded produced superior psychometric properties. Further research in an independent sample is needed to verify whether a revised, shortened version is advisable. Finally, the true-false nature of the questionnaire could be susceptible to guesses; therefore, recipients of the NPQ should be advised to use the undecided response option.

## References

1. Alagumalai S, Curtis DD, Hungi N: Applied Rasch Measurement: A Book of Exemplars: Papers in Honour of John P. Keeves. Dordrecht, NL, Kluwer Academic Publishers, 2005
2. Bond TG, Fox CM: Applying the Rasch Model: Fundamental Measurement in the Human Sciences, 2nd ed. Mahwah, NJ, Lawrence Erlbaum Associates Inc, 2007
3. Cameron L: Metaphors in the learning of science: A discourse focus. *Br Educ Res J* 28:673-688, 2002
4. Clarke CL, Ryan CG, Martin DJ: Pain neurophysiology education for the management of individuals with chronic low back pain: A systematic review and meta-analysis. *Man Ther* 16:544-549, 2011
5. D'Agostino RB, Pearson ES: Tests for departure from normality: Empirical results for the distributions of b2 and b1. *Biometrika* 60:613-622, 1973
6. Dansie EJ, Furberg H, Afari N, Buchwald D, Edwards K, Goldberg J, Schur E, Sullivan PF: Conditions comorbid with chronic fatigue in a population-based sample. *Psychosomatics* 53:44-50, 2012
7. Duit R: On the role of analogies and metaphors in learning science. *Sci Educ* 75:649-672, 1991
8. Ebel RL: Blind guessing on objective achievement tests. *J Educa Meas* 5:321-325, 1968
9. Fox CM, Jones JA: Uses of Rasch modeling in counseling psychology research. *J Couns Psychol* 45:30-45, 1998
10. Gallagher L, McAuley J, Moseley GL: A randomised-controlled trial of using a book of metaphors to reconceptualize pain and decrease catastrophizing in people with chronic pain. *Clin J Pain* 29:20-25, 2013
11. Hobart J, Cano S: Improving the evaluation of the therapeutic interventions in multiple sclerosis: The role of new psychometric methods. *Health Technol Assess* 13:1-177, 2009
12. Linacre JM: Sample size and item calibration stability. *Rasch Meas Transac* 7:328, 1994
13. Linacre JM: KR-20 or Rasch reliability: Which tells the "truth"? *Rasch Meas Transac* 11:580-581, 1997

**10 The Journal of Pain**

14. Linacre JM: A user's guide to Winsteps Ministep Rasch-model computer programs: Program manual 3.73.0, [www.winsteps.com](http://www.winsteps.com). 2011
15. Louw A, Diener I, Butler DS, Puentedura EJ: The effect of neuroscience education on pain, disability, anxiety, and stress in chronic musculoskeletal pain. *Arch Phys Med Rehabil* 92:2041-2056, 2011
16. Luce RD, Tukey JW: Simultaneous conjoint measurement: A new type of fundamental measurement. *J Math Psychol* 1:1-27, 1964
17. McMahon SB, Koltzenburg M, Wall PD, Melzack R: Wall and Melzack's Textbook of Pain. Philadelphia, PA, Elsevier, 2006
18. Meeus M, Nijs J, Elsemans KS, Truijen S, De Meirleir K: Development and properties of the Dutch Neurophysiology of Pain Test in patients with Chronic Fatigue Syndrome. *J Musculoskelet Pain* 18:58-65, 2010
19. Meeus M, Nijs J, Van Oosterwijck J, Van Alsenoy V, Truijen S: Pain physiology education improves pain beliefs in patients with chronic fatigue syndrome compared with pacing and self-management education: A double-blind randomised controlled trial. *Arch Phys Med Rehabil* 91:1153-1159, 2010
20. Moseley GL: The Psychophysiology of Pain and Motor Control. Sydney, AU, University of Sydney, 2002
21. Moseley GL: Unraveling the barriers to reconceptualization of the problem in chronic pain: The actual and perceived ability of patients and health professionals to understand the neurophysiology. *J Pain* 4:184-189, 2003
22. Moseley GL: Evidence for a direct relationship between cognitive and physical change during an education intervention in people with chronic low back pain. *Eur J Pain* 8:39-45, 2004
23. Moseley GL, Nicholas MK, Hodges PW: A randomised controlled trial of intensive neurophysiology education in chronic low back pain. *Clin J Pain* 20:324-330, 2004
24. Muijtjens AMM, van Mameren H, Hoogenboom RJI, Evers JLH, van der Vleuten CPM: The effect of a 'don't know' option on test scores: Number-right and formula scoring compared. *Med Educ* 33:267-275, 1999
- How Good Is the NPQ? Assessment With Rasch Analysis
25. Portney LG, Watkins MP: Foundations of Clinical Research: Applications to Practice. Upper Saddle River, NJ, Pearson/Prentice Hall, 2009
26. Raiche G: Critical eigenvalue sizes in standardized residual principal components analysis. *Rasch Meas Transac* 19:1012, 2005
27. Rasch G: Probabilistic Models for Some Intelligence and Attainment Tests. Chicago, IL, University of Chicago, 1960
28. Ryan CG, Gray HG, Newton M, Granat MH: Pain biology education and exercise classes compared to pain biology education alone for individuals with chronic low back pain: A pilot randomised controlled trial. *Man Ther* 15:382-387, 2010
29. Smith EV: Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2:205-231, 2002
30. Tasdemir M: A comparison of multiple-choice tests and true-false tests used in evaluating student progress. *J Instruct Psychol* 37:258-266, 2010
31. 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* 57:1358-1362, 2007
32. van Oosterwijck J, Nijs J, Meeus M, Truijen S, Craps J, Van den Keybus N, Paul L: Pain neurophysiology education improves cognitions, pain thresholds, and movement in people with chronic whiplash: A pilot study. *J Rehabil Res Dev* 48:43-58, 2011
33. Waugh R: An analysis of dimensionality using factor analysis (true score theory) and Rasch measurement: What is the difference? Which method is better? *J Appl Meas* 6:80-99, 2005
34. Woolf CJ: Central sensitization: Implications for the diagnosis and treatment of pain. *Pain* 152:S2-S15, 2011
35. Wright BD: Local dependency, correlations and principal components. *Rasch Measurement Transactions* 10:509-511, 1996