ORIGINAL PAPER

WILEY | Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

Rasch analysis of the Patient Participation in Rehabilitation Questionnaire (PPRQ)

Jeanette Melin PhD, RPT, Researcher[1] (ID) | Robin Fornazar MSc, Public Health[2] |
Martin Spångfors MScN[3,4] | Leslie Pendrill PhD, Docent[1] (ID)

[1] Metrology, RISE Research Institutes of Sweden, Gothenburg, Sweden

[2] Section for Epidemiology and Social Medicine (EPSO), The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[3] Department of Clinical Sciences Lund, Anaesthesiology and Intensive Care Medicine, Lund University, Lund, Sweden

[4] Department of Anaesthesiology and Intensive Care, Kristianstad Hospital, Kristianstad, Sweden

Correspondence
Jeanette Melin, Metrology, RISE Research Institutes of Sweden, Eklandagatan 86, Göteborg 41261, Sweden.
Email: jeanette.melin@ri.se

Abstract

Objective: To evaluate the Patient Participation in Rehabilitation Questionnaire (PPRQ) according to Rasch measurement theory.

Method: Five hundred twenty-two post-discharge patients from a neurological rehabilitation unit were included. The PPRQ questionnaire comprises 20 items rated by a cohort of 522 patients about their experiences of participating in rehabilitation. The measurement properties of the PPRQ were evaluated by Rasch analysis of the responses.

Results: The Rasch analysis of 20 items showed some major misfits, particularly three items addressing the involvement of family members. After removing those items, the model fit improved and no significant DIF remained. Despite improvements, person values (−2.96 to 4.86 logits) were not fully matched by the item values (−0.61 to 0.77 logits). Neither did the t test for unidimensionality meet the criterion of 5%, and local dependency was present. The unidimensionality and local dependency could, however, be accommodated for by four testlets.

Conclusion: The PPRQ-17 showed that a ruler with a reasonable and clinical hierarchy can be constructed, although the expectations of dimensionality and local dependency need to be evaluated further. Despite room for further development, PPRQ-17 nevertheless shows improved measurement precision in terms of patient leniency compared with previous evaluations with classical test theory. In turn, this can play a crucial role when comparing different rehabilitation programs and planning tailored care development activities.

KEYWORDS

patient-centred care, psychometrics, self-report, surveys and questionnaires

MELIN ET AL.

WILEY—Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

249

1 | INTRODUCTION

Person-centred care (PCC) and patient participation are crucial in modern health care. Although patient participation can be regarded as crucial, it should be underlined that not all patients prefer the same approach and the same patient may have varying preferences during the care and rehabilitation process.[1] Care and rehabilitation must therefore be tailored to each unique patient, without preconceived notions about the best approach for the patient.[2] In practice, health care professionals can get feedback and be aware of each patient's experiences of care and, in turn, improve their person-centredness accordingly.[3,4] Likewise, experiences rated post-discharge can be used as quality indicators in evaluating and developing care and rehabilitation programs or strategies.[5,6] This raises the corresponding demands on quality assured measurements.[7,8]

From a metrological, quality-assured measurement point of view, this implies invariance across groups, unidimensionality and equal measurement units across the scale continuum. The Danish mathematician Georg Rasch developed a model based on the same underlying principles as physical measurements, i.e. Rasch Measurement Theory, 50 years ago (RMT).[9] With RMT, data are evaluated against measurement criteria. Briefly, RMT allows separate estimates of person and item attribute values and their scaling on the same interval logit scale. In turn, this enables a more accurate measurement; the independence from measures from the validation sample; and more reliable decision making compared to measurements based on classical test theory (CTT).[7,10-12]

Moreover, as stated by Morel & Cano,[8] p.6, of all the measurement properties, 'content validity' is sine qua non, meaning that patient-reported measures should be founded on what patients find most important for their healthcare.[8,13] A recently developed questionnaire for assessment of patients' experience of participation is the Patient Participation in Rehabilitation Questionnaire (PPRQ).[2,5,14] Qualitative interviews with patients revealed five themes reflecting central aspects of patient participation.[2] Thus, the PPRQ was developed on the assumption that patient participation is a multidimensional concept, comprising five subscales: *respect and integrity*, *planning and decision making*, *information and knowledge*, *motivation and encouragement*, and *involvement of family*.[5] Also, the content of PPRQ seems to corroborate a broader patient perspective[15]; the understanding of staff[16]; theories of PCC[17,18]; and other similar questionnaires.[6,19,20] However, recent Rasch validations of two other similar questionnaires, the Patient Preference for Patient Participation tool (The 4Ps)[19] and Person-Centered Care in outpatient care in rheumatology (PCCoc/rheum),[6] suggest a unidimensionality for similar subcomponents as the PPRQ. Likewise, previous studies of the PPRQ using CTT revealed high correlations between the subscales,[5,14] indicating a potential full scale for patient participation. Likewise, previous studies of the PPRQ using CTT revealed high correlations between the subscales[5,14] that indicate a potential full scale for patient participation. This in turn could hypothetically mean that patient participation could be examined as a higher-order construct.[21] Previous research claimed reasonable measurement properties of the PPRQ according to CTT in SCI rehabilitation[5] and neurological rehabilitation.[14] However, CTT has some drawbacks,[1,9]

and it should therefore be beneficial to extend the CTT-based internal validity of the PPRQ by applying RMT. Hence, the purpose of this study was to evaluate the full PPRQ scale according to RMT.

2 | METHODS

2.1 | Data collection

Data were obtained from a previous study by Melin and Årestedt[14] consisting of a target population of patients aged 18 to 80 years with neurological conditions treated between 2006 and 2016 at a rehabilitation unit at the Sahlgrenska University Hospital. In total, 522 respondents were included, corresponding to a response rate of 41% in a post-discharge postal questionnaire survey. Data collection is reported in detail elsewhere,[14] while patient characteristics are summarized in Table 1. As reported in previous research,[14] the respondents were slightly older at the time of injury compared with the nonrespondents but were nevertheless considered close to representative of the target population.

TABLE 1 Sociodemographic and clinical characteristics of the respondents (n = 522)

	n (%)
Gender	
Men	316 (61)
Women	202 (39)
Age group	
≤44	106 (20)
45-64	373 (72)
≥65	41 (8)
Cause of injury	
Stroke	324 (62)
TBI	66 (13)
SCI	27 (5)
Other	102 (20)
Education	
Primary school	91 (17)
Secondary school	216 (42)
University	211 (41)
NPS-question	
Yes, totally	61 (12)
Partly	215 (42)
No	249 (26)

Abbreviations: NPS-question, National Patient Survey question (if the patient had been involved in decisions about his or her care and treatment as much as desired); SCI, spinal cord injury; TBI, traumatic brain injury.

250 | WILEY—Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

MELIN ET AL.

2.2 | Measurement

Patients were asked to report gender, age, education, and cause of injury as well as respond to one question from the Swedish National Patient Survey about the patient's feeling of being involved in decisions about his or her care and treatment (referred to as NPS-question).

The PPRQ is a questionnaire developed for the assessment of the central aspects of patient participation in rehabilitation.[5,14] With the PPRQ, the patients are asked to respond to each of the 20 items on a 5-step Likert scale and to indicate how often they have experienced the care described (ie, experience ratings ranging from 0 = never to 4 = always). For two of the items (E2 and E3), regarding the family involvement, two additional response options are given: *I did not have any family member to involve* and *I did not want to involve any family member*. If one of these options was filled in, the patient was then not supposed to rate items on the Likert scale. Nevertheless, there were patients answering the additional options on items E2 or E3, as well as making the rating within same item and these were considered as missing due to the ambiguity in what they had responded to.

2.3 | Statistical analysis

One important feature of RMT is coupling of item attribute to person characteristic for a certain response.[10] In the PPRQ, patients' experiences are considered to be a measure of the quality of care, including a measure of satisfaction which can be resolved with RMT into separate measures of a person characteristic, θ, patient leniency, and an item attribute, δ, the quality of care. The PPRQ responses were analysed using the software *Rasch Unidimensional Measurement model 2030* (RUMM). The analysis was structured around fundamental aspects of RMT[22,23]:

1. To evaluate the monotonicity of items, the threshold orders were evaluated, ie, the ratings to one item should be consistent with the metric estimate of the underlying construct. Collapsing categories was considered when disordered thresholds occurred.[22]

2. Item model fit was assessed according to fit residuals, chi-square, and item characteristic curve (ICC). The following guidelines were followed: mean fit residuals should be close to zero (0) and have standard deviations (SD) close to 1, the individual item fit residuals should be between −2.50 and +2.50; the chi-square values should not be statistically significant (Bonferroni corrected); the dots of the class intervals should follow the ICC to support good fit.[23] Moreover, chi-square testing of item-trait interaction was done, thereby minimizing the significance of such interactions and the risks of type 1 errors.[24]

3. Together with the fit statistics, examining how the items are distributed along the continuum is crucial for unidimensionality and when deciding whether a measurement ruler successfully could be constructed or not.[23] Considerations if the item distribution is consistent with clinical or theoretical expectations[25] must therefore be taken. Smith method for testing unidimensionality was also

applied,[26] ie, the patterning of residuals is evaluated in a principal component analysis (PCA). The first residual factor obtained is used to define two subsets of items by dividing positively and negatively correlated items. Person estimates for each subset are then compared by using an independent *t* test. To support unidimensionality, the percentage of tests outside the range −1.96 to 1.96 should not exceed 5%.

4. To ensure local independence, residual correlations were evaluated against a relative cut off, ie, residual correlations greater than 0.20 above the average correlations indicate local dependency.[27,28] To deal with local dependency, testlests were created, ie, sets of items were added together into new polytomous items, ie, "super items" with scores ranging from 0 to the maximum of the sum of the scores of the included items.[29] Thereafter, the analysis was repeated.

5. To evaluate targeting, the mean person location was compared with the mean item location (ie, 0 logits) indicating whether the person sample is off centred from the items.[23]

6. To evaluate the internal consistency reliability, the Person separation index (PSI) was used, which is equivalent to the Cronbach α, where zero (0) indicates all error and 1 implies no error. For group assessment, greater than 0.70 is required and greater than 0.85 for individual high-stake evaluations items.[30]

7. Analysis of variance was used to evaluate group differences for gender, age group, education, cause of injury, and response to the NPS-question. Based on the previous work with the PPRQ, significant differences were to be expected for the NPS-question[14] but not for the other comparisons.[31]

8. For differential item functioning (DIF) analysis, both main effects and interaction effects were taken into account, and they should be non-significant. Due to multiple tests, Bonferroni correction was applied. The baseline characteristics (gender, age group, education, and cause of injury) were used for the DIF analysis as clinically significant indicators of invariance in neurological rehabilitation. Due to the limited sample of patients with SCI (n = 27) and TBI (n = 66), data were merged into one subgroup with data from patients with other causes of injuries, ie, comparisons were done between patients with stroke vs others.

3 | RESULTS

3.1 | PPRQ full scale: 20 items

At the first stage of analysis, all 20 items were included and analysed together. Table 2 provides an overview of the statistics. There were several indicators of misfit for a full scale with 20 items: three items showed disordered thresholds: E1 (*Asked to involve family member*); E2 (*Family members invited to planning*); and E3 (*Family members invited to family meetings*). The effects of collapsing categories were evaluated, and disordered thresholds were collapsed into ordered thresholds. In addition, the response categories "sometimes,"

TABLE 2 Summary item statistics of the analyses for versions with 20 and 17 items, respectively[a]

Item		Location	CI	Fit Residuals	χ^2	Probability	Location	CI	Fit Residuals	χ^2	Probability
A1	Respected all context	−0.70	0.14	1.58	5.91	0.55	−0.61	0.15	2.10	10.39	0.17
A6	Treated as a unique individual	−0.10	0.12	**3.34**	17.28	0.02	0.00	0.13	**4.45**	36.57	**0.00**
A7	Took time to listen	−0.53	0.13	−0.43	13.91	0.05	−0.44	0.14	−0.23	8.22	0.31
A9	Sensitive to special wishes	−0.22	0.14	−2.67	26.68	**0.00**	−0.11	0.14	−2.48	20.67	**0.00**
A10	Took seriously	−0.56	0.13	0.71	7.90	0.34	−0.50	0.14	1.47	10.84	0.15
B1	Explained each moment	−0.08	0.12	0.27	11.60	0.11	0.05	0.13	2.22	13.57	0.06
B2	Shared decisions	0.07	0.12	0.27	11.03	0.14	0.23	0.13	1.50	4.25	0.75
B3	Encouraged own responsibility	0.21	0.12	1.36	13.77	0.06	0.37	0.13	**2.78**	11.98	0.10
B5	My expectations	0.54	0.12	−1.65	8.29	0.31	0.77	0.12	0.25	3.71	0.81
B6	My resources and capabilities	0.24	0.12	−2.20	21.75	0.00	0.39	0.13	−1.18	8.88	0.26
C1	Gave understandable information	−0.59	0.13	−0.53	5.04	0.65	−0.52	0.14	0.81	3.38	0.85
C2	Enough information to participate	−0.04	0.13	**−4.10**	29.36	**0.00**	0.10	0.13	**−3.28**	12.76	0.08
C3	Took time to answer my questions	−0.56	0.14	**−3.75**	26.98	**0.00**	−0.47	0.14	**−2.71**	9.74	0.20
C4	Information at "right moment"	0.05	0.13	−2.62	25.33	0.00	0.20	0.14	−2.13	13.52	0.06
D3	Gave hope	−0.11	0.12	0.79	8.49	0.29	0.03	0.13	1.26	3.29	0.86
D4	Enthused	−0.03	0.12	−1.41	8.06	0.33	0.11	0.13	−0.20	5.21	0.63
D5	Helped to set realistic goals	0.23	0.12	−1.78	15.70	0.03	0.40	0.13	0.07	6.03	0.54
E1	Asked to include family member	0.52	0.10	**5.89**	54.98	**0.00**					
E2	Family members invited to planning	0.71	0.10	**6.90**	77.57	**0.00**					
E3	Family members invited to family meetings	0.95	0.09	**11.11**	245.28	**0.00**					

[a]Bolded numbers indicate misfit: Fit residuals should ideally lie between −2.50 and 2.50, and χ^2 should not be significant after Bonnferroni correction (0.0005 for 20 items and 0.000588 for 17 items). Original subscales: A = respect and integrity; B = planning and decision making; C = information and knowledge; D = motivation and encouragement; and E = involvement of family.

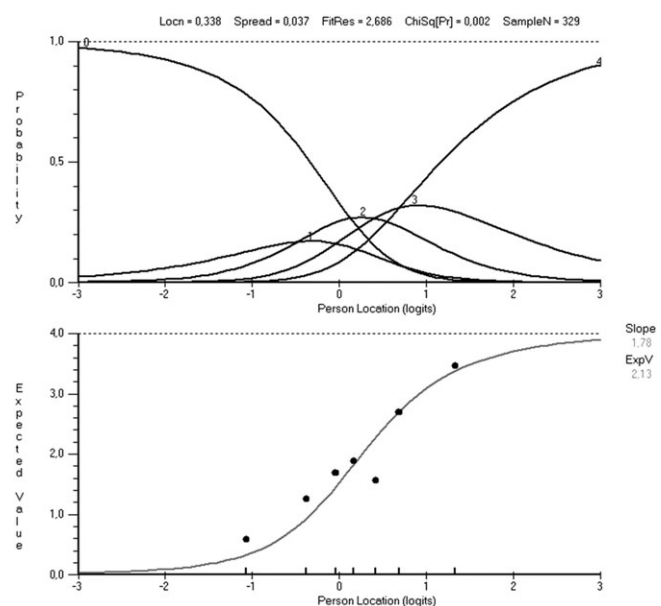Abbreviation: CI, confidence intervals for locations.

FIGURE 1 A, Category probability curves indicating disordered thresholds. B, Item characteristic curves (ICC) showing that the dots deviated from the ICC in item E3 (*family members invited to family meetings*)

"seldom," and "never" were grouped together (Figure 1A) despite being diverse. As shown in Table 2, for eight of the 20 items, the fit residuals were outside of the range −2.50 to +2.50; for three items, the chi-square values were significant; and three items failed to have dots for the inter classes close to the ICC (ICC curve exemplified in Figure 1B). The same items with disordered thresholds, E1, E2, and E3, showed misfit on all the three tests. Several misfit statistics where apparent, and when evaluating t test, 19% were outside the desired range of −1.96 to 1.96.

3.2 | PPRQ full scale: 17 items

There were several indicators of misfit for a full scale with 20 items. Especially, the three items belonging to the subscale *Involvement of family* were problematic. It has also been questioned if involvement of family members should be included in similar measures, which qualitatively is understandable. Hence, an additional analysis was conducted based on 17 items, ie, excluding E1 to E3. Table 2 provides an overview of the statistics, and Figure 2 shows the person-threshold distribution.

By removing items E1 to E3, there were no disordered thresholds, and the fit statistics were improved. The mean item fit residuals were
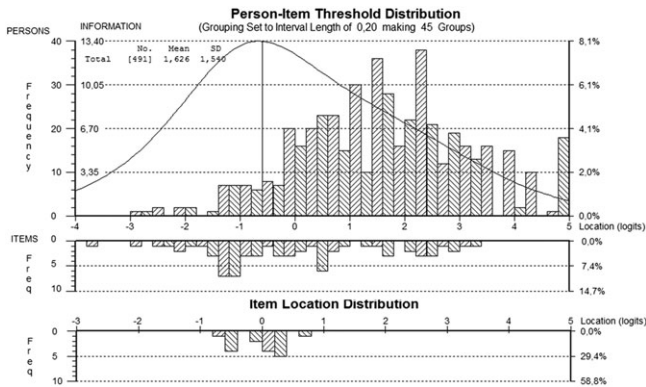
252 | WILEY— Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

MELIN ET AL.

FIGURE 2 Targeting of the PPRQ-17 with person-item thresholds distributions

close to 0 (0.28), although SD was large (2.13). As shown in Table 2, four items did not have item fit residuals within the range of −2.50 to +2.50. Regarding the chi-square values, item A6 (*Treated as a unique individual*) showed significant chi-square values. Likewise, item A6 had dots for the class intervals deviating from the ICC. Hence, item A6 could potentially overdiscriminate the patients' leniency as to whether they have been treated as unique individuals. According to the quality of care δ values, the items originating from the subscales *respect and integrity* and *information and knowledge* were considered as easier to satisfy and the items originating from the subscales *planning and decision making* and *motivation and encouragement* were considered as more difficult to satisfy (Table 2). This could be considered, qualitatively and when comparing to other scales,[6,19] as creating successful hierarchical rulers. Despite some minor misfits and significant item-trait interaction chi-square tests ($P < 0.001$), the cumulated fit statistics showed that the PPRQ-17 acceptably satisfies the RMT (Table 2). The *t* test failed to fulfil the criterion of 5%, since 11% of the percentage of tests were outside the range −1.96 to 1.96. No significant differences ($P > 0.05$) could be identified by studying the person factors (ie, gender, age, cause of injury, or education) of those outside the range compared with those inside the range. In total, nine of 136 residual correlations failed to meet the relative cut-off of (0.14). Item D3 *Gave hope* and item D4 *Enthused* showed the highest residual correlation (0.33).

Figure 2 shows the item and person thresholds; quality of care δ values (−0.61 to 0.77 logits) were covered by the patients' leniency θ values (−2.96 to 4.86 logits) but not the opposite. In total, 66 patients did not have fit residuals within ±2.50. The PSI was 0.93, ie, the scale's ability to discriminate correctly between person ability was well above the criterion of 0.85 to be used for individual, high-stake evaluation. As expected, there were significant differences regarding patients' ratings on the NPS-question ($P < 0.001$), ie, those who scored "no" (mean location 0.03, SD 1.75), respectively, "partly" (mean location 1.27, SD 1.29) showed lower leniency than those who scored "yes" (mean location 2.82, SD 1.60). Also, men had statistically significant higher leniency compared with women (mean location 2.01, SD 1.72 vs 1.570, SD 1.88). There were neither statistically significant DIF main effects nor interaction effects for any of the person variables for any of the items ($P > 0.00098$ after Bonferroni correction).

3.3 | PPRQ full scale: 17 items, four testlets

There were some indications of local dependency as well as of multidimensionality, which in turn could violate the reliability and internal validity of the 17-item PPRQ. By examining the residual correlations matrix, clusters of the subscales were identified. Consequently, items from the four subscales were grouped to form four testlets: respect and integrity (five items: A1, A6, A7, A9, and A10), planning and decision making (five items, B1, B2, B3, B5, and B6), information and knowledge (four items: C1, C2, C3, and C4), and motivation and encouragement (three items D3, D4, and D5). Repeating the analysis significantly improved dimensionality (from 11% to 4% outside −1.96 to 1.96), and item-trait interaction chi-square significance tests (from $P < 0.001$ to $P = 0.04$) and, as expected, reduced reliability (PSI from 0.93 to 0.86).

4 | DISCUSSION

Based on the additional insight gained by this RMT analysis of the PPRQ compared with former CTT-based studies, initial evidence that a higher-ordered patient participation score provided by the PPRQ questionnaire with reasonable fit to RMT has emerged. Previous recommendations to not calculate a full score for the PPRQ should therefore be revised. For PPRQ-17, the analyses showed questionable local dependency and *t* test statistics for unidimensionality, although this should not be considered as an absolute property but rather a relative one.[24] By creating four testlets stemming from the subscales, it was found that local dependency and multidimensionality could be accommodated for, and the reliability could be kept above the recommended value of 0.85 for individual high-stake evaluation items.[30] It could be dangerous to have a too hard-line data-driven approach as it heavily relies upon the quality of the data.[8] To provide evidence for the validity of a higher-order patient participation scale, it is recommended to pay attention to whether the item ordering is consistent with expectations.[21] For PPRQ-17, there are striking similarities in the logical order of similar items in the PPRQ-17 to the other questionnaires recently developed.[6,19] Similar to those two questionnaires, PPRQ-17 has items that "tell a story," from the easiest tasks of quality of care (eg, respecting the patient and providing information) to the more demanding tasks of quality of care (eg, involving the patient in decisions and goal-planning). However, this cascade of care tasks requires further evaluation across different contexts and samples.

By exploiting the RMT sound metrological underpinnings, ie, references for traceability and declarations of measurement uncertainties[11,23] are given to the PPRQ as a clinical tool for evaluation of patient participation. The benefit of using Rasch-transformed patient leniency θ values instead of raw sum scores according to CTT is shown clearly in Figure 3. Distortion of PPRQ-17 is evident towards both ends of the scale, meaning that patients with higher leniency are underestimated with CTT while patients with lower leniency are overestimated with CTT. Another benefit demonstrated in this paper is that no DIF could be found, ie, the assumption on invariance was confirmed. This implies that reported differences in leniency θ values are
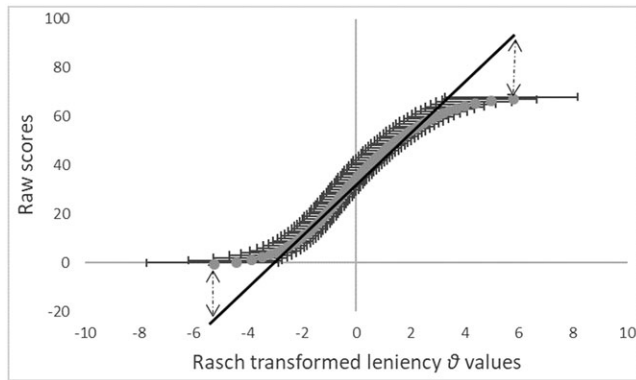
FIGURE 3 Distortion of the PPRQ-17 scale when raw scores on the y-axis are compared with Rasch-transformed patient leniency θ values on the x-axis. The black full line indicates a perfect correlation between mean scores and leniency θ values and the dashed arrows at the ends indicate underestimation and overestimation, respectively. Error bars indicates confidence intervals

not reflecting differences in the functioning of the PPRQ as the items work in the same way for the different sample groups to be compared.[32] In turn, this can play a crucial role when comparing different rehabilitation programmes and planning tailored care development activities.

As hypothesized, significant differences were shown between groups regarding the patients' responses to the NPS-question (their feelings about to what extent they are allowed to be involved in decisions about their care and treatment). This supports the criterion-based validity of the PPRQ. Previous evaluations with the PPRQ in SCI rehabilitation have not shown any significant differences between gender, age groups, or education level.[31] Nevertheless, in contrast to what was hypothesized previously, significant differences concerning leniency by men and women were demonstrated in the present study. Unfortunately, this study has not provided any information about why men and women have different experiences. The previous study with PPRQ,[31] likewise other studies assessing importance aspects in care and rehabilitation,[33-35] has shown that women assign higher importance than men to these issues. Consequently, this could affect respondents' leniency as they put higher expectations on their care delivery. On the other hand, it could be related to the fact that different groups have been unequally treated, and this warrants a more careful evaluation in forthcoming studies.

One major benefit of using RMT is that the outcomes of the analysis provide an understanding of the limitations of the current measurement as well as how to solve them.[36] Hence, some areas need to be improved in the PPRQ, especially involvement of family members as well as scale to sample targeting. Items concerning involvement of family showed disordered thresholds, ie, the probability of choosing one response category was equal to the probability of choosing the adjacent category. One explanation may be that some patients felt that those items were not applicable for their rehabilitation as they did not want to include a family member or did not have a family member to include.[2] Another reason, as argued by Bala et al,[6] is that involvement of family members may not be productive in the measurement of person-centredness. Hence, it is reasonable not to include those items in the measurement

of patient participation but, on the other hand, family involvement could still be a crucial part of the conceptual framework.[6,14] Therefore, rephrasing those response categories into fewer categories may solve the problem. Moreover, all three scales (PPRQ-17, 4Ps and PCCoc/rheum) have shown the same problems regarding targeting: Items are covered by the patients, but the patients are not fully covered by the items. This is especially true for those patients having the highest leniency for patient participation and implies that those patients are measured with a lower precision.[23] Similarly, at the other end of the continuum, where fewer patients were located, the items were compromised.[23] Additional data that include patients with a wider range of experiences could solve this.[6] Nevertheless, it might be difficult to find clinics not fulfilling the least demanding items (ie, tasks of quality of care such as respecting the patient and providing information). Another solution may be to include additional items placing higher demands on the quality of care, ie, more demanding tasks will allow to better differentiate between the patients with the highest leniency. Furthermore, one should also consider whether there is a "problem" to have patients with too high leniency. It could be a relatively minor problem if the purpose is to use the measure as a means of quality assurance and for identification of areas of clinical improvements.[6] Then again, lower measurement precision in terms of larger measurement uncertainty is without doubt giving important evidence about limitations in the reliability of decisions of conformity assessment about whether care fulfils requirements or not.[37] Therefore, the best option is preferably to include more demanding quality of care items, as in all metrology, wider item span allows better calibration of the psychometric ruler.

In addition to the mentioned limitations with unidimensionality and targeting, there are some other methodological considerations to bear in mind when interpreting the findings of this study. Firstly, the item-trait interaction chi-square test revealed significant effects and some misfit statistics were shown for some items in the PPRQ-17, but these items were not excluded when an all-embracing picture of the whole analysis was considered. This implies that there is a potential risk of overestimation for some items.[23] Nevertheless, as stated by McClimans et al,[38(p5)] "In order for theory driven measurement to proceed there should be as much attention paid to disorder as there is to order," which should be further evaluated in forthcoming studies. Secondly, local dependency was present for some items. However, this should be interpreted with caution as there are only 17 items and evaluations of local dependency seem to be less reliable when there are less items[27] than 20. Further studies should therefore assess the efficiency of the potential item redundancy and implications for the item and person estimates. Thirdly, about 10% of patients had extreme values. When these were excluded, the internal construct validity improved, while at the same time, concerns were raised about the external construct validity.[22] Lastly, it has been suggested that allowing patients to rate their own experiences during in-patient care might give a more accurate recall.[3] On the other hand, allowing some time for reflection and adaptation after a long and extensive rehabilitation to pass may be needed to reliably summaries one's experiences. Thus, the impact of recall bias will have to be evaluated in forthcoming studies, particularly in terms of measurement uncertainties.

254 | WILEY—Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

MELIN ET AL.

In conclusion, this study gives initial evidence that the PPRQ-17 shows a reasonably good fit to the RMT. Unidimensionality and local dependency could be resolved by testlets, but further evaluation is needed. Likewise, further exploration and development are needed to understand the construct and more demanding items for a better calibration of the psychometric ruler and in turn improvement of the targeting and reduction of measurement uncertainties. Anyhow, PPRQ-17 shows improved measurement precision in terms of patient leniency compared with previous evaluations with CTT and the PPRQ-17 leniency θ values is recommended.

DECLARATION OF INTEREST

The authors report no declarations of interest.

ORCID

Jeanette Melin https://orcid.org/0000-0002-3700-3921
Leslie Pendrill https://orcid.org/0000-0003-4349-500X

REFERENCES

1. Robinson A, Thomson R. Variability in patient preferences for participating in medical decision making: implication for the use of decision support tools. *Qual Health Care*. 2001;10(Suppl 1):i34-i38.

2. Lindberg J, Kreuter M, Taft C, Person LO. Patient participation in care and rehabilitation from the perspective of patients with spinal cord injury. *Spinal Cord*. 2013;51(11):834-837.

3. Gill S, Redden-Hoare J, Dunning T, Hughes A, Dolley P. Health service should collect feedback from the inpatients at the point of service: opinions from patients and staff in acute and subacute faciliteis. *International J Qual Health Care*. 2015;27(6):507-512.

4. Weath H, Horrell J, Valderas JM, Close J, Fosh B, Lloyd H. Can practitioners use patient reported measures to enhance person centred coordinated care in practice? A qualitative study. *Health Qual Life Outcomes*. 2018;16:223.

5. Lindberg J, Kreuter M, Person LO, Taft C. Patient participation in rehabilitation questionnaire (PPRQ)—development and psychometric evaluation. *Spinal Cord*. 2013;51(11):838-842.

6. Bala S-V, Forslind K, Fridlund B, Hagell P. Measuring person-centred care in nurse-led outpatient rheumatology clinics. *Musculoskeletal Care*. 2018;16(2):296-304. https://doi.org/10.1002/msc.1234

7. Pendrill LR. Assuring measurement quality in person-centred healthcare. *Meas Sci Technol*. 2018;29(3):034003.

8. Morel T, Cano SJ. Measuring what matters to rare disease patients—reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. *Orphanet J Rare Dis*. 2017;12(1):171.

9. Rasch G. Probabilistic Theorys for some intelligence and attainment tests. Copenhagen; 1960.

10. Pendrill L. Man as a measurement instrument. *NCSLI Meas*. 2014; 9(4):24-35.

11. Andrich D. Controversy and the Rasch theory: a characteristic of incompatible paradigms? *Med Care*. 2004;42(1 Suppl):I7-I16.

12. Cano SJ, Hobart JC. The problem with health measurement. *Patient Prefer Adherence*. 2011;5:279-290.

13. Øvretveit J, Zubkoff L, Nelson E, Frampton S, Lehman Knudsen J, Zimlichman E. Using patient-reported outcome measures to improve patient care. *International J Qual Health Care*. 2017;29(6):874-879.

14. Melin J, Årestedt K. The patient participation in rehabilitation questionnaire (PPRQ): psychometric evaluation of the revised version for use in neurological rehabilitation. *Disabil Rehabil*. 2019;1-8. https://doi.org/10.1080/09638288.2018.1528303

15. The Swedish Agency for Health and Care Services Analysis. Från mottagare till medskapare. Report 2018:8.

16. Melin J, Persson LO, Taft C, Kreuter M. Patient participation from the perspective of staff members working in spinal cord injury rehabilitation. *Spinal Cord*. 2018;56(6):614-620.

17. Leplege A, Gzil F, Cammelli M, Lefeve C, Pachoud B, Ville I. Person-centredness: conceptual and historical perspectives. *Disabil Rehabil*. 2007;29(20-21):1555-1565.

18. Ekman I, Swedberg K, Taft C, et al. Person-centered care—ready for prime time. *Eur J Cardiovasc Nurs*. 2011;10(4):248-251.

19. Luhr K, Eldh AC, Nilsson U, Holmefur M. Patient preferences for patient participation: psychometric evaluation of the 4Ps tool in patients with chronic heart or lung disorders. *Nord J Nurs Res*. 2017;38:68-76.

20. Cott CA, Teare G, McGilton KS, Lineker S. Reliability and construct validity of the client-centred rehabilitation questionnaire. *Disabil Rehabil*. 2006;28(22):1387-1397.

21. Andrich D. Implications and applications of modern test theory in the context of outcomes based education. *Stud Educ Eval*. 2002;28(2): 103-121.

22. Tennant A, Conaghan PG. The Rasch measurement theory in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum*. 2007;57(8): 1358-1362.

23. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* (Winchester, England). 2009;13:iii, ix-x:1-177.

24. Hagell P, Westergren A. Sample size and statistical conclusions from tests of fit to the Rasch model according to the Rasch unidimensional measurement model (Rumm) program in health outcome measurement. *J Appl Meas*. 2016;17(4):416-431.

25. Hagell P. Testing rating scale unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch theory: the primacy of theory over statistics. *Open J Stat*. 2014;4(06):456-465.

26. Smith RM. A comparison of methods for determining dimensionality in Rasch measurement. *Struct Equ Theorying*. 1996;3(1):25-40.

27. Marais I. Local dependence. In: Christensen KB, Krainer S, Mesbah M, eds. *Rasch Theorys in Health*. London, UK: Hoboekn NJ: ISTE ltd and John Wiley & Sons. Inc; 2013.

28. Christensen KB, Makransky G, Horton M. Critical values for Yen's Q3: identification of local dependence in the Rasch theory using residual correlations. *Appl Psychol Measur*. 2017;41(3):178-194.

29. Andrich D, Marinas I. Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *J Appl Meas*. 2008;9:200-215.

30. Nunnally JC. In: Bernstein IH, ed. *Psychometric Theory*. 3rd ed. New York: New York: McGraw-Hill; 1994.

MELIN ET AL.

WILEY—Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

255

31. Lindberg J, Kreuter M, Persson L, Taft C. Family members' perspectives on patient participation in spinal cord injury rehabilitation. *Int J Phys Med Rehabil*. 2014;2:223.

32. Hagquist C, Andrich D. Recent advances in analyses if differential item functioning in health research using the Rasch model. *Health Qual Life Outcomes*. 2017;15:181.

33. Say R, Murtagh M, Thomson R. Patients' preference for involvement in medical decision making: a narrative review. *Patient Educ Couns*. 2006;60(2):102-114.

34. Hamann J, Neuner B, Kasper J, et al. Participation preferences of patients with acute and chronic conditions. *Health Expect*. 2007;10(4): 358-363.

35. Florin J, Ehrenberg A, Ehnfors M. Clinical decision-making: predictors of patient participation in nursing care. *J Clin Nurs*. 2008;17(21): 2935-2944.

36. Hobart J, Cano S, Posner H, et al. Putting the Alzheimer's cognitive test to the test II: Rasch measurement theory. *Alzheimers Dement*. 2013;9(1): S10-S20.

37. Pendrill LR. Using measurement uncertainty in decision-making and conformity assessment. *Metrologia*. 2014;51(4):S206-S218.

38. McClimans L, Browne J, Cano S. Clinical outcome measurement: Models, theory, psychometrics and practice. *Stud Hist Philos Sci*. 2017;65-66:67-73.