VALUE IN HEALTH ∎ (2014) ∎∎∎–∎∎∎

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Construct Validity of the Swedish Version of the Revised Piper Fatigue Scale in an Oncology Sample—A Rasch Analysis

Åsa Lundgren-Nilsson, PhD[1,2], Anna Dencker, PhD[1,], Sofie Jakobsson, PhD[1], Charles Taft, PhD[1], Alan Tennant, PhD[1,3]*

[1]*Centre for Person-centred Care, Institute of Health and Care Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden;* [2]*Department of Clinical Neuroscience and Rehabilitation, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden;* [3]*Faculty of Medicine & Health, Academic Department of Rehabilitation Medicine, University of Leeds, Leeds, UK*

ABSTRACT

Objectives: Fatigue is a common and distressing symptom in cancer patients due to both the disease and its treatments. The concept of fatigue is multidimensional and includes both physical and mental components. The 22-item Revised Piper Fatigue Scale (RPFS) is a multidimensional instrument developed to assess cancer-related fatigue. This study reports on the construct validity of the Swedish version of the RPFS from the perspective of Rasch measurement. **Methods:** The Swedish version of the RPFS was answered by 196 cancer patients fatigued after 4 to 5 weeks of curative radiation therapy. Data from the scale were fitted to the Rasch measurement model. This involved testing a series of assumptions, including the stochastic ordering of items, local response dependency, and unidimensionality. A series of fit statistics were computed, differential item functioning (DIF) was tested, and local response dependency was accommodated through testlets. **Results:** The Behavioral, Affective and Sensory domains all satisfied the Rasch model expectations. No DIF was observed, and all domains were found to be unidimensional. The Mood/Cognitive scale failed to fit the model, and substantial multidimensionality was found. Splitting the scale between Mood and Cognitive items resolved fit to the Rasch model, and new domains were unidimensional without DIF. **Conclusions:** The current Rasch analyses add to the evidence of measurement properties of the scale and show that the RPFS has good psychometric properties and works well to measure fatigue. The original four-factor structure, however, was not supported.

Keywords: cancer-related fatigue, Rasch analysis, Revised Piper Fatigue Scale.

Introduction

Fatigue is common and can be very distressing in cancer patients due to both the disease and its treatments [1]. The concept of fatigue is multidimensional and includes both physical and mental components and has impact on physical energy levels as well as the patient's social life [2,3]. The Revised Piper Fatigue Scale (RPFS) is a multidimensional instrument developed to assess cancer-related fatigue and is one of the most widely used instruments internationally [4]. The RPFS was developed within a theoretical framework known as the Integrated Fatigue Model. The theory encompasses factors, of subjective and objective character, that are assumed to affect the development and manifestations of fatigue [5].

The RPFS has coverage of fatigue domains in 22 items divided into four subscales: Behavioral/Severity, Affective meaning, Sensory, and Cognitive/Mood [6,7]. It also contains open-ended questions to assess patients' beliefs about what contributes to their fatigue and what they do to alleviate their fatigue. Subscales and total score range from 0 to 10 in the original version, with higher values indicating more intense fatigue. To date, the RPFS has been translated into Chinese [8] and Greek [9] and validated for use in France [10], The Netherlands [11], Brazil [12], and Italy [13,14].

Swedish Version of Piper

In 2007, the RPFS was translated and culturally adapted for use in Sweden [15]; however, no psychometric evaluation was performed at that time. Before initiating later psychometric assessments, the Swedish version was reevaluated in cognitive interviews with 29 cancer patients. Based on comments and suggestions made by patients in the cognitive interviews, some changes were made to earlier item translations, time frame ("now" to "during the past week"), and response scale (0–10 to 1–10). The psychometric evaluation included content and concurrent validity as well as internal consistency including exploratory factor analysis and multitrait scaling analysis. The factor analysis failed to support a four-dimensional model of fatigue as conceptualized in the original RPFS, but rather support was found within a three-factor solution [16]. Validation studies of other language versions of the RPFS have reported similar results [13].

* *Address correspondence to:* Anna Dencker, Centre for Person-centred Care, Institute of Health and Care Sciences, Sahlgrenska Academy, University of Gothenburg, Sweden.

E-mail: anna.dencker@gu.se.

http://dx.doi.org/10.1016/j.jval.2014.02.010

Table 1 – Demographic characteristics of the study group (n = 196).

Characteristic	n (%)
Sex	
Women	133 (68)
Men	63 (32)
Age (y)	
<60	79 (40)
61–70	74 (38)
>70	43 (22)
Site of radiotherapy	
Breast	109 (56)
Pelvic	56 (28)
Thorax	9 (5)
Head and neck	22 (11)

Furthermore, to justify the use of interval-level scores, further assessments of the construct validity of the RPFS are needed with appropriate techniques. The aim of this study was to examine the internal construct validity of the Swedish version of the RPFS using Rasch analysis.

Methods

Setting and Patients

Data for this analysis were collected from patients undergoing curative radiation therapy, with or without concomitant chemotherapy treatment, within two oncology outpatients' settings. Patients were eligible when they were planned to undergo radiotherapy against breast, pelvic, thorax, or head and neck. The patients answered on their level of fatigue after 4 to 5 weeks of radiotherapy. Within a 3-month period, 300 patients were informed on the study and were sent the questionnaire by mail. Evaluable questionnaires were completed and returned from 65% of the patients (n = 196), as described elsewhere [16]. The Affective meaning subscale had the highest ratings of missed items (n = 14). Characteristics of study participants are shown in Table 1.

Rasch Analysis

Data from the scale were fitted to the Rasch measurement model [17]. This involved testing a series of assumptions, including the stochastic ordering of items, local response dependency, and unidimensionality [18]. Stochastic ordering is evaluated through fit to the model, which reflects a probabilistic Guttman ordering [19]. A series of fit statistics are used to indicate adequacy of fit, and their ideal values are shown at the bottom of the summary fit table (Table 2).

The overall summary statistics (item trait interaction), with standardized mean person and item fit, was evaluated by using χ^2 statistics with nonsignificant χ^2 probability values. A significant χ^2 value indicates that the hierarchical ordering of the items varies across the trait being measured (i.e., fatigue), which comprise the required property of invariance. The standardized mean values of the person and item fit residual by a mean ± SD score of 0.0 ± 1.0 indicates a good fit. Values outside this range indicate problems and render further examination of the individual fit of persons and items residuals. A nonsignificant χ^2 probability value and standardized fit residuals of between −2.5 and +2.5 (99% confidence interval) indicate adequate fit of individual persons and items residuals [20]. A good fit to the Rasch model would expect that for each item of the Piper scale, persons who are severely affected by fatigue would rate higher scores whereas persons who are less affected would rate lower scores. To examine the category function of each item, the threshold ordering was analyzed. A threshold is the point between two categories in which either response is equally probable. When disordered thresholds occur, the items can be rescored by collapsing the categories [21].

The process of Rasch analysis also allows for an investigation of differential item functioning (DIF) [22]; the response to an item (dichotomous or polytomous), given the same level of the trait, should not differ across group membership such as diagnosis. The presence of DIF can be adjusted by "splitting" items in the latter case such that, for example, when there are two diagnostic groups, an item becomes two items, one for each diagnosis, with structural missing values for the excluded diagnosis. A reliability index (Person Separation Index) is also reported.

Local response dependency is where items are linked in some way, for example, two items asking about the distance walked,

Table 2 – Fit of the Revised Piper Fatigue Scale to the Rasch model.

Analysis no.	Name	Item residual*	Person residual*	χ^2		Unidimensionality		
				Value	P	PSI	test %	LCI
1	Behavioral	0.19 ± 2.24	−0.65 ± 1.36	10.20	0.60	0.94	2.6	0.0
2	Affective	−0.09 ± 1.52	−0.59 ± 1.23	12.19	0.07	0.91	7.8	4.4
3	Sensory	0.30 ± 1.89	−0.68 ± 1.38	15.13	0.13	0.93	7.6	4.2
4	Mood/Cognition	0.45 ± 1.23	−0.46 ± 1.31	12.16	0.43	0.90	12.1	9.5
5	Cognition	0.36 ± 0.76	−0.61 ± 1.07	8.36	0.21	0.88	6.1	2.7
6	Mood	0.14 ± 0.98	−0.70 ± 1.20	4.70	0.58	0.86	4.3	1.0
7	Initial 22 items	0.61 ± 2.60	−0.43 ± 2.11	127.10	0.00	0.96	25.9	22.6
8	Rescored	0.48 ± 2.37	−0.48 ± 2.09	121.75	0.00	0.96	25.9	22.6
9	Five testlets	0.37 ± 1.40	−0.46 ± 1.18	8.36	0.59	0.91	6.2	2.8
10	12-item form, four testlet solution	0.26 ± 1.55	−0.44 ± 1.08	7.95	0.44	0.91	8.7	5.3
	Ideal values	0.0 ± <1.4†	0.0 ± <1.4		>0.05‡	>0.85		(LCI <5%)

LCI, lower confidence interval; PSI, Person Separation Index.

* Mean ± SD.

† May be higher when unequal length testlets present.

‡ Bonferroni adjusted.

where one item asks about difficulty walking 100 m and the second about walking 500 m. If a respondent has no difficulty in walking 500 m, then he or she must also have no difficulty in walking 100 m. This breaches the local independence assumption that says that, depending on the trait being measured, responses to items must be independent; otherwise, the presence of local dependency inflates reliability and compromises parameter estimation [23]. Local response dependency can be accommodated through testlets where the items are simply summed together into a "super item" or testlet (in the walking example, this would form the equivalent of one question asking how far the respondent could walk without difficulty) [24]. When all items are grouped into testlets, this is formally equivalent to a bifactor solution [25]. The latent correlation between testlets can also be determined, as well as the proportion of nonerror variance associated with the total score when the testlets are added together to make that score [26].

The latent estimate (in this case of fatigue) so derived from the analysis will be that associated with this common variance, having discarded the unique variance associated with multidimensionality. Consequently, any transformation table produced to show the raw score to interval scale estimate will be based on this common variance.

The local dependency assumption can also be violated by multidimensionality. A basic assumption of summating any set of items to make a total score is that the set is unidimensional [27]. In RUMM2030, the software used in the current study, Smith's test of unidimensionality, is implemented whereby items loading positively and negatively on the first principal component of the residuals are used to make independent person estimates (in this case, fatigue), and these are contrasted through a series of independent *t* tests [28].

Less than 5% of such tests should be significant to support unidimensionality. A binomial confidence interval of proportions can be used to show that the lower confidence interval of the observed proportion falls below the 5% level.

RUMM 2030 was used for the Rasch analysis [29]. All other analyses were undertaken with SPSS20.

The study was approved by The Regional Ethical Review Board in Gothenburg (diary no: 618-10) and conducted in compliance with the Helsinki declaration.

Results

Rasch Analysis

Each domain was initially analyzed separately. The Behavioral, Affective, and Sensory domains all satisfied the Rasch model expectations (Table 2, analyses 1–3). No DIF was observed, and all domains were found to be unidimensional. The Mood/Cognitive scale, however, failed to fit the model (analysis 4), and substantial multidimensionality was found. Splitting the scale between Mood and Cognitive items resolved fit to the Rasch model (analyses 5 and 6), and new domains were unidimensional without DIF.

Following this, the full scale was fitted to the model. Initial fit was poor, with a significant misfit to the model (analysis 7). Two items displayed disordered thresholds (item 5 "Engage in sexual activity" and item 22 "Unable to remember"). Consequently, these items were rescored, item 5 into four categories and item 22 into six categories (item 5: 0,1,1,1,1,2,2,2,2,3; item 22: 0,1,1,1,2,2,2,3,4,5). This solution gave ordered thresholds in all items; however, fit to the model was still poor (analysis 8). Analysis of correlations in residuals gave an indication that the items clustered within the five domains: Behavioral, Affective, Sensory, Mood, and Cognition. Grouping of the items into five testlets gave fit to the model and, with a lower confidence interval for the number of significant *t* tests overlapping 5%, supports unidimensionality (analysis 9). There was no DIF by age, sex, or diagnosis. Applying the testlet-based bifactor solution within the Rasch framework showed that the average latent correlation, between five sets of items (testlets), and adjusted for error, was 0.82. In summating a total score based on the five testlets (subscales—the original four but with cognition and mood separate), the proportion of common true score variance relative to the total nonerror variance was 0.94. Consequently, the total score can be considered unidimensional and the profiles of respondents can be adequately summarized by their score when it is transformed to the testlet-based metric.

Finally, a newly reported 12-item short form was tested against the Rasch model [7] but failed to satisfy the unidimensionality assumption, even after applying the bifactor testlet solution (analysis 10). Once again the cognitive items retained in the short form appear to be the problem.

Conclusions

The internal construct (factorial) validity of the Swedish adaptation of the RPFS was assessed using the Rasch model. Individual subscale scores for the Behavioral, Affective, and Sensory domains appear robust on the whole, although the Affective scale appears to be affected by local dependency. A previously performed factor analysis of the Swedish version [16] also indicated a possible division, reflected by the magnitudes of the factor loadings of the first two items compared with those of the latter three, even though the subscale held together. A single domain comprising Cognitive/Mood failed to be supported, corroborating earlier factor analytical studies both for the Swedish version [16] and for other international versions [10,13,14]. A bifactor solution with each of the three original domains treated as testlets, and with Cognitive/Mood treated as separate testlets (so making five in all), does support a dominant common factor, and with most of the nonerror variance being common, the summated raw score is, in practice, a sufficient statistic for estimating fatigue.

The lack of fit of the cognitive items suggests that multidimensionality may be inherent in fatigue, with respect to physical and cognitive/mental components of the construct. This has been found, to a certain extent, in recent Rasch-built scales in other conditions, although some communality has also been observed [30]. Although such multidimensionality may cause problems with scale viability, it may not be such a problem in practice where it may be more important to model the dual effects of physical and mental fatigue on some outcome, such as physical functioning, or job retention [31]. Thus, having both physical and mental components of fatigue measured separately could be important. Indeed, the Piper scale offers several domains as independent viable constructs, and this may offer the opportunity to investigate how these may differentially contribute to some defined outcome.

This study failed to confirm a 12-item short form of the RPFS that was suggested after a confirmatory factor analysis and item reduction in a group of cancer survivors [7]. In the current Rasch analysis, the 12-item solution was not fit to the Rasch model. The benefit of the Rasch model is that scale scores can be reliably compared between persons and groups regardless of whether they experience high or low levels of the construct measured (i.e., fatigue). It is important to establish the measurement properties of the scale to be able to reliably summarize the scale score [32]. It may be a worthwhile objective to reduce the number of items included in a scale, but according to the Rasch model, the 22-item RPFS seems to function better.

One interesting aspect of the Rasch analysis in the current study was just how much the numeric rating scale format of the response options remain fully ordered, with just occasional marginal disordering of the thresholds. Typically, scales with many categories often show disordering of the thresholds such that, for example, when a high score represents high levels of fatigue, the transition between categories 2 and 3 represents a higher level of fatigue than does the transition between categories 3 and 4, which should not be the case [33]. This, however, does not seem to be the case with the numeric rating scale in the current study and suggests that it may not be the number of categories, per se, that increases the risk of disordering, but rather instead to have to make the choice between the semantic descriptions of the categories.

Limitations of the current study could be that the RPFS may not fit the Rasch model when applied in other groups of patients. This population was a sample of patients, aged 35 to 87 years, with different cancer diagnoses who all received radiotherapy and the majority were women. Consequently, the analyses need to be confirmed with data from other groups of patients.

The strength of this study is that Rasch analyses seek to test data against the most rigorous standards for constructing invariant measurement. Failure to satisfy these requirements can contribute to a better understanding of how the scale works and, as in this study, the improvement in the Swedish version of the RPFS, through detailing the construct validity (or absence of) of the scale. The testing for invariance by group membership, such as age and sex, is also a key aspect of evaluation, such that the absence of DIF gives confidence to the comparison of groups. The Rasch method also enables transformation of ordinal data to interval scaling, which makes it possible to summarize the scale scores into a sum score and transform this into an interval scale latent estimate that can be used to calculate means and SDs to compare patients and groups [32].

Finally, the RPFS is an increasingly used instrument to assess cancer-related fatigue and despite demonstrating good internal consistency, several factor analyses have shown problems in the predefined four-factor structure but rather a three- or a five-factor structure [10,12,13,16]. The current Rasch analyses add to the evidence of measurement properties of the scale and show that the RPFS, which is grounded in theory, has good psychometric properties and works well to measure the domain of fatigue.

Source of financial support: The authors have no other financial relationships to disclose.

REFERENCES

[1] Berger AM, Gerber LH, Mayer DK. Cancer-related fatigue: implications for breast cancer survivors. Cancer 2012;118(8, Suppl):2261–9.
[2] de Raaf PJ, Sleijfer S, Lamers CH, et al. Inflammation and fatigue dimensions in advanced cancer patients and cancer survivors: an explorative study. Cancer 2012;118:6005–11.
[3] Paiva CE, Paiva BS. Prevalence, predictors, and prognostic impact of fatigue among Brazilian outpatients with advanced cancers. Support Care Cancer 2013;21:1053–60.
[4] Seyidova-Khoshknabi D, Davis MP, Walsh D. Review article: a systematic review of cancer-related fatigue measurement questionnaires. Am J Hosp Palliat Care 2011;28:119–29.
[5] Piper BF, Lindsey AM, Dodd MJ. Fatigue mechanisms in cancer patients: developing nursing theory. Oncol Nurs Forum 1987;14:17–23.
[6] Piper BF, Dibble SL, Dodd MJ, et al. The revised Piper Fatigue Scale: psychometric evaluation in women with breast cancer. Oncol Nurs Forum 1998;25:677–84.
[7] Reeve BB, Stover AM, Alfano CM, et al. The Piper Fatigue Scale-12 (PFS-12): psychometric findings and item reduction in a cohort of breast cancer survivors. Breast Cancer Res Treat 2012;136:9–20.
[8] So WK, Dodgson J, Tai JW. Fatigue and quality of life among Chinese patients with hematologic malignancy after bone marrow transplantation. Cancer Nurs 2003;26:211–9; quiz 20–1.
[9] Lavdaniti M, Patiraki E, Dafni U, et al. Prospective assessment of fatigue and health status in Greek patients with breast cancer undergoing adjuvant radiotherapy. Oncol Nurs Forum 2006;33:603–10.
[10] Gledhill JA, Rodary C, Mahe C, Laizet C. French validation of the revised Piper Fatigue Scale [in French]. Rech Soins Infirm 2002;68:50–65.
[11] Dagnelie PC, Pijls-Johannesma MC, Pijpe A, et al. Psychometric properties of the revised Piper Fatigue Scale in Dutch cancer patients were satisfactory. J Clin Epidemiol 2006;59:642–9.
[12] Mota DD, Pimenta CA, Piper BF. Fatigue in Brazilian cancer patients, caregivers, and nursing students: a psychometric validation study of the Piper Fatigue Scale-Revised. Supportive Care Cancer 2009;17:645–52.
[13] Giacalone A, Polesel J, De Paoli A, et al. Assessing cancer-related fatigue: the psychometric properties of the Revised Piper Fatigue Scale in Italian cancer inpatients. Support Care Cancer 2010;18:1191–7.
[14] Annunziata MA, Muzzatti B, Mella S, et al. The revised piper fatigue scale (PFS-R) for Italian cancer patients: a validation study. Tumori 2010;96:276–81.
[15] Ostlund U, Gustavsson P, Furst CJ. Translation and cultural adaptation of the Piper Fatigue Scale for use in Sweden. Eur J Oncol Nurs 2007;11:133–40.
[16] Jakobsson S, Taft C, Ostlund U, Ahlberg K. Performance of the Swedish version of the Revised Piper Fatigue Scale. Eur J Oncol Nurs 2013;17:808–13.
[17] Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago, IL: University of Chicago Press, 1960.
[18] Gustafsson JE. Testing and obtaining fit of data to the Rasch model. Br J Math Stat Psy 1980;33:205–33.
[19] Guttman LA. The basis for Scalogram analysis. In: Stouffer SA, Guttman LA, Suchman FA, et al., eds. Studies in Social Psychology in World War II: Vol 4. Measurement and Prediction. Princeton: Princeton University Press, 1950.
[20] Andrich D. Rasch Models for Measurement Series: Quantitative Applications in the Social Sciences No. 68. London: Sage, 1988.
[21] Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007;57:1358–62.
[22] Teresi JA, Kleinman M, Ocepek-Welikson K. Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. Stat Med 2000;19:1651–83.
[23] Wright BD. Local dependency, correlations and principal components. Rasch Meas Trans 1996;10:509–11.
[24] Wainer H, Kiely GL. Items clusters and computerized adaptive testing—a case for testlets. J Educ Meas 1987;24:185–201.
[25] Reise SP, Ventura J, Keefe RS, et al. Bifactor and item response theory analyses of interviewer report scales of cognitive impairment in schizophrenia. Psychol Assess 2011;23:245–61.
[26] Andrich D. Cronbach's alpha in the presence of subscales. Presented at: International Conference on Outcome Measurement, Bethesda, MD, September 1–3, 2010.
[27] Thurstone L. Attitudes can be measured. Am J Sociol 1928;33:529–54.
[28] Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J App Measurement 2002;3:205–31.
[29] Andrich D, Lyne A, Sheridan B, Luo G. RUMM 2030. Perth. RUMM Laboratory [SPSS Inc. SPSS 18.0 for Windows 2009.], 2010.
[30] Mills RJ, Young CA, Pallant JF, Tennant A. Rasch analysis of the Modified Fatigue Impact Scale (MFIS) in multiple sclerosis. J Neurol Nneurosurg Psychiatry 2010;81:1049–51.
[31] de Boer AG, Bruinvels DJ, Tytgat KM, et al. Employment status and work-related problems of gastrointestinal cancer patients at diagnosis: a cross-sectional study. BMJ 2011;1:e000190.
[32] Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (2nd ed.). Mahwah, NJ: Erlbaum, 2007.
[33] Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. Value Health 2004;7(Suppl. 1):S22–6.