



LUND UNIVERSITY

Psychometric properties of a Swedish version of the Pearlin Mastery Scale in people with mental illness and healthy people.

Eklund, Mona; Erlandsson, Lena-Karin; Hagell, Peter

Published in:
Nordic Journal of Psychiatry

DOI:
[10.3109/08039488.2012.656701](https://doi.org/10.3109/08039488.2012.656701)

2012

[Link to publication](#)

Citation for published version (APA):
Eklund, M., Erlandsson, L.-K., & Hagell, P. (2012). Psychometric properties of a Swedish version of the Pearlin Mastery Scale in people with mental illness and healthy people. *Nordic Journal of Psychiatry*, 66(6), 380-388. <https://doi.org/10.3109/08039488.2012.656701>

Total number of authors:
3

General rights

Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Running head: Rasch analysis of Mastery-S

Psychometric Properties of a Swedish Version of the Pearlin Mastery Scale in People with Mental Illness and Healthy People

Eklund, Mona, Ph.D., Reg. O.T., Professor ¹⁾

Erlandsson, Lena-Karin, Ph.D., Reg. O.T., Associate Professor ^{1, 2)}

Hagell, Peter, Ph.D., R.N., Associate Professor ¹⁾

¹⁾ Department of Health Sciences, Lund University, Sweden.

²⁾ The Vårdal Institute, The Swedish institute for Health Sciences, Lund, Sweden

Corresponding author:

Mona Eklund, Department of Health Sciences,

Lund University, P.O. Box 157, SE-221 00 Lund, Sweden.

Tel: +4646 222 1957

Fax: +4646 222 1959

Email: mona.eklund@med.lu.se

Abstract

BACKGROUND. Mastery refers to the degree to which people perceive that they can control factors that influence their life situation, and has been found important for people's quality of life and well-being. It is thus essential to be able to measure mastery in a valid and reliable way.

AIM. This study aimed at using the Rasch measurement model for investigating the psychometric properties of a Swedish version of the Pearlin Mastery Scale (Mastery-S).

METHODS. A sample of 300 healthy individuals and 278 persons with mental illness responded to the Mastery-S. Item responses were Rasch analyzed regarding model fit, response category functioning, differential item functioning (DIF) and targeting, using the partial credit model.

RESULTS. The Mastery-S items represented a logical continuum of the measured construct but one item displayed misfit. Reliability (Person Separation Index) was 0.7. The response categories did not work as expected in three items, which could be corrected for by collapsing categories. Three items displayed DIF between the two subsamples, which caused a bias when comparing mastery levels between subsamples, suggesting the Mastery-S is not truly generic.

CONCLUSIONS. The Mastery-S may be used to obtain valid and reliable data, but some precautions should be made. If used to compare groups, new analyses of DIF should first be made. Users of the scale should also consider exempting item 6 from the scale and analyze it as a separate item. Finally, rewording of response categories should be considered in order to make them more distinct and thereby improve score reliability.

Key words: Control, coping, validity, reliability, psychometrics.

Word count: 3652

Introduction

It is increasingly recognised that the ability to manage daily activities and live an independent and satisfying life is dependent on the degree to which people perceive that they can control the factors that influence their life situation (1-3). The phenomenon expressed in such a belief is often termed mastery (4). Mastery is an important factor in public health since it is closely related to quality of life and well-being (5). It is thus essential to be able to measure mastery in a valid and reliable way, and it is also important to scrutinise the theoretical underpinnings of the mastery construct.

A few studies have elaborated on theoretical aspects of the mastery construct. In the original works of Pearlin and associates (4, 6) mastery was understood as a coping mechanism when dealing with stress. It was also seen as a mainly stable personal resource. DeSocio, Kitzman and Cole (7) used mastery to operationalise self-agency, seen as the conceptual understanding of self as an agent capable of shaping motives, behaviour and future possibilities. Self-agency also encompasses a continuous self, existing from the past into the future, which allows for the envision of a future self, compatible with current motivations and intentions (7). More recently, the concept of empowerment has been used to denote the control of one's life and recovery process (8), which has clear similarities with mastery. Mastery has also been theoretically linked with well-being, and according to Ryff and Singer (9), mastery is one of six components that comprise well-being. Research thus indicates that the construct of mastery may theoretically be seen as a personal resource related to coping, self-agency, the existence of a continuous self, empowerment and well-being.

A commonly used instrument when assessing mastery is the Pearlin Mastery

Scale (6). Still, although it has been widely used (1, 7, 10-11), there are very few studies of its psychometric properties and no study appears to have been performed in a Swedish context. Therefore, this study aimed at using the Rasch measurement model (12) to investigate the psychometric properties of a Swedish version of the Pearlin Mastery Scale. The Rasch model was used because it mathematically defines the requirements for measurement and allows scales to be examined in a way that is freed from the distribution of the study sample (12-14). By determining the extent to which observed data conform with model specifications, Rasch analysis provides a powerful means of assessing a scale's measurement properties (13-15).

Methods

This study was based on re-analysis of data collected for other studies (16-18), which all complied with stipulations in the Swedish act regulating research ethics and were approved by the local ethical review board. All procedures followed were in accordance with the Helsinki Declaration.

Participants and procedure

The participants were sought in two contexts – a working sample of women and men employed in a state-run organisation and another sample consisting of people in outpatient psychiatric care.

The working sample consisted of employees from six South Swedish national health insurance offices. The sample was selected for a study that aimed to explore salutogenetic factors at work. All employees at the selected insurance offices received an information letter that invited them to the study, detailed the aims and

study design and explained the principles of voluntariness and confidentiality. They were invited to fill out a computer based questionnaire and participation was taken as the agreement to participate. This is in line with the Swedish act regulating research ethics. At the time of data collection, a total of 424 persons (75 men and 349 women) were employed at the six offices. A total of 382 individuals logged into the questionnaire, of whom 64 (16 men, 48 women) decided not to participate. A further 18 persons had to be excluded because they either did not complete the Mastery Scale or did not fill in information about gender and age. In all 300 individuals (250 females), representing 71 % of the total staff group, responded to the questionnaire, including the mastery instrument (16). The data collection process did not allow for any further analyses of non-responders. Socio-demographic data for the participants are presented in Table I. They worked in average 37.5 hours per week, and the majority was administrative officials and investigators working with consulting assignments with clients. The data was collected in 2006.

The sample of people with mental illness was recruited from different units. Presumptive participants received an information letter and then gave their written informed consents. The data was collected face-to-face. One subsample was formed by patients visiting an outpatient unit for people with psychosis, previously used in studies investigating relationships between satisfaction with daily activities and well-being. The selection procedure, described in more detail elsewhere (17), was based on systematic sampling among patients of working ages with at least two years of history of psychiatric illness. Diagnoses were set by the psychiatrists responsible for the clients, according to the ICD-10 (19). Those data were collected 2001-2002. A second subsample was formed by people of working ages visiting day centres for people with psychiatric disabilities in three Swedish municipalities. The inclusion

criterion was visiting a day centre for four hours per week or more. A third subsample was composed of patients visiting outpatient psychosis units in the same municipalities, and inclusion criteria were two years of history of psychiatric illness and not visiting a day centre or the like for four hours or more. Both of these latter subsamples are described more fully in Eklund and Sandlund (18). No records were kept, since the day centres were not part of the medical care, and the diagnoses were based on self-report in both of these latter subsamples. The self-reported diagnoses were subsequently grouped according to the ICD-10 system by an independent experienced psychiatrist. Data from subsamples two and three were collected 2006-2008. In all, the sample of people with mental illness comprised 278 individuals. Based on psychiatrists' assessments for the first subsample and self-report in the other two, 59% had schizophrenia or other psychoses, 26% had a mood or anxiety disorder and 15% had other psychiatric diagnoses. See Table I for sociodemographic characteristics.

< Table I >

The Mastery scale

A Swedish version of the Pearlin Mastery Scale was used. The original American scale (4, 6) has shown good construct and predictive validity and good internal consistency according to classical test theory criteria (20). It consists of seven items with four ordered response categories (1=strongly agree; 2=agree; 3=disagree; 4=strongly disagree), where 4 indicates the highest level of self-mastery. The Swedish version tested here (Mastery-S) was initially developed for a project investigating the consequences following job loss and unemployment (21). It has been used in a number of studies (1, 11, 22-23), but psychometric testing appears to

be lacking. For this study, the Mastery-S was back-translated into English by a professional translator and compared to the original, which revealed some deviances in four out of the seven items. For example, the original wording of item 3 is “I have little control over the things that happen to me” while the retranslation was “I haven’t much say in matters that concern my life”. Although the Mastery-S may not be fully comparable with the original, its frequent use warrants investigation of its psychometric properties.

Data analysis

The Mastery-S was analyzed psychometrically according to the Rasch measurement model (12-13) for ordered response categories (the partial credit model) (14, 24-25).

The Rasch model

The Rasch model (12-14, 24) defines, mathematically, what is required from data (item responses) for total scores to express valid measurement. In terms of mastery, the Rasch model is based on the notion that people with low mastery levels will have low probabilities to affirm items expressing high levels of mastery. The model separately locates persons and items on a common logit (log-odd units) metric, with mean item location set at zero logits. The Rasch model requires unidimensionality (items represent a common underlying latent variable) and local independence (each item response provides unique information). Both these aspects are reflected in the fit of data to the model and violation of either distorts measurement (26-27).

Model fit is assessed by examining the accordance between expected and observed responses across person locations (class intervals) on the measured construct (13, 24). Item fit is supported by non-significant standardised residuals (i.e., the discrepancy between observed and expected item responses) that range between -2.5

and +2.5 (13, 24). Large negative residuals signal local dependency, whereas large positive residuals suggest violation of unidimensionality. The associated chi-squared value represents an overall indication of the interaction between item responses across class intervals and the measured variable; the closer the empirical responses accord with model expectations, the smaller the chi-squared value and the larger the associated P-value. However, the interpretation of chi-squared values is complicated by the facts that they are sensitive to sample size (15, 24). It is therefore recommended that sample size should be set to $n=500$ when large data sets are analyzed (preserving all other empirical aspects of the data), and that chi-squared values should be used as an order statistic (15, 24). That is, a smooth increase in chi-squared values followed by a larger increase would suggest misfit of the item(s) representing the increase. In addition, fit statistics are complemented by judgement of the correspondence between observed and expected item responses through visual inspection of the item-characteristic curve (ICC).

The Rasch model also provides a means to assess whether response categories work as assumed (24). Ordered response categories (e.g., 0 – 1 – 2 – 3) are expected to reflect an increasing amount of the variable under investigation. The threshold between two adjacent categories is the point where there is a 50/50 probability of scoring, e.g., 2 or 3. Disordered thresholds indicate that the response categories do not work as intended. This may be due to, e.g. too many response categories or ambiguous wording. Collapsing categories with disordered thresholds may provide clues regarding how the scale may be improved (15, 24).

Differential item functioning (DIF) is an additional aspect of fit to the Rasch model that may bias scale scores (13). DIF analyses assess whether subgroups of people with similar levels on the measured construct respond systematically different

to items (28). That is, the presence of DIF suggests that an item does not work the same way in subgroups of people. When DIF is uniform (i.e. item responses differ uniformly between subgroups across the measured construct) this can be adjusted for by splitting the item into two new items, one for each subgroup (24).

Within the Rasch framework, reliability can be assessed by the person separation index (PSI). The PSI is analogous to coefficient alpha and indicates how many statistically distinct groups of people (separated by three standard errors) a scale is able to separate between (14, 29).

Targeting assesses how well a scale corresponds to the levels of, e.g., mastery experienced by respondents. If a scale is well targeted to the sample, the mean sample location approximates the mean item location (i.e., zero). Examination of the relationship between the locations of people and items also reveals how successful a set of items is in mapping out a continuum of the measured variable (14-15).

Analysis plan

Sample size was set to $n=500$ in all analyses of model fit (15, 24). We then examined model fit for each item by means of fit residuals, chi-squared values and associated P-values, and inspection of ICCs.

Functioning of response categories was examined and if disordered thresholds were found, categories were collapsed to explore if a smaller number of categories would improve the scale. The presence of DIF was assessed by means of a 2-way ANOVA of the differences between item response functions among healthy and mentally ill people. Items with significant F-values ($P<0.05$ following Bonferroni correction) were considered to have DIF. In case of DIF, these items were split into two new items (one for each subgroup). The impact of any observed DIF was

assessed by testing whether DIF influenced estimated differences in mastery levels between healthy and mentally ill people by means of t-tests and effect sizes (differences between groups divided by the overall SD).

Reliability was assessed by the person separation index (PSI). We also assessed targeting and the extent to which the points of measurement (i.e. the locations of response category thresholds) mapped out an evenly spaced quantitative continuum, without significant gaps (indicating compromised measurement ability and larger measurement error) or clustering (indicating item measurement redundancy) (15). The logic of the hierarchical ordering of item locations was considered in order to assess internal content and construct validity.

Finally, the scale was analyzed among healthy and mentally ill individuals separately.

All analyses were conducted using the RUMM2030 software (Rumm laboratory Pty Ltd., Perth). P-values were adjusted according to Bonferroni.

Results

Data quality was good with an average of 1.2% missing responses per item (range, 0.9-1.5%). Table II shows item fit to the Rasch model. Two items (numbers 2 and 6) displayed signs of misfit to the model. In both instances, the nature of the misfit suggested that these items may represent a different construct than that measured by the scale as a whole. Misfit was most pronounced for item 6 (fit residual, 4.36), whereas it was marginal and non-significant for item 2 (fit residual, 2.68). This was also reflected in the ICCs for the respective items (Figure 1). Accordingly, inspection of the ordered chi-squared values identified item 6 as misfitting (Figure 2).

< Table II >

< Figure 1 >

< Figure 2 >

Examination of the response categories revealed disordered thresholds in items 1, 3 and 5. In all instances, disordering involved the third category. Following collapsing of the second and third categories in these items, the rating scale worked as intended, as exemplified by item 5 (Figure 3).

< Figure 3 >

Analyses of DIF did not reveal any non-uniform DIF but three items (numbers 2, 5, 6) displayed uniform DIF between healthy and mentally ill responders. For items 2 and 5, healthy people scored systematically higher than people with mental illness despite having the same levels of mastery. Item 6 showed DIF in the opposite direction. To explore the impact of the observed item level DIF we compared mastery levels between healthy and mentally ill responders using the DIF-adjusted and the original (non-DIF adjusted) item sets (Table III). Although both comparisons resulted in the same conclusion (small but statistically significant difference), the estimated difference was more pronounced for the non-DIF-adjusted item set. This suggests that DIF contributed to a bias favouring healthy people.

< Table III >

Reliability (PSI) was 0.704, indicating that the scale can separate people into two statistically distinct strata. Figure 4 illustrates the locations of item response category thresholds relative to the locations of the sample. The mean person location is 0.59 (SD, 0.98) logits (bars above the x-axes) relative to the mean item location of 0 logits (bars below the x-axes). This means that there is a general tendency for the items to

represent less mastery than that experienced by the persons. While the thresholds are able to map out a continuum, there are also several gaps as well as clusters along the continuum (Figure 4). Specifically, the scale fails to represent higher levels of mastery and tends to be somewhat redundant around the centre of the person distribution.

< Figure 4 >

The hierarchy that the mastery items represent is illustrated in Table II, which displays the items in location order. In general, the ordering appears logical as for example only people with high levels of mastery would say that they “can do just about anything”, whereas those with low levels of mastery feel that they have “little control over things that happen” to them.

Finally, the scale was analyzed separately among healthy and mentally ill individuals, respectively. In the healthy sample, there was only marginal misfit for item 2 (fit residual, 2.68), whereas item 6 did not misfit (fit residual, 1.57). However, reliability was relatively low (PSI, 0.62) and the third response category thresholds were disordered in all items but numbers 4 and 6. Mean (SD) person location was 0.917 (SD,0.897; range, -0.552 - 4.158) logits. Among people with mental illnesses, item 6 showed misfit similar to that in the main analysis (fit residual, 4.18) and there was marginal misfit for item 5 (fit residual, -2.64). However, reliability was acceptable (PSI, 0.75) and there was no response category threshold disordering. Mean (SD) person location was 0.390 (SD,1.080; range, -2.146 - 3.478) logits.

Discussion

This study reached its aim of clarifying the psychometric properties of the Mastery-S,

by application of the Rasch measurement model. Although the Mastery-S displays some differences in item wording compared with the original Pearlin Mastery Scale, it showed acceptable reliability. Moreover, while there was some scarcity of items representing higher levels of mastery, the scale represents a logical continuum of the measured construct.

There were, however, signs of misfit to the Rasch model. Specifically, data suggest that item 6 (“what happens to me in the future depends on me”) may represent a different construct than that measured by the scale as a whole. The content of this Mastery-S item agreed with the original, so this observation does not appear to be an artefact of the translation. Notably, however, the misfit of this item appears to have emanated from responses provided by mentally ill people. This item also displayed DIF between people with mental illness and healthy people. Specifically, those with a mental illness had a higher probability of endorsing this item than healthy people, regardless of their levels of mastery. Imagining a realistic future self may be problematic for people with a severe mental illness, which may explain the DIF.

The fact that there was DIF and that this manifested in a small but clear difference in effect sizes (0.10) of the differences between healthy and mentally ill people’s mastery levels as derived from the original and DIF adjusted scale suggests that the Mastery-S is not truly generic. DIF may yield misleading conclusions, and in the present case it produced an artificially large difference between the two groups. Therefore DIF should be checked and controlled for when the scale is used to compare groups.

The response scale showed to be non-optimal, since response category thresholds were reversed for three items (1, 3 and 5). Collapsing the second and third response alternatives produced a better response scale, with a logical threshold order.

However, post hoc exercises such as collapsing response categories are exploratory and need empirical verification. Future modifications of the scale should therefore consider revising its response categories. This could be done either by reducing the number of response categories from four to three, or by modifying their wording in order to obtain more clearly distinctive meanings. In general, the latter option would be preferable since reducing response categories also reduces measurement precision, person separability and score reliability (14-15). These concerns appear to primarily apply when using the scale among healthy individuals, since scores had poor reliability and response category threshold disordering was more pronounced in this subsample. The latter finding is consistent with observations regarding the interpretation of response categories among people with Parkinson's disease and neurologically healthy control subjects, where the perceived difference between "disagree" and "totally disagree" was marginal, particularly among control subjects (30).

The fact that there was some redundancy regarding items at the average level of "difficulty" and scarcity of mastery items of greater "difficulty", i.e. items that are less likely to be endorsed, means that ceiling effects can be expected in groups likely to experience high levels of mastery. This may have some clinical and practical implications, such as that the Mastery-S will be less suitable for detecting differences between groups with stronger sense of Mastery or follow changes over time in such samples.

Methodological discussion

The sample with no known illness consisted of mainly women. That skewed gender distribution, not deliberately sought, may be seen as a methodological problem and

may have contributed to the observed DIF. Another factor that may have influenced the DIF is the variation between the samples regarding family situation. Very few among those with a mental illness were married or had children, compared with the healthy sample. Thus, this study leaves some uncertainty to whether the DIF identified was caused solely by the difference in health condition between the groups and to what extent socio-demographic factors contributed. Additional studies of samples that are more balanced in these respects are needed to firmly address these issues. Moreover, the administration format differed between the samples and this may have affected the results in some unknown way. For example, it cannot be excluded that the computerized administration contributed to response category disordering. Another discrepancy between the two samples was that reliability was relatively low among healthy people. The explanation for this most probably lies in the larger spread in Mastery-S scores among people with mental illnesses (24, 31).

Conclusion and recommendation

We suggest that the Mastery-S may be used to obtain valid and reliable data in the context of public health, but some precautions should be made. Firstly, if used to compare groups, DIF analyses should initially be performed to estimate its impact. Similarly, reliability needs to be verified when using the scale; we found it only marginally acceptable in the main analysis and below acceptable levels among healthy people. Thirdly, users of the scale should consider exempting item 6 from the scale and analyze it as a separate item. Furthermore, in future development of the scale rewording of response categories should be considered in order to make them more distinct and thereby improve score reliability and potentials to separate mastery levels among people. Finally, given that no previous study appears to have conducted this type of analyses on the Pearlin Mastery Scale, our observations have important

implications for the use and further scientific inquiry on the original version, as well as any of its translations.

References

1. Eklund M. Perceived control: how is it related to daily occupation in patients with mental illness living in the community? *Am J Occup Ther* 2007; 61:535-42.
2. Koplas PA, Gans HB, Wisely MP, Kuchibhatla M, Cutson TM, Gold DT, et al. Quality of life and Parkinson's disease. *J Gerontol A Biol Sci Med Sci* 1999;54:M197-202.
3. Riazi A, Thompson AJ, Hobart JC. Self-efficacy predicts self-reported health status in multiple sclerosis. *Mult Scler* 2004;10:61-6.
4. Pearlin LI, Menaghan EG, Lieberman MA, Mullan JT. The stress process. *J Health Social Behav* 1981;22:337-56.
5. Eklund M, Bäckström M, Hansson L. Personality and self-variables: important determinants of subjective quality of life in schizophrenia out-patients. *Acta Psychiatr Scand* 2003;108:134-43.
6. Pearlin LI, Schooler C. The structure of coping. *J Health Soc Behav* 1978;19:2-21.
7. DeSocio J, Kitzman H, Cole R. Testing the relationship between self-agency and enactment of health behaviors. *Res Nurs Health* 2003;26:20-9.
8. Rogers ES, Chamberlin J, Ellison ML, Crean T. A consumer-constructed scale to measure empowerment among users of mental health services. *Psychiatr Serv* 1997;48:1042-7.
9. Ryff CD, Singer B. Psychological well-being: meaning, measurement, and implications for psychotherapy research. *Psychother Psychosom* 1996;65:14-23.
10. Shanahan MJ, Bauer DJ. Developmental properties of transactional models: the case of life events and mastery from adolescence to young adulthood. *Dev Psychopathol* 2004;16:1095-117.
11. Erlandsson L-K, Eklund M. The relationships of hassles and uplifts to experience of health in working women. *Women Health* 2003;38:19-37.
12. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danmarks Paedagogiske Institut; 1960.
13. Andrich D. Rasch models for measurement. Beverly Hills: Sage Publications, Inc; 1988.
14. Wright BD, Masters GN. Rating scale analysis. Chicago: MESA Press; 1982.
15. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009;13:iii, ix-x, 1-177.
16. Erlandsson L-K, Carlsson G, Horstmann V, Gard G, Holmström E. Health factors in daily life and work of public service employees in Sweden. *Work*. in press.

17. Eklund M, Hansson L, Ahlqvist C. The importance of work as compared to other forms of daily occupations for wellbeing and functioning among persons with long-term mental illness. *Comm Mental Health J* 2004;40:465-77.
18. Eklund M, Sandlund M. The life situation of people with persistent mental illness visiting day centers – a comparative study. *Comm Ment Health J*. in press.
19. WHO. The ICD-10 classification of mental and behavioural disorders. Geneva: World Health Organization; 1993.
20. Marshall GN, Lang EL. Optimism, self-mastery, and symptoms of depression in women professionals. *J Pers Soc Psychol* 1990;59:132-9.
21. Brenner S-O, Arnetz B. Effekter av otrygghet i arbetet. Förlust av arbetet och arbetslöshet.: Statens institut för psykosocial miljömedicin, stressforskning Lfk;1983. Report No: 170.
22. Bengtsson-Tops A. Mastery in patients with schizophrenia living in the community: relationship to sociodemographic and clinical characteristics, needs for care and support, and social network. *J Psychiatr Mental Health Nurs* 2004;11:298-304.
23. Hildingh C, Fridlund B, Baigi A. Sense of coherence and experiences of social support and mastery in the early discharge period after an acute cardiac event. *J Clin Nurs* 2008;17:1303-11.
24. Andrich D, Sheridan B, Luo G. Interpreting RUMM. Perth: RUMM Laboratory Pty Ltd; 2004-2005.
25. Luo G. The relationship between the Rating Scale and Partial Credit Models and the implication of disordered thresholds of the Rasch models for polytomous responses. *J Appl Meas* 2005;6:443-55.
26. Marais I, Andrich D. Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *J Appl Meas* 2008;9:200-15.
27. Marais I, Andrich D. Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model. *J Appl Meas* 2008;9:105-24.
28. Holland PW, Wainer H. Differential item functioning. Mahwah, NJ: Lawrence Erlbaum Associates; 1993.
29. Andrich D. An index of person separation in latent trait theory, the traditional KR-20 index, and the Guttman scale response pattern. *Education Research and Perspectives* 1982;9:95-104.
30. Knutsson I, Rydstrom H, Reimer J, Nyberg P, Hagell P. Interpretation of response categories in patient-reported rating scales: a controlled study among people with Parkinson's disease. *Health Qual Life Outcomes* 2010;8:61.
31. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. New York: Oxford University Press Inc., 2008.

Acknowledgment

We wish to thank all participating respondents for their cooperation. The study was financially supported by the Swedish Research Council, grant number K2009-69X-20067-04-3.

Disclosure of interest

We have no disclosure of interest to report.

Figure legends

Figure 1.

Graphical illustration of Rasch model fit of items 2, 6 and 4 of the Mastery-S scale.

The item characteristic curve (ICC; grey curve) represents the expected item responses (y-axis) at various levels of the measured construct (x-axis). Black dots represent the observed responses in the sample as divided into ten class intervals according to their locations on the measured construct, indicated by the marks on the x-axis. Item 2 (panel A) has a somewhat large but non-significant fit residual value (2.68) and item 6 (panel B) exhibits a more pronounced deviation from model expectations (fit residual, 4.36). This item shows under discrimination as the empirical observations tend to parallel the x-axis. This suggests that the item may not represent the same construct as the test as a whole, i.e. it violates unidimensionality. For comparison, panel C depicts an item (number 4) displaying good model fit statistics (fit residual, 1.99).

Figure 2.

Chi-squared values (y-axis) of the item-trait interaction goodness-of-fit for items of the Mastery-S scale in ascending order (x-axis). A large “jump” from a smooth increase in chi-squared values suggestive of model misfit (24).

Figure 3.

Example category probability curves from the Mastery-S scale. Location on the measured construct is indicated on the x-axis (with threshold locations centered at

zero; negative values = less mastery) and the y-axis represents the probability of affirming response categories 1 through 4 relative to the location on the measured construct (x-axis). Panel A shows an item (no. 5) representing the typical pattern with disordered thresholds between response categories 2-to-3 and 3-to-4. Panel B displays response category functioning after collapsing the second and third categories (old categories 2 and 3 rescored into new category 2, and old category 4 becomes new category 3).

Figure 4.

Distributions of the locations of people (upper histogram: mean, 0.59; SD, 0.98) and response category thresholds (lower histogram: mean, 0; SD, 0.26) on the common logit metric. Response category thresholds are the locations where there is a 50/50 probability of endorsing either of two adjacent categories and represent the “notches” on the latent ruler defined by the items. Mean item locations are tabulated below the graph. This represents the observed item hierarchy for the Mastery-S scale.

Table I. *Socio-demographic data for the two samples.*

	<i>Healthy sample N=300</i>	<i>Sample with mental illness N=278</i>	<i>P-value</i>
Age, mean (SD)	47.7 (9.9)	43.7 (10)	<.001
[min - max]	[25-66]	[25-64]	
Sex			<.001
Males, n (%)	50 (16.7)	147 (53.1)	
Females, n (%)	250 (83.3)	130 (46.9)	
Marital status			<.001
Married/cohabitant, n (%)	238 (79.6)	68 (24.9)	
Single, n (%)	61 (20.4)	205 (75.1)	
Having children living at home			<.001
Yes, n (%)	135 (45.5)	44 (17.1)	
No, n (%)	162 (54.5)	214 (82.9)	

Table II. *Rasch item and fit statistics for the Mastery-S scale*^a.

Item		Item statistics ^b		Fit statistics	
No.	Content (abridged)	Location	SE	Residual ^c	Chi squared ^{d,e}
3	Little control over things that happen to me	-0.395	0.050	-0.825	20.663
7	Little I can do to change important things in life	-0.141	0.051	-0.472	14.804
5	Feel helpless in dealing with problems	-0.077	0.049	-2.100	22.012
1	No way I can solve some of my problems	-0.033	0.049	-1.492	11.859
6	What happens to me in the future depends on me	-0.005	0.052	4.360	41.059
2	Feel I'm being pushed around	0.263	0.049	2.684	8.928
4	Can do just about anything	0.388	0.056	1.991	5.244

^a Performed with the sample divided into ten class intervals according to person locations on the measured construct.

^b Expressed in linear log-odds units (logits). Mean item location is zero with positive values representing higher levels of mastery.

^c Residuals summarise the deviation of observed from expected responses. Deviation from the recommended range of -2.5 to +2.5, indicating item misfit, are bold.

^d Chi-square values summarise the deviation of observed from expected responses across the three class intervals of people. Higher values represent larger deviations.

^e Analysed with sample size set to n=500. Bonferroni corrected (0.05 / 7) statistically significant deviations across class intervals (i.e. $P < 0.007$), indicating item misfit, are bold.

SE, standard error.

Table III. *Comparisons (t-tests) of mastery levels among healthy and mentally ill responders according to the original and DIF adjusted Mastery-S scale.*

	Healthy (n=300) ^a	Mentally ill (n=349) ^a	P-value	Effect size ^b
Original scale	0.77 (0.86)	0.43 (1.06)	<0.0001	0.35
DIF adjusted scale	0.74 (0.89)	0.49 (1.07)	0.0012	0.25

^a Data are mean (SD) logit values.

^b Mean difference between groups divided by the overall SD (original scale, 0.985; DIF adjusted scale, 0.998).

DIF, differential item functioning.

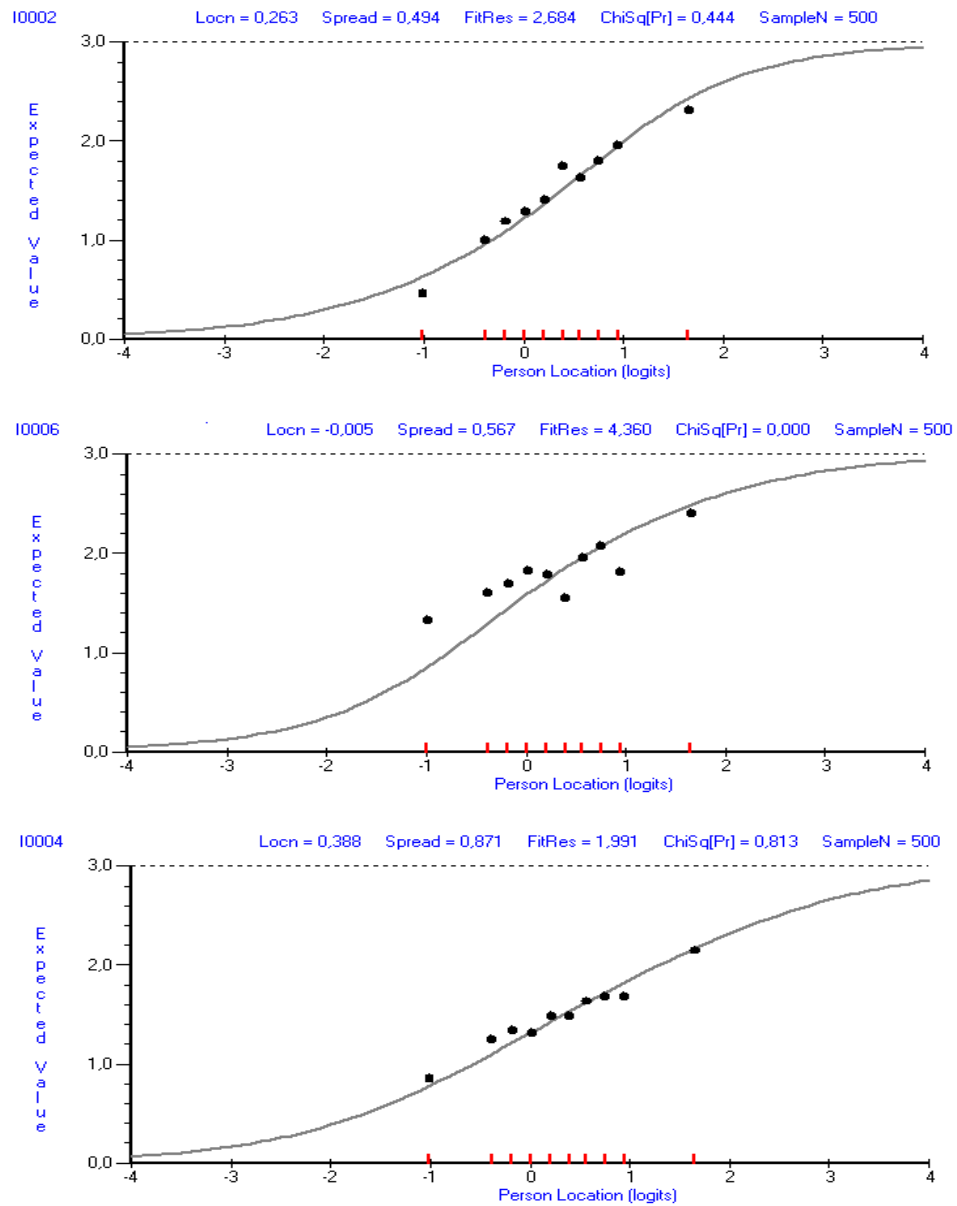


Figure 1

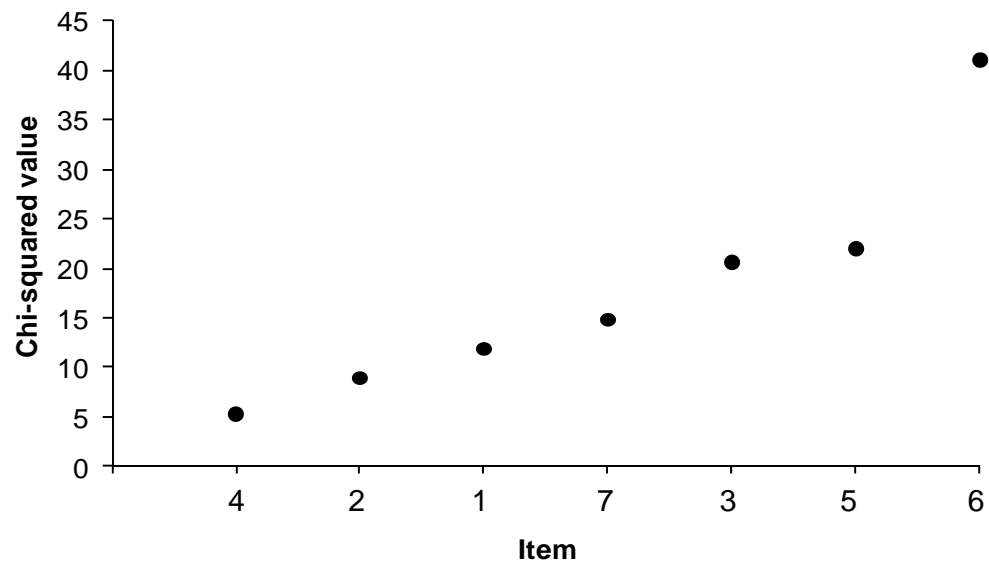


Figure 2

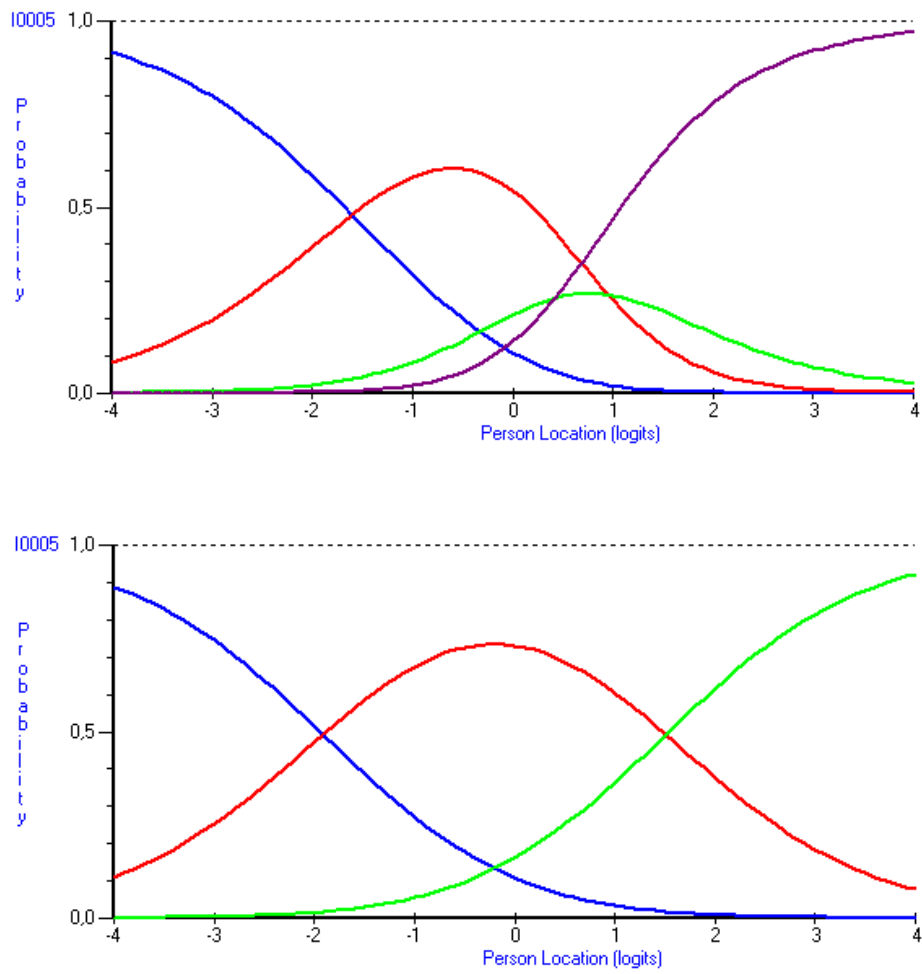


Figure 3

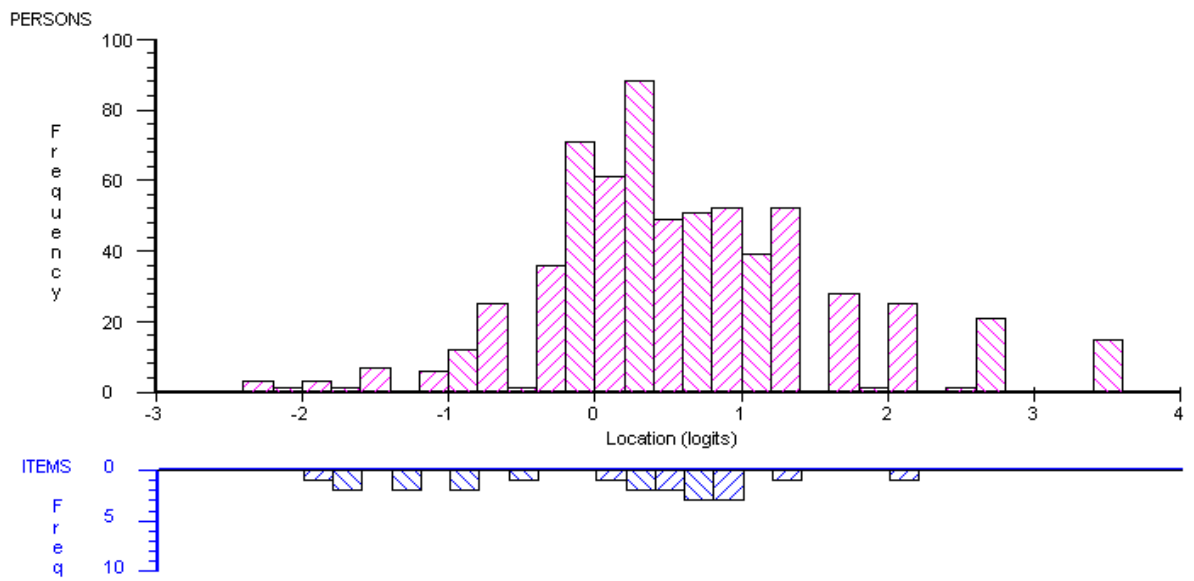


Figure 4