Quality of Life Research (2019) 28:221–231
https://doi.org/10.1007/s11136-018-2005-2

The 88-item Multiple Sclerosis Spasticity Scale: a Rasch validation of the Italian version and suggestions for refinement of the original scale

Leonardo Pellicciari[1] · Marcella Ottonello[2] · Andrea Giordano[3] · Caterina Albensi[4] · Franco Franchignoni[5]

Accepted: 15 September 2018 / Published online: 20 September 2018
© Springer Nature Switzerland AG 2018

Abstract

Background In multiple sclerosis (MS), the impact of spasticity on the patient's life is a key issue, and it is fundamental that existing tools measuring the patient's perspective undergo psychometric analysis and refinement to optimize confidence in their use in clinical practice and research.

Objective We examined—by Rasch analysis (RA)—the main metric characteristics of the 88-item Multiple Sclerosis Spasticity Scale (MSSS-88) to: (i) further validate its Italian version (MSSS-88-IT), previously validated through classical test theory methods only and (ii) independently verify the measurement properties of the original scale.

Methods MSSS-88 data from a convenience sample of 232 subjects with MS underwent RA, mainly examining item fit, reliability indices, test information function, dimensionality, local item independence, and differential item functioning (DIF).

Results Most items fitted the Rasch model, but 13/88 items showed a misfit in infit and/or outfit values. Rasch reliability indices were high (> 0.80). Test information functions in most subscales showed a sharp decrease in measurement precision as the ability level departs from the quite limited central range of maximal information. The unidimensionality of each subscale was confirmed. Thirteen item pairs showed local dependency (residual correlations > 0.30) and three items presented DIF.

Conclusion Reliability, dimensionality and some internal construct validity characteristics of the MSSS-88-IT were confirmed. But, drawbacks of the original MSSS-88 emerged related to some item misfit, redundancy, or malfunctioning. Thus, further large independent studies are recommended, to verify the robustness of previous findings and examine the appropriateness of a few targeted item replacements.

Keywords Psychometrics · Rasch model · Patient-reported outcome measure · Spasticity · Multiple sclerosis

Introduction

Spasticity is one of the most common disabling symptoms of multiple sclerosis (MS), and its severity increases as the disease progresses [1].This symptom is characterized by a velocity-dependent increase in muscle tone, typically in the

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11136-018-2005-2) contains supplementary material, which is available to authorized users.

✉ Leonardo Pellicciari
leonardo.pellicciari@gmail.com

[1] Unit of Functional Rehabilitation, Azienda USL Toscana Centro, Via Rozzalupi, 57, 50053 Empoli, FI, Italy

[2] Department of Physical & Rehabilitation Medicine, ICS Maugeri SpA SB, Via Missolungi, 14, 16167 Nervi, GE, Italy

[3] Bioengineering Service, ICS Maugeri SpA SB, Via Revislate, 13, 28010 Veruno, NO, Italy

[4] Specialty School in Physical and Rehabilitation Medicine, Department of Clinical Sciences and Translational Medicine, University of Rome "Tor Vergata", Viale Montpellier, 1, 00133 Rome, Italy

[5] Department of Physical Medicine and Rehabilitation, ICS Maugeri SpA SB, Via Bernasconi 16, 20035 Lissone, MB, Italy

lower legs, which produces impairment of walking ability and balance [2, 3] and has a marked negative impact on the patient's well-being and quality of life [4]. In clinical trials on MS and therapeutic interventions, the patient's own perspective on the impact of spasticity on their life is widely considered to be a key outcome [5]. Hence, the measurement of the patient's experiences and perceptions is of vital importance in their clinical management, and patient-reported outcome measures have been indicated as a key tool for measuring treatment benefit (or risk) in medical product clinical trials [6].

To investigate the perceived impact of spasticity on the patient's life, an 88-item patient-reported outcome measure has been developed, the Multiple Sclerosis Spasticity Scale (MSSS-88) [7]. The MSSS-88 quantifies the perceived burden of spasticity in three main domains (spasticity-specific symptoms, physical functioning, and psychosocial impact), with a total of eight subscales [7]. The original version of the MSSS-88 (developed according to Rasch measurement principles) has been demonstrated to be a reliable and valid interval-level measure of the impact of spasticity in MS [7], and the scale has been used in several clinical studies [8–13]. Serbian [14], German [15] and Italian [16] versions of the scale have been developed, each version demonstrating its own reliability and validity when analyzed by classical theory test (CTT) methods.

However, the only study [17] that—to our knowledge—replicated a Rasch analysis (RA) of the MSSS-88 subscales reported a series of psychometric flaws. Thus, a further RA of the MSSS-88 seems indicated to ensure more confident use of the scale in clinical practice and research. The aim of this study was: (a) to perform a secondary psychometric validation (through RA) of the Italian version of the MSSS-88, so far analyzed only by CTT [16] and (b) to independently verify (in a different dataset and applying additional analyses) the original findings by Hobart et al. on the measurement properties of the scale.

Methods

Subjects

From January 2008 to March 2014, all patients with a clinical diagnosis of MS with spasticity, aged over 18 years, who accessed the day-hospital rehabilitation service of the Department of Clinical Sciences and Translational Medicine of the Tor Vergata Polyclinic Foundation in Rome, Italy, were recruited for this study. Exclusion criteria were problems with reading and understanding the Italian language and the presence of severe cognitive or communication impairments. Further details on the study methodology (including the enrolment procedure) are available in the previous report on validation of the Italian version of MSSS-88 [16]. Patients were recruited by the participating physicians at the time of their regular visit if they met the inclusion criteria. This paper is a secondary study of a previous research, which was carried out with ethical approval of the local institutional review board [11].Written signed informed consent was obtained from all participants.

Outcome measure

The MSSS-88 [7] is a patient-reported outcome measure designed to quantify the impact of spasticity on the MS patient's life. It is composed of 88 items subdivided into eight subscales: three concerning spasticity-specific symptoms ('muscle stiffness': 12 items; 'pain and discomfort': 9 items; 'muscle spasms': 14 items), three concerning physical functioning ('activities of daily living': 11 items; 'walking': 10 items; 'body movements': 11 items), and two concerning the psychosocial impact ('emotional health'; 13 items; and 'social functioning': 8 items). Each item is rated on a 4-level Likert scale, from 1 (not at all bothered) to 4 (extremely bothered). The validated Italian version of the MSSS-88 [16] was administered to all patients immediately before the start of their day-hospital physical therapy program.

Statistical analysis

Descriptive statistics were used to describe the characteristics of the sample; continuous variables were expressed as mean ± standard deviation, categorical variables as absolute frequency and percentage, and ordinal variables as median and 25th and 75th percentiles. RA was performed on the matrix of item responses for each subscale of the MSSS-88 (rating scale model), using Winsteps software analysis program, version 3.69.1.96 [18]. RA is a statistical methodology to test whether the properties of a questionnaire comply with a wide range of psychometric requirements (assessment of response format, item content, appropriate targeting, unidimensionality, etc.) which cannot be analyzed using CTT techniques [19]. An unconditional maximum likelihood estimation procedure was performed to estimate the model parameters. The steps of analysis were as follows:

Rating scale diagnostics First, we investigated whether each of the rating scale categories was being used in the expected manner, using the following criteria suggested by Linacre [20]: (1) at least 10 observations per category; (2) monotonic increase in both average measures across rating scale categories and thresholds (the average measure for a category is the average ability of the people who respond in that category; thresholds, also known as step calibrations, are the points at which the probability of a response in one or another of 2 adjacent categories is equally likely, i.e., they represent the transition from one category to the next; (3) threshold differences > 1 and < 5 logits; and (4) category outfit mean-square values (MnSq) < 2.

Item fit statistics Item validity was assessed by evaluating the fit of individual items to the latent trait as per the Rasch model and examining whether the pattern of item difficulties was consistent with the model expectancies [19]. Information-weighted (infit) and outlier-sensitive (outfit) MnSq for each item were calculated. Infit values are based on the information-weighted Chi square statistics (i.e., each observation is weighted by its statistical information); conversely, outfit values are based on outlier-sensitive Chi square statistics (more sensitive to unexpected responses to items that are far from the respondent's measure level) [21].Given the sample size, we considered MnSq > 0.75 and < 1.30 as an indicator of acceptable fit [18]. Items outside this range were considered underfitting (MnSq > 1.30 suggested the presence of an expectedly high variability) or overfitting (MnSq < 0.75 indicated a too predictable pattern).

Reliability was evaluated in terms of separation, defined as the ratio of the true spread of the measures with their measurement error [21, 22], in order to determine whether each subscale of MSSS-88 could divide the studied sample into different levels of spasticity. Along the measurement construct, the item separation index gives an estimate (in SE units) of the spread or separation of items, whereas the person separation index gives an estimate of the spread of persons. A separation of 2.0 is considered good [22], and a related index is the reliability of these separation indices, providing the degree of confidence that can be placed in the consistency of the estimates (range 0–1; coefficients > 0.80 are considered good and > 0.90 excellent) [21].

Targeting and test information function Targeting was addressed informally comparing the distribution of the person parameter estimates together with the distribution of the item thresholds (on the person-item map). The test information function was calculated to examine the amount of information (degree of measurement precision) yielded by the test at any person-ability level along the latent variable (i.e., how well the test is estimating ability over the whole range of ability scores). The Fisher information value is independent of the particular patient sample undergoing the test.

Dimensionality and local dependence A principal component analysis (PCA) on the standardized residuals was performed to analyze: (1) the proportion of variance attributable to both the Rasch factor and the first residual factor (as a further confirmation of the unidimensionality of each subscale). The following criteria were used to determine whether additional factors were likely to be present in the residuals: a cut-off of 50% of the variance, explained by the Rasch factor, and eigen value of the first residual factor smaller than 3; and (2) the local independence of items. Large positive correlations of Rasch residuals for 2 items (usually above 0.30) indicates that they may not be locally independent, either because they duplicate some feature of each other or because they both incorporate some other shared dimension [18].

Differential item functioning (DIF) DIF analysis was carried out to examine the stability of item hierarchy and difficulty across each of the following subsamples: males vs. females; younger vs. older subjects (split at the sample median age of 46 years). DIF in each subsample was investigated separately, calibrating the scale for each group, to obtain an estimate of the item difficulties in each group, using as anchor values the person calibrations on the full sample, and then performing pairwise *t*-tests between the two sets of item difficulties (two sided, $\alpha = 0.05$ with Bonferroni's correction). The general 'a-priori' hypothesis was to not find DIF between the analyzed groups (apart from possible minor differences attributable to typical gender- or age-related characteristics, such as in household chores).

Results

Sample characteristics

The study sample consisted of 232 patients with definite diagnosis of MS. Their main demographic and clinical characteristics are reported in Table 1.

Rasch analysis

The 4-level rating scale fulfilled the category functioning criteria in terms of monotonic increase in both average measure for categories and their related thresholds. Moreover, each threshold difference was between 1 and 5 logits and the category outfit mean-square values were always lower than 2. Item fit statistics showed that most items of each of the eight MSSS-88 subscales fitted the respective measure constructs. Just 1–2 items for each subscale were either under fitting (MnSq > 1.3) or overfitting (MnSq < 0.75) in infit and/or outfit values (Tables 2, 3, 4, 5). In detail: (a) nine items showed an underfit (i.e., they elicited some unexpected responses), of which four had infit values (those more sensitive to unexpected responses on items that are near the person's ability level) ≤ 1.40 and (b) four items were overfitting in both fit values.

When underfitting persons with infit and/or outfit MnSq higher than 2.0 (5–9% of the sample, in the different subscales) were temporarily excluded from the computation

Table 1 Main demographic and clinical characteristics of the sample ($N=232$)

Characteristics	Mean ± SD	Frequency (%)	Median (25th–75th percentile)
Age (years)	45.3 ± 11.2		
Sex			
Female		134 (57.8%)	
Male		98 (42.2%)	
Disease type			
Relapsing-remitting		140 (57.8%)	
Secondary progressive		63 (26.1%)	
Primary progressive		39 (16.1%)	
Duration of symptoms (years)	11.4 ± 9.8		
Time since MS diagnosis (years)	8.2 ± 7.8		
EDSS score			3.5 (1.0–4.5)
Modified Ashworth Scale at lower limbs (0–24[a])			3.1 (1.0–7.8)
Educational level			
No formal education		2 (0.9%)	
Primary		2 (0.9%)	
Secondary		62 (26.7%)	
High school		142 (61.2%)	
University degree		24 (10.3%)	
Employment			
Employed		156 (67.2%)	
Not employed		36 (15.5%)	
Retired		40 (17.3%)	
Marital status			
Married		142 (61.2%)	
Unmarried		90 (38.8%)	

SD standard deviation, *EDSS* Expanded Disability Status Scale, *MS* multiple sclerosis

[a]24 = highest spasticity level

(because they create more noise than information), all underfitting items maintained their underfit except the one in walking subscale.

Table 6 shows for each of the eight MSSS-88 subscales: (a) person ability (mean and range); (b) the span of item difficulty, which ranged from 1.78 logits (walking) to 4.03 logits (muscle spasms), and the item threshold values; and (c) the reliability indices: in particular, the item separation reliability spanned from 0.95 to 0.99 and the person separation reliability from 0.82 to 0.91.

The graphs related to test information function for each subscale are shown in the Online Appendix. The maximum value of the test information function was acceptable but not high (around 5, equivalent to a classical test theory reliability estimate of about 0.80) in 6 out of the 8 MSSS-88 subscales, i.e., all but ‘emotional health’ and ‘muscle spasms’ where that value was about 7 and 8, respectively. All functions were bell-shaped, with a peak in the central area (mostly, from -1 to 1 of person ability, θ). As the ability levels moved towards the two extremes of each subscale, there was a rather sharp decline in information with a corresponding significant increase of the standard error of the ability estimates.

As for dimensionality and local dependence, the PCA of standardized residuals showed that:

(1) the variance explained by RA in the different subscales of the MSSS-88 ranged from 50.1% (muscle spasms) to 63.8% (walking).The unexplained variance in the first contrast was below 10%, and the corresponding eigen values ranged from 1.8 to 2.4 in all subscales except emotional health, which showed an unexplained variance of 11.2% in the first contrast, with an eigen value of 3.3.
(2) six out of the eight subscales had at least a couple of locally-dependent items, i.e., displaying positive correlations of their residuals > 0.30 (range 0.31–0.66) (Tables 2, 3, 4, 5).

Table 2 Summary of Rasch analysis related to "muscle stiffness" and "pain and discomfort" subscales, containing item-difficulty measures (with standard error, SE), fit information, and local dependence between items

Item number and label	Measure (SE)	Fit (MnSq)		Local item dependence
		Infit	Outfit	
Muscle stiffness subscale				
1. Stiffness when walking?	−1.07 (0.11)	1.01	0.97	2 (0.66)
2. Stiffness anywhere in your lower limbs?	−1.10 (0.11)	0.82	0.77	1 (0.66)
3. Stiffness when you are in the same position for a long time?	−0.70 (0.11)	0.88	1.01	
4. Stiffness first thing in the morning?	0.12 (0.11)	1.06	1.09	
5. Tightness anywhere in your lower limbs?	−0.32 (0.11)	0.89	0.85	
6. Your lower limbs feeling rigid?	−1.04 (0.11)	*0.57*	*0.57*	
7. Stiffness when standing up?	−0.06 (0.11)	0.82	0.86	
8. Tightness in your muscles?	−0.19 (0.11)	1.00	1.02	
9. Stiffness that is unpredictable?	0.40 (0.11)	1.15	1.11	
10. Feeling that your muscles are pulling?	0.17 (0.11)	1.19	1.22	
11. Stiffness in your whole body?	1.95 (0.12)	1.20	1.21	12 (0.50)
12. Your whole body feeling rigid?	1.83 (0.12)	**1.49**	**1.62**	11 (0.50)
Pain and discomfort subscale				
13. Feeling restricted and uncomfortable?	−1.01 (0.10)	**1.59**	**1.59**	
14. Feeling uncomfortable sitting for a long time?	−0.17 (0.11)	1.11	1.24	
15. Painful or uncomfortable spasms?	0.26 (0.11)	0.89	0.85	
16. Pain when in the same position for too long?	−0.57 (0.10)	0.76	0.75	
17. Feeling uncomfortable lying down for a long time?	0.33 (0.11)	0.89	0.86	18 (0.31)
18. Difficulties finding a comfortable position to sleep in bed?	0.30 (0.11)	1.17	1.09	17 (0.31)
19. Pain in the muscles on getting out of bed in the morning?	0.10 (0.11)	0.88	0.92	
20. Pain in the muscles provoked by movement?	−0.09 (0.10)	0.82	0.79	
21. Constant pain in the muscles?	0.84 (0.12)	0.90	0.85	

The higher the item measure (from negative to positive values), the more difficult that item was for the group to endorse (i.e., more negative measure = higher scores and more frequently bothering the patient; more positive measure = lower scores and less frequently bothering the patient). The misfitting items are in bold (underfit) or italics (overfit). In the "Local Item Dependence" column, the correlated item and the degree of correlation of the residuals are reported

Finally, in general, the items did not show DIF apart from: (i) item 3 'stiffness when you are in the same position for a long time' (muscle stiffness subscale) between males and females; (ii) item 37 'Doing housework such as cooking or cleaning' (activities of daily living subscale) between males and females; and (iii) item 85 'feeling reluctant to go out' (social functioning subscale) between younger and older persons (borderline difference).

In the Online Appendix, for each subscale there is a graph showing the conversion function between raw scores and corresponding Rasch measures, with their error estimate.

Discussion

Spasticity is one of the most common physical signs and symptoms experienced by people with MS. The presence of spasticity can significantly impair motor performance, functional activities and participation [4]. As such, the treatment of this symptom represents an important therapeutic target, and an accurate measurement of spasticity (and its impact on functions) is crucial to assess the effectiveness of treatment [23, 24].

The MSSS-88 was developed in 2006 to examine the impact of spasticity on quality of life in people with MS, and to evaluate their subjective perception of the impact of spasticity on different clinical areas, including both physical and psychological [7].The scale's validation was performed using RA, a relatively modern psychometric technique for developing and refining rating scales and questionnaires with sound measurement properties. To the best of our knowledge, only one study [17] replicated the RA of the MSSS-88 subscales, reporting a series of psychometric flaws, even though the global performance was generally good. The flaws, unfortunately, were not examined in detail and only briefly discussed. The present study provides a further replication of RA of the MSSS-88, to both verify the results of the previous studies [7, 17] and gather new elements for further refinement of the scale.

Table 3 Summary of Rasch analysis of the "muscle spasms" and "activities of daily living" subscales, containing item-difficulty measures (with standard error, SE), fit information, and local dependence between items

Item number and label	Measure (SE)	Fit (MnSq)		Local item dependence
		Infit	Outfit	
Muscle spasms subscale				
22. Spasms that come on unpredictably?	−0.95 (0.11)	0.96	0.98	23 (0.40)
23. Powerful or strong spasms?	−0.20 (0.12)	1.07	0.99	22 (0.40)
24. Spasms when first getting out of bed in the morning?	−0.18 (0.12)	0.90	0.72	25 (0.42)
25. Spasms provoked by changing positions?	−0.36 (0.12)	0.75	0.76	24 (0.42)
26. Spasms provoked by movement?	−0.66 (0.11)	0.80	0.82	
27. Spasms where your leg kicks out in front of you?	−0.69 (0.11)	1.24	1.23	
28. Spasms provoked by certain positions?	−0.68 (0.11)	*0.50*	*0.51*	
29. Spasms disturbing sleep?	−0.05 (0.12)	**1.37**	**1.58**	
30. Spasms when doing certain tasks?	−0.59 (0.11)	0.75	0.77	
31. Spasms when traveling over bumps or cobbles?	0.20 (0.13)	1.14	1.17	
32. Spasms where your knees pull up?	0.48 (0.14)	1.15	0.88	
33. Spasms causing legs to hit things?	0.51 (0.14)	1.25	1.15	
34. Spasms provoked by touch?	0.70 (0.13)	**1.65**	**1.67**	
35. Spasms pushing you out of a chair or wheelchair?	3.08 (0.34)	1.28	0.78	
Activities of daily living subscale				
36. Putting on your socks or shoes?	−0.73 (0.12)	**1.54**	**1.61**	
37. Doing housework such as cooking or cleaning?	−0.90 (0.12)	**1.33**	**1.35**	
38. Getting in and out of a car?	−1.12 (0.12)	0.79	0.77	
39. Getting in and out of shower and/or bath?	−1.11 (0.12)	0.86	0.89	
40. Sitting up in bed?	−0.56 (0.12)	0.78	0.80	41 (0.44) 42 (0.33)
41. Getting into or out of bed?	−0.08 (0.13)	0.91	0.86	40 (0.44)
42. Turning over in bed?	0.38 (0.13)	1.06	1.05	40 (0.33)
43. Getting into or out of a chair?	0.47 (0.13)	0.81	0.77	
44. Getting dressed or undressed?	0.68 (0.13)	0.77	0.76	
45. Getting on or off the toilet seat?	0.99 (0.13)	0.91	0.87	
46. Drying yourself with a towel?	1.98 (0.14)	**1.36**	1.25	

For details, see caption of Table 2

We used the Italian version of the MSSS-88. This version was obtained according to well-established international guidelines [25], involving forward/backward translations, discussion with experts, and cognitive debriefing. In our previous study [16], we had validated the MSSS-88 using factor analysis (to determine its dimensionality), and CTT methods, which examined its properties of internal consistency (Cronbach's alpha and item-total correlation) and validity (correlation with other widely-used and well-respected measurement tools). In each MSSS-88 subscale, a good internal consistency and criterion-related validity were found, as well as an acceptable unidimensionality, as expected given the thorough methodology used for their construction. In line with the statement by the scale creators specifying that each subscale represents a stand-alone measurement instrument [7], in the present study RA was performed on each MSSS-88 subscale, for an additional detailed analysis of: scale diagnostics (category functioning), internal construct validity, reliability issues, scale dimensionality, local independence of items, and DIF. We think that our results provide useful indications for refining the measurement properties of the questionnaire.

As a first step, we examined whether rating scale categories were being used in the expected manner [20]. According to RA diagnostics, respondents were able to correctly discern between the four response categories of the questionnaire. Each category was well defined, mutually exclusive, sufficiently univocal and exhaustive. This result confirmed the findings of the original study related to the tool's development [7], showing the appropriateness of the use of four response options in all subscales.

Next, for each subscale, we fitted the data to the Rasch model [19]: most items fitted the model. Nine items showed underfit (in most of them there was an unexpectedly high variability in both infit and outfit values): however, according to Linacre [26], all their mean-square values indicate an unproductive but not degrading impact on measurement system. Regarding the overfit of four items (one in each of four

Table 4 Summary of Rasch analysis of the "walking" and "body movements" subscales, containing item-difficulty measures (with standard error, SE), fit information, and local dependence between items

Item	Measure (SE)	Fit (MnSq)		Local item dependence
		Infit	Outfit	
Walking subscale				
47. Difficulties walking smoothly?	−0.78 (0.12)	0.95	0.93	
48. Being slow when walking?	0.07 (0.12)	1.20	1.20	
49. Having to concentrate on your walking?	0.86 (0.12)	1.09	1.03	
50. Having to increase the effort needed for you to walk?	0.39 (0.12)	0.76	0.78	
51. Being slow when going up or down stairs?	−0.92 (0.12)	0.83	0.85	
52. Being clumsy when walking?	−0.24 (0.12)	*0.61*	*0.61*	
53. Tripping over or stumbling when walking?	0.25 (0.12)	0.99	1.05	
54. Feeling like you are walking through treacle?	−0.15 (0.12)	**1.40**	**1.58**	
55. Losing your confidence to walk?	−0.04 (0.12)	0.89	0.82	
56. Feeling embarrassed to walk?	0.57 (0.12)	1.21	1.17	
Body movements subscale				
57. Difficulties moving freely?	−0.82 (0.13)	0.92	0.90	58 (0.31)
58. Difficulties moving smoothly?	−1.44 (0.11)	0.89	0.87	57 (0.31)
59. Limited range of movement?	−0.80 (0.11)	0.97	0.93	
60. Difficulties moving parts of your body?	−0.43 (0.11)	0.71	0.70	
61. Difficulties bending your limbs?	−0.14 (0.11)	1.00	0.97	
62. Your body being resistant to movement?	0.34 (0.12)	0.78	0.76	
63. Your body or limbs feeling locked?	−0.04 (0.11)	0.92	0.85	
64. Awkward or jerky movement?	1.52 (0.13)	**1.67**	**1.47**	
65. Difficulties straightening your limbs?	0.67 (0.13)	0.94	0.81	
66. Difficulties relaxing parts of your body?	−0.04 (0.11)	1.27	1.29	
67. No control over your body?	1.17 (0.12)	1.0	1.00	

For details, see caption of Table 2

MSSS-88 subscales), their pattern of response was probably too predictable from the overall pattern of response to the other items [18] and thus they contribute little extra information over the other items in the scale, but they do not degrade the quality of measurement. Indeed, also in both previous studies on MSSS-88, the item fit statistics indicated some item misfit [7, 17], though the Authors did not provide further details or comment. Worthy of note, the items 'awkward or jerky movement' (body movement subscale) and 'lower limbs feeling rigid' (muscle stiffness subscale) were misfitting both in our study and in the original one [7].

The hierarchy of item difficulty did not exhibit any major divergence from Hobart et al.'s study [7]. The span of item difficulty (ranging from 1.78 logits in the Walking subscale to 4.03 logits in the muscle spasms subscale) was acceptable but for some subscales—as suggested by the original Authors [7]—an extension of their range of measurement would be beneficial. Regarding the targeting (the extent to which the items were appropriately difficult for our sample), in some subscales ('muscle spasms', 'activities of daily living', 'emotional health', and 'social functioning') the mean subject 'ability' (here, impact of spasticity on the patient's life) of our sample was < 1 logit lower than the mean item difficulty (conventionally set at 0 logits). This mismatch is mainly due to the low disability levels of our patient sample. However, the present finding confirms Ball et al.'s [17] observation that in their study most shapes of person measurement distribution were skewed to some extent, with the potential risk of underestimating improvement or worsening; moreover, the percentage of their extreme scores in one subscale (muscle spasms subscale) was > 15%.

In order to examine how well the test is doing in estimating ability over the whole range of ability scores, we analyzed the test information function [18, 21], an indicator independent of the particular patient sample taking the test. The ideal test information function should be rather flat or at least show a plateau over a large region of the ability scale. Conversely, most of the MSSS-88 subscales showed similar bell-shaped test information function distributions, with maximal information of about 5 (not very high) in 6 out of 8 subscales (all but 'muscle spasms' and 'emotional health'), and a measurement precision rather steadily and sharply decreasing as the ability level departs from the quite limited central range (about 4 logits) of maximal information (i.e., where the test works

Table 5 Summary of Rasch analysis of the "emotional health" and "social functioning" subscales, containing item-difficulty measures (with standard error, SE), fit information, and local dependence between items

Item	Measure (SE)	Fit (MnSq)		Local item dependence
		Infit	Outfit	
Emotional health subscale				
68. Feeling frustrated?	0.02 (0.11)	1.10	1.10	
69. Feeling less confident in yourself?	−0.07 (0.11)	0.96	1.04	
70. Feeling inadequate?	0.18 (0.11)	0.92	0.97	
71. Feeling low?	1.09 (0.13)	0.84	0.83	
72. Feeling irritated?	−0.63 (0.11)	0.94	0.88	73 (0.47) 80 (0.37)
73. Feeling angry?	−0.75 (0.11)	1.04	1.00	72 (0.47) 80 (0.33)
74. Feeling depressed?	−0.01 (0.11)	0.85	1.00	
75. Loss of self-worth?	0.65 (0.12)	1.00	1.00	76 (0.36)
76. Feeling like a failure?	1.43 (0.13)	1.06	0.91	75 (0.36)
77. Feeling frightened?	−0.50 (0.11)	1.26	1.21	79 (0.40)
78. Crying (tearful)?	−0.20 (0.11)	*0.65*	*0.64*	
79. Feeling panicky?	−0.15 (0.11)	1.18	1.12	77 (0.40)
80. Feeling nervous?	−1.06 (0.11)	1.12	1.13	72 (0.37) 73 (0.33)
Social functioning subscale				
81. Difficulties going out?	−0.66 (0.12)	0.96	0.94	
82. Feeling isolated?	0.57 (0.13)	0.76	0.75	
83. Feeling vulnerable?	−1.03 (0.11)	1.13	1.27	
84. Difficulties finding energy for other people?	−0.74 (0.12)	1.00	0.98	
85. Feeling reluctant to go out?	−0.44 (0.12)	1.06	1.02	
86. Feeling less sociable?	0.27 (0.13)	0.86	0.83	
87. Difficulties with relationships with other family members?	1.12 (0.14)	1.29	1.14	
88. Difficulties interacting with people?	0.92 (0.14)	0.90	0.76	

For details, see caption of Table 2

Table 6 Person-ability (mean and range) levels, person separation index and reliability, item difficulty (range), and item separation index and reliability of each of the eight subscales of the Multiple Sclerosis Spasticity Scale-88

MSSS-88 subscales	Mean person ability (range)	PSI	PSR	Item-difficulty range (span)	Item threshold values	Item separation index	Item separation reliability
Muscle stiffness	−0.22 (−4.77 to 5.20)	3.15	0.91	−1.10 to 1.95 (3.05)	−3.26 to 4.35	8.29	0.99
Pain and discomfort	−0.94 (−3.91 to 4.13)	2.29	0.84	−1.01 to 0.84 (1.85)	−2.85 to 2.86	4.43	0.95
Muscle spasms	−1.76 (−4.13 to 3.05)	2.15	0.82	−0.95 to 3.08 (4.03)	−2.44 to 4.67	5.91	0.97
ADL	−1.86 (−4.83 to 2.77)	2.45	0.86	−1.12 to 1.98 (3.10)	−3.43 to 4.33	6.49	0.98
Walking	0.28 (−4.60 to 4.85)	2.99	0.90	−0.92 to 0.86 (1.78)	−3.3 to 3.45	4.07	0.94
Body movements	−0.81 (−4.67 to 4.60)	2.85	0.89	−1.44 to 1.52 (2.96)	−3.61 to 3.62	6.63	0.98
Emotional health	−1.16 (−4.32 to 3.55)	2.72	0.88	−1.06 to 1.43 (2.49)	−2.84 to 3.23	5.80	0.97
Social functioning	−1.74 (−4.15 to 3.34)	2.11	0.82	−1.03 to 1.12 (2.15)	−3.09 to 3.19	5.78	0.97

ADL activities of daily living, *MSSS-88* Multiple Sclerosis Spasticity Scale, *PSI* Person separation index, *PSR* Person separation reliability

best). This means that outside that optimal range the scale measures 'ability' with much less precision (i.e., with quite high standard error of estimation).

For this reason, we agree with Ball et al.'s concerns [17] about "the possibility that all 11 scales may have the potential to underestimate the true change in target variables".

Rasch reliability indexes confirmed their high values, which give us a good degree of confidence in the consistency of both person-ability and item-difficulty estimates.

Unidimensionality is a core assumption of the Rasch model [21]. It ensures that a response to any item is explained only by the respondent's amount of the trait being investigated and not by other factors. In our study, the unidimensionality testing with robust psychometric techniques confirmed that each MSSS-88 subscale measures a single underlying dimension [18]. Only the 'emotional health' subscale showed the presence of a small secondary dimension (11.2% of unexplained variance in first contrast, with a related eigen value of 3.3).

As for DIF [18], just three items showed a significant difference between the subgroups analyzed: (i) item 3 'stiffness when you are in the same position for a long time' (muscle stiffness subscale) where females were more bothered than males; (ii) item 37 'doing housework such as cooking or cleaning' (activities of daily living subscale) where males were more bothered than females; and (iii) item 85 'feeling reluctant to go out' (social functioning subscale) where older persons were more bothered than younger ones. According to our local culture, we considered at least the latter two differences as referable to typical gender-related (doing housework) or age-related (going out) characteristics, and as such negligible.

Conversely, additional information regarding optimal item selection came from the residual correlation matrix [27]. They showed inter-item residual correlations higher than 0.30 in one to two pairs of items in five subscales ('body movements', 'pain discomfort', 'muscle stiffness', 'muscle spasms', and 'activities of daily living'), while the 'emotional health' subscale showed as many as five pairs of items with large residual correlations. In the Rasch model, this inter-item dependency indicates a violation of the basic assumption of statistical local independence by the corresponding pair of items, and suggests that the items concerned tend to duplicate some feature of each other, or to incorporate some other shared dimension [18]. This seems the case for item 11 'stiffness in your whole body?' and item 12 'your whole body feeling rigid?' (muscle stiffness subscale), and for items 72 'feeling irritated?' and 73 'feeling angry?', or item 77 'feeling frightened' and 79 'feeling panicky' (emotional health subscale), as well as for some pairs of items in other subscales. There is empirical evidence that they are synonyms or near-synonyms. Thus, some of these 13 pairs of locally-dependent items (particularly those with higher local dependence and similar item difficulty) seem candidates for further consideration, such as possible testlet aggregation, or even deletion/replacement, because their presence degrades the measurement process and increases standard errors [18, 27].This result is not surprising because in Ball et al.'s study 33 item biases were found (i.e., with correlations among residuals $r > .30$).

Overall, our findings principally show the presence in MSSS-88 subscales of:

(a) a few misfitting items, some of them already misfitting in the original study [7];
(b) 13 pairs of locally-dependent items, some of them containing words or phrases that are synonyms or near-synonyms, and/or with a similar item-difficulty estimates;
(c) an acceptable but not wide item-difficulty range in some subscales, and sometimes several items with similar location;
(d) test information functions in most subscales characterized by a rather quick decrease in information (precision) and a corresponding significant increase of the standard error of the estimate when measuring subjects with extreme ability levels at both ends.

In line with Ball et al.'s findings of a suboptimal performance of some MSSS-88 items and subscales [17], all these results point to the need for further examinations and attempts to provide substantive explanations for the anomalies detected, to better understand the constructs for measurement, and then probably decide in some subscales a replacement of 1–2 poorer items with well-targeted ones (i.e., with higher statistical information), for both psychometric and clinical reasons [21, 28].

Care should be taken in interpreting our data. First, our analysis was conducted in a convenience sample of quite young Italian-speaking patients with MS (consecutively recruited when undergoing a day-hospital rehabilitation program), showing a low level of disability; this could lessen the generalizability of some of our findings. Second, we cannot exclude that some (linguistic, cultural or technical) characteristics of the Italian version of the MSSS-88 might have influenced our results. Although our version was checked using a thorough procedure of 'forward/backward translation' followed by pilot testing and expert revision, without significant semantic difficulty emerging (apart from the need to adapt the term 'treacle', which is unfamiliar in the Italian culture), a full check of a cross-cultural adaptation would require that metric properties of the new language version are assessed in comparison with those of the source version, applying DIF analysis to test whether the difficulty hierarchy of the items is similar across versions (as well as other testing situations).

Conclusion

In conclusion, the main results of our Rasch study are as follows:

(1) The MSSS-88 confirmed several positive metric characteristics; however, a series of drawbacks emerged related to the measurement performance of some items.
(2) The reliability, dimensionality, and some internal construct validity characteristics of the Italian version of the MSSS-88 were confirmed by the present study, thus strengthening the positive results of our previous validation of the questionnaire using CTT methods [16].

We recommend that further large independent studies—analyzing subjects with a wide range of disease severity and in different contexts—be carried out on MSSS-88, to verify the robustness of previous findings and examine the usefulness of a few targeted item replacements. The aim should be to create a MSSS 2.0 able to optimize the scale's coverage and technical quality with the most convenient type and number of items.

Acknowledgements This paper is based on research conducted by LP and MO during their PhD program in "Advanced Sciences and Technologies in Rehabilitation Medicine and Sports", at the University of Rome "Tor Vergata", Rome, Italy. They express special thanks to Prof. Calogero Foti and Prof. Diego Centonze for their helpful support in the PhD theses.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Rizzo, M. A., Hadjimichael, O. C., Preiningerova, J., & Vollmer, T. L. (2004). Prevalence and treatment of spasticity reported by multiple sclerosis patients. *Multiple Sclerosis, 10*, 589–595. https://doi.org/10.1191/1352458504ms1085oa.
2. Sosnoff, J. J., Socie, M. J., Boes, M. K., Sandroff, B. M., Pula, J. H., Suh, Y., Weikert, M., Balantrapu, S., Morrison, S., & Motl, R. W. (2011). Mobility, balance and falls in persons with multiple sclerosis. *PLoS ONE, 6*, e28021. https://doi.org/10.1371/journal.pone.0028021.
3. Flachenecker, P., Henze, T., & Zettl, U. K. (2014). Spasticity in patients with multiple sclerosis—Clinical characteristics, treatment and quality of life. *Acta Neurologica Scandinavica, 129*, 154–162. https://doi.org/10.1111/ane.12202.
4. Hughes, C., & Howard, I. M. (2013). Spasticity management in multiple sclerosis. *Physical Medicine and Rehabilitation Clinics North America, 24*, 593–604. https://doi.org/10.1016/j.pmr.2013.07.003.
5. Anwar, K., & Barnes, M. P. (2009). A pilot study of a comparison between a patient scored numeric rating scale and clinician scored measures of spasticity in multiple sclerosis. *NeuroRehabilitation, 24*, 333–340. https://doi.org/10.3233/NRE-2009-0487.
6. Food and Drug Administration (FDA). (2009). Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed 26 May 2018.
7. Hobart, J. C., Riazi, A., Thompson, A. J., Styles, I. M., Ingram, W., Vickery, P. J., Warner, M., Fox, P. J., & Zajicek, J. P. (2006). Getting the measure of spasticity in multiple sclerosis: The Multiple Sclerosis Spasticity Scale (MSSS-88). *Brain, 129*, 224–234. https://doi.org/10.1093/brain/awh675.
8. Mehta, L. R., McDermott, M. P., Goodman, A. D., & Schwid, S. R. (2010). A randomized trial of memantine as treatment for spasticity in multiple sclerosis. *Multiple Sclerosis, 16*, 248–251. https://doi.org/10.1177/1352458509355462.
9. Schyns, F., Paul, L., Finlay, K., Ferguson, C., & Noble, E. (2009). Vibration therapy in multiple sclerosis: A pilot study exploring its effects on tone, muscle force, sensation and functional performance. *Clinical Rehabilitation, 23*, 771–781. https://doi.org/10.1177/0269215508101758.
10. Sosnoff, J., Motl, R. W., Snook, E. M., & Wynn, D. (2009). Effect of a 4-week period of unloaded leg cycling exercise on spasticity in multiple sclerosis. *NeuroRehabilitation, 24*, 327–331. https://doi.org/10.3233/NRE-2009-0486.
11. Mori, F., Ljoka, C., Magni, E., Codecà, C., Kusayanagi, H., Monteleone, F., Sancesario, A., Bernardi, G., Koch, G., Foti, C., & Centonze, D. (2011). Transcranial magnetic stimulation primes the effects of exercise therapy in multiple sclerosis. *Journal of Neurology, 258*, 1281–1287. https://doi.org/10.1007/s00415-011-5924-1.
12. Balantrapu, S., Sandroff, B. M., Sosnoff, J. J., & Motl, R. W. (2012). Perceived impact of spasticity is associated with spatial and temporal parameters of gait in multiple sclerosis. *ISRN Neurology, 2012*, 675431. https://doi.org/10.5402/2012/675431.
13. Zajicek, J. P., Hobart, J. C., Slade, A., Barnes, D., & Mattison, P. G., MUSEC Research Group (2012). Multiple sclerosis and extract of cannabis: Results of the MUSEC trial. *Journal of Neurology, Neurosurgery & Psychiatry, 83*, 1125–1132. https://doi.org/10.1136/jnnp-2012-302468.
14. Rodic, S. Z., Knezevic, T. I., Kisic-Tepavcevic, D. B., Dackovic, J. R., Dujmovic, I., Pekmezovic, T. D., Drulovic, J. S., & Konstantinovic, L. M. (2016). Validation of the Serbian version of Multiple Sclerosis Spasticity Scale 88 (MSSS-88). *PLoS ONE, 11*, e0147042. https://doi.org/10.1371/journal.pone.0147042.
15. Henze, T., von Mackensen, S., Lehrieder, G., Zettl, U. K., Pfiffner, C., & Flachenecker, P. (2014). Linguistic and psychometric validation of the MSSS-88 questionnaire for patients with multiple sclerosis and spasticity in Germany. *Health and Quality of Life Outcomes, 12*, 119. https://doi.org/10.1186/s12955-014-0119-y.
16. Ottonello, M., Pellicciari, L., Centonze, D., Foti, C., Pistarini, C., Albensi C., & Giordano, A. (2017). The cross-cultural adaptation and psychometric validation of the MSSS-88 for use in Italian patients with multiple sclerosis. *Disability and Rehabilitation, 25*, 1–7. https://doi.org/10.1080/09638288.2017.1393699.
17. Ball, S., Vickery, J., Hobarth, J., Wright, D., Green, C., Shearer, J., Nunn, A., Cano, M. G., MacManus, D., Miller, D., Mallik, S., & Zajicek, J. (2015). The Cannabinoid Use in Progressive Inflammatory Brain Disease (CUPID) trial: A randomised double-blind placebo-controlled parallel-group multicentre trial and economic evaluation of cannabinoids to slow progression in multiple sclerosis. *Health Technology Assessment, 19*, 1–187. https://doi.org/10.3310/hta19120.
18. Linacre, J. M. (2009). *A User's guide to Winsteps-ministep: Rasch-model computer programs. Program Manual 3.68.0*. Chicago: Winsteps.
19. Tesio, L. (2003). Measuring behaviors and perceptions: Rasch analysis as a tool for rehabilitation research. *Journal of Rehabilitation Medicine, 35*, 105–115.

20. Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*, 85–106.
21. Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. New York: Routledge, Taylor & Francis Group.
22. Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: Mesa Press.
23. Bethoux, F., & Marrie, R. A. (2016). A cross-sectional study of the impact of spasticity on daily activities in multiple sclerosis. *Patient, 9*, 537–546. https://doi.org/10.1007/s40271-016-0173-0.
24. Milinis, K., Tennant, A., Young, C. A., & TONiC study group (2016). Spasticity in multiple sclerosis: Associations with impairments and overall quality of life. *Multiple Sclerosis and Related Disorders, 5*, 34–39. https://doi.org/10.1016/j.msard.2015.10.007.
25. Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976), 25*, 3186–3191.
26. Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878.
27. Wright, B. D. (1996). Local dependency, correlations and principal components. *Rasch Measurement Transactions, 10*, 509–511.
28. Wolfe, E. W., & Smith, E. V. Jr. (2000). Instrument development tools and activities for measure validation using Rasch models: Part II—Validation activities. *Journal of Applied Measurement, 8*, 204–234.