Multiple Sclerosis and Related Disorders 36 (2019) 101400

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

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

Examining the validity of the multiple-sclerosis walking scale-12 with Rasch analysis: Results from an Italian study

D. Marengo[a], R. Rosato[a,*], G. Gamberini[b], P. Cavalla[c], M. Gironi[d,e], F. Patti[f], L. Prosperini[g], C. Solaro[b]

[a] Department of Psychology, University of Turin, Via Verdi 10, Turin 10126, Italy
[b] Department of Rehabilitation, "Mons. L. Novarese" Hospital, Moncrivello (VC), Italy
[c] Department of Neurosciences, AOU S. Giovanni Battista, University of Turin, Torino, Italy
[d] INSPE, Neurosciences Division, IRCCS HSR, Milano, Italy
[e] CAM, Polidiagnostic Center, Monza, Italy
[f] Department DANA, "GF Ingrassia", Neuroscience Section (Multiple Sclerosis Centre), University of Catania, Catania, Italy
[g] Department of Neurosciences, S. Camillo-Forlanini Hospital, Roma, Italy

ARTICLE INFO

Keywords:
Multiple sclerosis walking scale-12
Walking
Mobility limitation
Outcome Assessment
Rasch model

ABSTRACT

Background: The Multiple Sclerosis Walking Scale (MSWS-12) has showed good psychometric properties in reports exploring its validity using Classical Test Theory methods. Findings from recent studies using Item Response Theory methods advance the idea that some aspects of the MSWS-12 does not fully comply with some requirements of sound measurement.

Research question: The present study investigated whether the measurement properties of the Italian version of the MSWS-12 met the assumptions of the Rasch model.

Methods: Sample consisted of 199 patients with a diagnosis of MS (50% female; Mean age (SD) = 48.15 (12.33)). Analyses were performed using both unidimensional and multidimensional Rasch modeling techniques.

Results: Rating scale for items assessing need for support when walking, and ability to run, showed significant functioning problems. A revision of the rating scale improved the measurement properties of these items. Items assessing need for support showed signs of multidimensionality and differential item functioning when controlling for patients' disease course and EDSS score. Additionally, included items did not appear to cover the full range of impairment as observed in the sample.

Significance: Emerging findings are consistent with those from previous studies in highlighting the need for a revision of the current content of the MSWS-12, and the inclusion of new items assessing impairment at the lower end of the disability continuum.

1. Introduction

Multiple sclerosis (MS) is a debilitating, neurodegenerative, and disabling disease of the central nervous system (CNS), which affects a wide range of body systems including neuromuscular, motor, and cognitive systems. Impairment of mobility due to progression of MS has a significant impact on the functional independence and quality of life for persons with MS (Compston and Coles, 2008).

The Multiple Sclerosis Walking Scale (MSWS-12), the most widely used patient report measure of walking ability in MS, has generally demonstrated good psychometric properties in reports exploring its validity using Classical Test Theory (CTT) methods, such as exploratory or confirmatory analysis (Hobart et al., 2003; Kieseier and Pozzilli, 2012; McGuigan and Hutchinson, 2004; Motl and Snook, 2008). Recent studies examining the functioning of the MSWS-12 using Item Response Theory (IRT) methods have advanced the idea that some items included in the MSWS-12 do not fully comply with some requirements of sound measurement (Cano et al., 2015; Mokkink et al., 2016; Engelhard et al., 2016). In particular, problems have emerged concerning the functioning of the response format of items assessing ability to run (Item 2), and need for supports (e.g., Item 9). Patients tend to respond to these items using extreme categories (i.e., indicating the lowest or highest levels of disability), while intermediate responses are seldom used (Cano et al., 2015; Mokkink et al.,

* Corresponding author.
E-mail address: rosalba.rosato@unito.it (R. Rosato).

https://doi.org/10.1016/j.msard.2019.101400
Received 17 October 2018; Received in revised form 30 August 2019; Accepted 16 September 2019

2016; Engelhard et al., 2016). Moreover, items assessing need for supports (i.e., Item 8-9) have been shown to not comply with the assumption of measurement invariance across groups differing by severity of MS diagnosis (Mokkink et al., 2016), while age was found to be a source of measurement bias when assessing ability to run (Item 2) (Engelhard et al., 2016). Results from these studies highlight the usefulness of using of IRT methods for the examination of validity of measurement of walking ability using the MSWS-12, as well as for improving its' measurement performance.

In a recent validation using CTT methods, the Italian version of the MSWS-12 has demonstrated high internal consistency, test-retest reliability, and convergent validity with EDSS scores, the timed 25-foot walk, and the Fatigue Severity Scale (Solaro et al., 2015). The present study aimed at exploring the psychometric characteristics of the Italian version of the MSWS-12 using IRT methods, and evaluating possible revisions to the scoring procedure of the scale that would improve its suitability for use in clinical practice, and providing evidences of criterion validity.

2. Material and methods

2.1. Procedure and sample

The present study involved six MS outpatient clinics across Italy. Inclusion criteria were a confirmed MS diagnosis according to McDonald's (Polman et al., 2011), age greater than 18 years, ability to walk with or without the use of supports, and an EDSS range from 0 to 7. Patients experiencing an exacerbation within 30 days prior to the assessment, with an additional neurological disease or with one or more concomitant illnesses, were excluded from the study. Patients signed informed consent prior to enrollment in the study according to the Declaration of Helsinki. Ethical approval was obtained by a multicenter research ethics committee (P.R.196REG2015). For the present study, recruited patients were 199; Table 1 shows descriptive statistics for patients' demographic and clinical characteristics.

3. Instruments

3.1. Multiple-Sclerosis walking scale (MSWS-12)

Patients rated the severity of their walking disability using the Italian adaptation of the MSWS-12 walking-related items (Solaro et al., 2015). A common stem introduces the items: "In the past 2 weeks, how much has your MS…". Each item assesess a specific symptom or functional change, e.g., "Limited your ability to walk?" (Item 1). Items are scored on a five-point Likert scale: 1 (Not at all), 2 (A little), 3 (Moderately), 4 (Quite a bit), and 5 (Extremely). The MSWS-12 score is obtained by summing the item scores, and rescaling on a 0–100 range by subtracting the minimum score, dividing by the range of scores, and then multiplying by 100.

Table 1
Descriptive statistics for recruited patients.

Patient characteristics	
Number of patients	199
Gender (Female)	50.30%
Age, mean (SD); median (range)	48.15 (12.33); 47 (20 − 81)
MS subtype	
Relapsing remitting MS	60.40%
Primary progressive MS	12.20%
Secondary progressive MS	27.40%
Disease duration, median (range)	12 (1–38)
EDSS, median (range)	4.50 (0–7.00)
MSWS-12 score, median (range)	50 (0–100)

3.2. Strategy of analyses

As a preliminary step, reliability of the MSWS-12 was assessed using Cronbach's alpha coefficient; principal component analysis (PCA) was also carried out for a preliminary exploration of dimensionality.

The Rasch partial credit model (Masters, 1982) which allows for polytomous response data, was then applied in order to investigate: item and rating scale functioning, unidimensionality, measurement invariance, reliability and test-targeting, and criterion validity of the MSWS-12 scale.

Item functioning was evaluated using mean-square Infit and Oufit statistics, with values in the range 0.60–1.40 indicating items had adequate fit to the Rasch model (Bond and Fox, 2015; Streiner et al., 2015; Wright, 1994). In presence of items with a malfunctioning rating scale (i.e., disordered and overlapping thresholds), and in order to meet the criterion for adequate functioning of the rating scale (i.e., ordered thresholds), we examine the functioning of a revised scoring procedure in which response categories showing problematic overlap are collapsed into a single response category (Bond and Fox, 2015; Streiner et al., 2015).

Then, we evaluated possible violations of Rasch unidimensionality assumption by examining the presence of local item dependency (LID) between items. A positive residual correlation above 0.20 between items was considered an indication of undesired LID (Christensen et al., 2017). In order to determine further the relevance of emerging LID between pair of items, the fit of the unidimensional Rasch model was compared with an alternative multidimensional IRT (MIRT) model accounting for the existence for LID, i.e. the Rasch testlet model (Wang and Wilson, 2005). In the Rasch testlet model items showing LID are modeled to load on both a specific testlet dimension and on the general dimension. The variance of the estimated testlet effects provide information about their magnitude: simulation studies indicate values smaller than 0.25 as a sign of negligible LID (Wang and Wilson, 2005; Glas et al., ; Wang et al., 2002; Zhang et al., 2010). The fit of the models was compared by testing the significance of the G^2 likelihood-ratio statistic (Briggs and Wilson, 2003). Analyses were implemented using Conquest 3.0 (Adams et al., 2012).

Measurement invariance was evaluated according to gender, age (dichotomized at the median age of 47 years), walking impairment as assessed by EDSS (EDSS > 4.5 vs. EDSS ≤ 4.5), and disease course (relapsing remitting vs. primary and secondary progressive). A significant ($p < 0.05$) difficulty contrast ≥ 0.50 logit was considered an indication of non-negligible DIF across the examined groups (Bond and Fox, 2015).

Next, we examined reliability of person measures, and adequacy of item-person targeting of the MSWS-12. A reliability coefficient of 0.80 or higher was considered acceptable for both persons and items. Test targeting was examined by comparing the means of the item and person measures, and by visual inspection of the person-item map. We also examined possible ceiling and floor effects, which were considered significant if we found more than 15% of patients reporting extreme scores (McHorney and Tarlov, 1995; Terwee et al., 2007). Differences in demographic and clinical characteristics between patients reporting extreme and non-extreme scores were examined using Chi-square tests (categorical variables), and one-way ANOVA with post-hoc Bonferroni tests (continuous scores).

Finally, criterion validity of the revised MSWS-12 score was investigated by examining correlations with EDSS scores and disease duration. Criterion validity with disease course was examined using Student's *t*-test.

Except where indicated otherwise, Rasch analyses were performed using Winsteps 3.68.2 (Linacre, 2009). All other analyses were performed in SPSS version 18 (SPSS, 2009).

Table 2
Item parameters estimates and fit statistics for the revised MSWS-12 after recode of items 2, 8, 9, and 12.

Item	Item location (logit) Measure	S.E.	Item fit Outfit	Infit	Threshold measure (logit) 1 Little	2 Fairly	3 Very	4 Extremely	Rating scale
1 Limited your ability to walk	0.29	0.13	0.82	0.80	−3.77	−0.91	1.40	3.28	01234
1 Limited your ability to run*	−2.91	0.28	1.07	1.06					00111
1 Limited your ability to climb up and down stairs	−0.20	0.12	1.19	1.13	−3.55	−0.95	1.44	3.06	01234
1 Made standing when doing things more difficult	0.37	0.13	0.96	0.99	−3.86	−0.88	1.08	3.67	01234
1 Limited your balance when standing or walking	−0.07	0.13	1.16	1.13	−3.86	−0.65	0.88	3.63	01234
1 Limited how far you are able to walk	−0.55	0.12	0.71	0.75	−2.96	−1.08	1.17	2.87	01234
1 Increased the effort needed for you to walk	−0.23	0.12	0.89	0.84	−3.39	−0.92	1.48	2.83	01234
1 Made it necessary for you to use support when walking indoors*	1.84	0.23	0.84	0.88					00011
1 Made it necessary for you to use support when walking outdoors*	1.07	0.22	0.95	1.01					00011
1 Slowed down your walking	−0.29	0.12	0.85	0.97	−2.95	−0.37	1.05	2.26	01234
1 Affected how smoothly you walk	−0.44	0.12	0.90	0.89	−3.49	−0.23	0.86	2.85	01234
1 Made you concentrate on your walking*	1.12	0.15	1.37	1.40	−3.19		0.41	2.78	01123

* Recoded items.

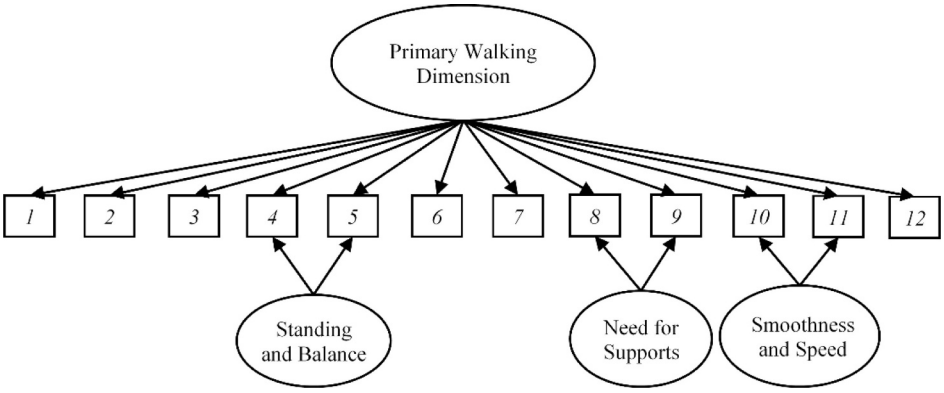

Fig. 1. Diagram for the Rasch testlet model for the MSWS-12.

4. Results

4.1. Preliminary reliability and exploratory factor analysis

The Cronbach's alpha for the 12 items was excellent at 0.97. Exploratory factor analysis (EFA) performed using PCA indicated that the each of the 12 items loaded onto the first component (Range: 0.77–0.92), which explained 76.7 percent of the variance. Eigenvalues for the rest of the extracted components was < 1.00. However, it is worthy to note that items 8 and 9 also showed loadings ≥0.40 on a second component, which explained an additional 5.6 percent of variance.

4.2. Rasch analyses

4.2.1. Functioning of items and rating scale

Rasch analysis highlighted that items number 2, 8 and 9 showed poor fit to the model, i.e., Infit and Outfit statistics sitting outside the acceptable range (0.60–1.40). Moreover, items 2, 8, 9, and 12 showed disordered threshold estimates or overlapping response categories, suggesting the need for a recode of the rating scale. Concerning item 2, 8, and 9 we found most of the patients answered using either the lowest or highest response category ("1–Not at all", "5-Extremely"), while significant disorder and overlap emerged between non-extreme response categories. Concerning item 12, significant overlap emerged between the two lowest response categories (1-Not at all, 2-A little).

4.2.2. Revision of the rating scale

We revise the rating scale of items showing problematic functioning (i.e., item 2, 8, 9, 12) by merging response categories showing significant overlap into a single response category. Table 2 shows the final scoring strategy used in the revised version of the MSWS-12. As regards items 2, 8, and 9, existing problems required a dichotomization of the rating scale, while a 4-point rating scale was required for item 12. After recoding, all items reported Infit and Outfit statistics within the acceptable range, and showed ordered thresholds. We performed following analyses on this revised version of the MSWS-12.

4.2.3. Dimensionality and local dependency analyses

LID analyses performed by examining the residual correlation matrix showed a few items had positive residual correlations > 0.20 (Items 4, 5 = 0.28; Items 8, 9 = 0.47; Items 10, 11 = 0.40), suggesting the presence of LID between these items. In order to significance of the violation of Rasch assumptions due to multidimensionality and LID, we compared the fit of the unidimensional model Rasch and that of an alternative. Fig. 1 depict the diagram of the Rasch testlet model tested to examine LID between items. Based on results of the G^2 LR test, the Rasch testlet model showed a small but significant increase in fit over the unidimensional Rasch model (χ^2 (3) = 13.36, $p < 0.01$). Estimated testlet effects for items assessing smoothness and speed of walking (Item 10, 11), and balance and standing problems (Item 4, 5) were negligible (Range: 0.01–0.03). In turn, testlet effect for items assessing need for supports (Item 8, 9), at 1.47, was substantially higher than the proposed cut-off, supporting the presence of LID between these items.

4.2.4. Measurement invariance

No significant DIF emerged when controlling for gender and age. However, we found significant DIF on item assessing need for support for walking indoor (item 8) and outdoor (item 9) when controlling EDSS, and diagnosis status. Individuals with a EDSS score < 4.5 and a diagnosis of relapsing remitting MS were less likely to report need of support when walking indoors (Item 8) and outdoors (Item 9) due to MS, than those with higher EDSS (Item 8: DIF = -2.02, $p < 0.05$; Item

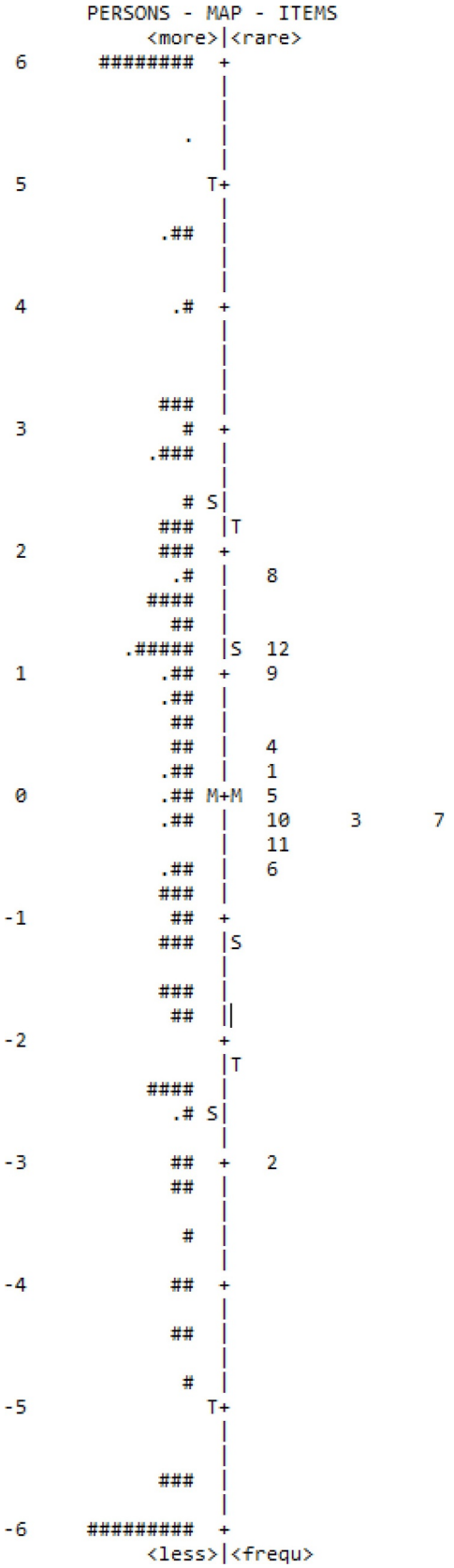

Fig. 2. Item-person map for the revised MSWS-12.

9: DIF = −2.84, $p < 0.05$) and primary or secondary progressive course. (Item 8: DIF = 1.56, $p < 0.05$; Item 9: DIF = 1.24, $p < 0.05$).

4.2.5. Reliability and targeting

Reliability of person measures was excellent, at 0.94. The item reliability index was also excellent, at 0.98. The item-person map (Fig. 2) showed the MSWS-12 had adequate item/person targeting, as we found the mean of the sample measures to coincide with the mean of the item calibrations. At 1.84 logit, Item 8 ("Made it necessary for you to use support when walking indoors") was the most difficult to endorse for patients, while Item 2 ("Limited your ability to run"), at −2.91 logit, was the easiest to endorse. Items did not distribute evenly along the measurement continuum, as all items except for item 2 were located between the average and the upper end of the walking disability continuum. Further, we found several patients without aligned items, as person measures covered a wider range (−6.98–6.67 logit) than item measures (−2.91–1.85 logit). There was no significant (< 15% of criterion) evidence of ceiling (8%, $n = 16$) or floor (9%, $n = 18$) effects. Table 3 shows demographic and clinical characteristics of the two groups of participants reporting the highest or lowest possible scores when compared with participants with non-extreme scores. Patients reporting the lowest possible score where more likely to have relapsing/remitting MS, and lower EDSS scores and disease duration, while participants which achieved the highest possible score were more like to have primary or secondary progressive MS and had higher EDSS scores.

5. Criterion validity

The score for the revised version of the MSWS-12 on a 0–100 scale was obtained using the same scoring rules of the original one (Mean = 49.44, SD = 31.34). We found a strong positive correlation ($r = 0.75$, $p < 0.001$) between this revised MSWS-12 score and the EDSS score, indicating that people reporting lower walking ability on the scale had also a higher level of disability as rated by clinicians. The revised MSWS-12 score did also show a small positive correlation with disease duration ($r = 0.22$, $p < 0.001$). Finally, we found a significant mean difference in score based on disease course, indicating patients with primary or secondary MS had significantly higher MSWS-12 scores than relapsing remitting patients (t (196) = 8.92, $p < 0.001$).

6. Discussion

The main aim of the present study was to perform a thorough examination of the psychometric properties of the Italian adaption of the MSWS-12. In particular, using IRT methods, we performed and investigation aimed at determining item and scale functioning, scale dimensionality, measurement invariance, reliability, and adequacy of targeting of the MSWS-12. Additionally, based on emerging results, we sought to introduce changes to the scale's scoring procedure that would improve its' suitability for clinical use. Finally, we examined criterion validity of the revised scale.

Preliminary analyses with EFA supported the expected unidimensional structure of the scale. However, Rasch analyses indicated a few items showed poor fit to the model. In line with previous findings (Mokkink et al., 2016), items assessing patients' need for supports when walking indoors (Item 8) and outdoors (Item 9), and ability in running (Item 2) showed problematic rating scale functioning. Concerning these items, we found most of the patients answered using extreme response categories, while significant disorder and overlap emerged between non-extreme response categories, suggesting a dichotomization of the rating scale was required to improve functioning of the items. Similarly, Item 12 (Need to concentrate when walking) showed no misfit but problematic overlap between lower response categories. To deal with these issues, we revised the rating scale of problematic items by collapsing adjacent response categories showing problematic functioning

Table 3
Patient characteristics by MSWS-12 score group.

Patient characteristics	Score group Minimum	Non-Extreme	Maximum	test
	n/M(SD)	n/M(SD)	n/M(SD)	
Sex				χ^2 (2) = 0.01, p = 0.99
Male	9	82	9	
Female	9	83	8	
Diagnosis				χ^2 (2) = 29.82, p < 0.001
Relapsing/remitting	18	100	1	
Primary or secondary progressive	0	65	14	
Age	46.56 (14.28)	48.26 (11.7)	48.89 (16.57)	F (2196) = 0.18, p = 0.83
EDSS	1.94 (0.8)*	4.48 (1.63)	6.72 (0.52)*	F (2196) = 42.59, p < 0.001
Disease duration	7.39 (4.84)*	14.04 (8.32)	15.06 (5.87)	F (2196) = 6.07, p = 0.03

* Contrast significant at p < .05 (Reference = Non-Extreme).

(i.e., threshold disorder or overlap).

Overall, the revised version of the MSWS-12 scale showed adequate measurement properties. Still, minor indications of local item dependence emerged. Analyses with the Rasch teslet model indicated items 8 and 9, which assess need for supports when walking indoors and outdoors, represented a small, but non-negligible source of multidimensionality. Examination of measurement invariance also indicated item 8 and 9 showed differential functioning when comparing individuals differing by EDSS score, and disease course. In line with previous findings (Mokkink et al., 2016; Ansari et al., 2015), these results suggest that items assessing need for supports does not fully comply with the requirement of unidimensionality of measurement. Nonetheless, a correlation coefficient close to 1 (result not showed) between scale scores as computed including and excluding problematic items showed the impact of removing these items on scores was negligible. Additionally, due to the brevity of the scale, removing items could compromise construct validity of the MSWS-12. For these reasons, we decided against removing problematic items from the scale.

Next, the revised MSWS-12 showed good reliability and adequate targeting to the recruited sample. However, items did not distribute evenly along the continuum of walking disability, as we found all items except for Item 2 (“Limited your ability to run”) to cluster at the medium-to-high area of the disability continuum. Further, the scale revealed only minor floor and ceiling effects, indicating that the extreme items (for assessing lowest or highest walking impairment) were not able to cover the full range of impairment as observed in the sample, suggesting the need for inclusion of additional items assessing impairment at both the upper and lower end of the continuum.

Lastly, criterion validity of this revised MSWS-12 scores was examined by investigating its association with clinician-rated EDSS score, and results indicated a strong convergence between these two measures. As expected, the MSWS-12 score also positively correlated with disease duration, and was significantly higher among patients with primary or secondary MS than relapsing remitting patients.

Overall, findings from the present study indicate that after a revision of the scoring procedure, the Italian adaptation of the MSWS-12 shows adequate fit to Rasch assumptions and provides a reliable assessment of walking disability for MS patients, in particular in the medium-to-high area of the disability continuum. Still, in line with previous findings (Mokkink et al., 2016; Ansari et al., 2015), items assessing need for supports when walking showed signs multidimensionality, and differential functioning when comparing patients differing in clinician-rated disability and disease course, questioning their inclusion in the scale.

7. Limitations

The present study is not without limitations. First and foremost, at 199 patients, sample size for the study was small, but close to the proposed sample size of 200 required for stable parameter estimates in Rasch analyses (Suen, 1990) (Suen, 1990). At the same time, it is worthy to note that Rasch mean-square Infit and Outfit statistics are expected to provide reliable information even at low sample sizes (N < 50) (Smith et al., 2008).Further, emerging findings are mostly in support of previous findings reported by studies employing larger sample sizes (Cano et al., 2015; Mokkink et al., 2016; Engelhard et al., 2016). Thus, we expect our results concerning the presence of problematic functioning of some items to be confirmed by future studies. Additional, due to the lack of inclusion of alternative assessments of walking ability (e.g., Time 25 foot walk test), in the present study we were not able to perform a thorough investigation of the criterion validity of the scale.

8. Conclusions

In summary, this study indicates that, after introducing minor changes to the scale's scoring procedure, the revised version of the Italian adaptation of the MSWS-12 provides an assessment of walking disability in MS that shows adequate fit to Rash measurement. After recoding items showing problematic functioning (i.e., 2, 8, 9, 12), the scale showed adequate fit to the Rasch model. Still, items assessing need for supports (i.e. items 8 and 9) showed minor signs of multidimensionality, as well as signs of differential functioning when comparing patients with different levels of walking disability and disease course. The scale showed also limited targeting to individuals at a low disability level, as only Item 2 (“Limited your ability to run”) appeared to target this level of disability. As such, present findings are consistent with those by previous studies (Mokkink et al., 2016; Engelhard et al., 2016; Ansari et al., 2015) in suggesting a revision of the actual content of the MSWS-12 might help improve its measurement properties. In spite of the aforementioned problems, the revised version of the MSWS-12 scale shows properties which make it suitable for use in clinical practice, in particular with patients at medium-to-high levels of walking disability.

Role of funding source

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

None

References

Adams, R., Wu, M.L., Haldane, S., Sun, X., 2012. ACER ConQuest (Version 3.0. 1) [*Computer Software*]. Australian Council for Educational Research & University of California, Berkeley.

Ansari, N.N., Naghdi, S., Mohammadi, R., Hasson, S., 2015. Multiple sclerosis walking scale-12, translation, adaptation and validation for the persian language population.

Gait. Posture 41 (2), 420–424.
Bond, T., Fox, C.M., 2015. Applying the Rasch model: Fundamental Measurement in the Human Sciences, 3rd ed. Routledge, New York, NY.
Briggs, D.C., Wilson, M., 2003. An introduction to multidimensional measurements using Rasch models. J. Appl. Meas. 4, 87–100.
Cano, S., Cleanthous, S., Marquis, P., et al., 2015. Measuring the impact of multiple sclerosis: enhancing the performance of the MSIS-29 and MSWS-12. Value Health 18 (7), A709–A710.
Christensen, K.B., Makransky, G., Horton, M., 2017. Critical values for Yen's Q 3: identification of local dependence in the Rasch model using residual correlations. Appl. Psychol. Meas. 41 (3), 178–194.
Compston, A.,., Coles, A, 2008. Multiple sclerosis. Lancet 372 (9648), 1502–1517.
Engelhard, M.M., Schmidt, K.M., Engel, C.E., et al., 2016. The e-MSWS-12: improving the multiple sclerosis walking scale using item response theory. Qual. Life Res. 25 (12), 3221–3230.
Glas, C.A., Wainer, H., Bradlow, E.T., 2000. MML and eap estimation in testlet-based adaptive testing. Computerized Adaptive Testing: Theory and Practice. Springer, Dordrecht, pp. 271–287.
Hobart, J.C., Riazi, A., Lamping, D.L., et al., 2003. Measuring the impact of MS on walking ability the 12-Item MS walking scale (MSWS-12). Neurology 60 (1), 31–36.
Kieseier, B.C.,., Pozzilli, C., 2012. Assessing walking disability in multiple sclerosis. Mult. Scler. 18 (7), 914–924.
Linacre, J.M., 2009. Winsteps (Version 3.68. 0)[Computer software]. Chicago, IL.
Masters, G.N., 1982. A Rasch model for partial credit scoring. Psychometrika 47 (2), 149–174.
McGuigan, C.,., Hutchinson, M., 2004. Confirming the validity and responsiveness of the multiple sclerosis walking scale-12 (MSWS-12). Neurology 62 (11), 2103–2105.
McHorney, C.A., Tarlov, A.R., 1995. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual. Life Res. 4 (4), 293–307.
Mokkink, L.B., Galindo-Garre, F., Uitdehaag, B.M., 2016. Evaluation of the multiple sclerosis walking scale-12 (MSWS-12) in a dutch sample: application of item response theory. Mult. Scler. 22 (14), 1867–1873.
Motl, R.W.,., Snook, E.M., 2008. Confirmation and extension of the validity of the multiple sclerosis walking scale-12 (MSWS-12). J. Neurol. Sci. 268 (1), 69–73.
Polman, C.H., Reingold, S.C., Banwell, B., et al., 2011. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann. Neurol. 69 (2), 292–302.
Smith, A.B., Rush, R., Fallowfield, L.J., Velikova, G., Sharpe, M., 2008. Rasch fit statistics and sample size considerations for polytomous data. BMC Med. Res. Methodol. 8 (1), 33.
Solaro, C., Trabucco, E., Signori, A., et al., 2015. Italian validation of the 12-item multiple sclerosis walking scale. Mult. Scler. Int ID 540828.
SPSS I, 2009. PASW Statistics 18. SPSS Inc, Chicago.
Streiner, D.L., Norman, G.R., Cairney, J., 2015. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford University Press, USA.
Suen, H.K.,., 1990. Principles of Test Theories. Lawrence Erlbaum, Hillsdale, NJ.
Terwee, C.B., Bot, S.D., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., Bouter, L.M., de Vet, H.C., 2007. Quality criteria were proposed for measurement properties of health status questionnaires. J. Clin. Epidemiol. 60 (1), 34–42.
Wang, W.C., Wilson, M.,., 2005. The Rasch testlet model. Appl. Psychol. Meas. 29 (2), 126–149.
Wang, X., Bradlow, E.T., Wainer, H., 2002. A general bayesian model for testlets: theory and applications. Appl. Psychol. Meas. 26 (1), 109–128.
Wright, B., 1994. Reasonable mean-square fit values. Rasch. Meas. Trans. 8, 370.
Zhang, O., Shen, L.,., Cannady, M., 2010. Polytomous IRT or Testlet Model: An Evaluation of Scoring Models in Small Testlet Size Situations. University of Florida, USA Doctoral dissertation.