

German Translation, Cultural Adaption and Validation of the Unidimensional Self-efficacy Scale for Multiple Sclerosis Using Rasch Analysis

Barbara Seebacher (✉ barbara.seebacher@i-med.ac.at)

Clinical Department of Neurology Medical University of Innsbruck, Austria <https://orcid.org/0000-0001-5699-9077>

Roger Mills

Walton Centre NHS Foundation Trust, Department of Neurology, Liverpool

Markus Reindl

Medical University of Innsbruck, Clinical Department of Neurology

Laura Zamarian

Medical University of Innsbruck, Clinical Department of Neurology

Simone Kircher

University Hospital Innsbruck, Clinical Department of Neurology

Christian Brenneis

Clinic for Rehabilitation Münster, Department of Neurology and Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster

Rainer Ehling

Clinic for Rehabilitation Münster, Department of Neurology and Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster

Florian Deisenhammer

Medical University of Innsbruck, Clinical Department of Neurology

Research

Keywords: Multiple Sclerosis, Self-Efficacy, Patient Reported Outcome Measures, Austria, Cross-Cultural Comparison, Validation Studies as Topic

Posted Date: August 18th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-59390/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

Background

Self-efficacy refers to people's beliefs in their ability to perform relevant activities to achieve personal goals. In people with multiple sclerosis (PwMS), self-efficacy has been shown to significantly impact health-related behaviour. So far, a validated German language self-efficacy scale for PwMS is missing. Therefore, the aims his study were to translate the Unidimensional Self-Efficacy Scale for Multiple Sclerosis (USE-MS) into German, establish face and content validity and cultural adaption of the German version for PwMS in Austria. Further aims were to validate the German USE-MS (USE-MS-G) in PwMS using Rasch analysis.

Methods

Formal permission to translate and validate the USE-MS was received from the scale developers. Following guidelines for translation and validation of questionnaires and applying Bandura's concept of self-efficacy, the USE-MS was forward-backward translated, and content and face validity established. Cultural adaption for Austria was performed using cognitive interviews of 30 PwMS. The validation by Rasch analysis, of the final USE-MS-G involved 309 PwMS with minimal to severe disability using scales for resilience, general self-efficacy, anxiety and depression, MS fatigue and health-related quality of life. Re-test was conducted within 14–21 days after the initial test. Data were also pooled with an historic English dataset.

Results

Using a bi-factor solution, excellent external and construct validity, internal consistency, person separation reliability, test-retest reliability and fit to the Rasch model was demonstrated. The USE-MS-G was shown to be unidimensional, well-targeted and free from differential item functioning. Pooling of the English and German datasets confirmed equivalence of the two language versions.

Conclusion

The USE-MS-G is a robust, valid and reliable scale to assess self-efficacy in PwMS. Rasch model fit given, an interval scale transformation table is available for clinical settings and research.

Trial registration:

ISRCTN Registry; ISRCTN14843579; prospectively registered on 02. 01. 2019;
<http://www.isrctn.com/ISRCTN14843579>

Background

Multiple Sclerosis is a chronic demyelinating disease of the central nervous system, with accumulating disability and loss in quality of life (1). It appears crucial for people with MS (PwMS) to preserve their autonomy, despite functional limitations and an unpredictable disease course. Bandura's social cognitive theory proposes constructs such as people's personal agency, self-reflection and self-efficacy (2). Self-efficacy concerns the belief about one's capability to perform relevant tasks, overcome challenges and achieve meaningful goals. Importantly, self-efficacy is not related to people's level of functioning or their skills but rather their judgement of what they can achieve (2). This implies that PwMS who are confident in their ability to master challenges and reach their goals may cope with the disease more effectively. Higher levels of self-efficacy may enhance people's motivation to be physically active as there is a strong relationship with health promoting behaviour and perceived quality of life in PwMS (3).

Several scales have been developed to assess self-efficacy in PwMS (4–6). The Unidimensional Self-Efficacy Scale for Multiple Sclerosis (USE-MS) (5) was developed from the Liverpool Self-efficacy Scale (LSES) (6) and the Multiple Sclerosis Self-Efficacy Scale (MSSS) (4), both of which were based on patients' in-depth interviews and underpinned by Bandura's theory of self-efficacy. Only the USE-MS met the stringent criteria of Rasch analysis (7) for assessing its psychometric properties in a large sample of PwMS. Accordingly, the USE-MS is a valid and reliable instrument for use in clinical practice and research. So far, however, no validated German language version of the USE-MS is available. Therefore, the main purpose of this study was to translate the USE-MS into German and validate the German language version (USE-MS-G) using Rasch analysis. Further purpose was to examine any differential item functioning, including that for language to equate the original English USE-MS and German USE-MS-G.

Methods

Study design and participants

This was a bi-centre prospective cross-sectional translation and validation study with repeated measures, consisting of Phase 1 and Phase 2. In Phase 1, the forward-backward translation, establishment of face and content validity and cross-cultural adaption of the pre-final USE-MS-G was performed. In Phase 2, examination of construct validity and reliability of the USE-MS-G was done. The study was conducted at the outpatient MS-Clinic of the Clinical Department of Neurology, Medical University of Innsbruck, Austria and Department of Neurology, Clinic for Rehabilitation Münster, Austria from 12.2.2019 to 15.06.2020. Ethical approval was received from the Ethics Committee of the Medical University of Innsbruck (reference number EK1260/2018; 13.12.2018).

Information brochures and invitations for study participation were displayed in the MS-Clinic, the Clinic for Rehabilitation, the Austrian MS Society patient magazine and on their website; they were also forwarded to MS support groups. Upon agreement, severely disabled PwMS (Expanded Disability Status Scale (EDSS)(8) ≥ 8) were visited at home to facilitate their participation. Additionally, during their regular

visits, PwMS were notified about the study by Clinic staff. All procedures followed the tenets of the Declaration of Helsinki and written informed consent was obtained from all participants. Research data are available on reasonable request (barbara.seebacher@i-med.ac.at).

A random cross-sectional cohort of patients with clinically definite MS according to the McDonald's criteria (9) version valid at the time of diagnosis and with any MS phenotype was recruited into this study. PwMS of any ethnicity with very good German language skills, aged ≥ 18 years with different levels of functioning were included (EDSS scores from 0 (no disability) to 9.0 (severe disability); see (10) for a detailed study protocol.

Exclusion criteria were comorbidities potentially affecting subjective self-efficacy ratings (e.g., malignant diseases, other neurological or psychiatric disorders), a relapse of MS within 2 months prior to the study or any change in medication within 4 weeks of the study commencement. A relapse between test and retest required the exclusion of the participant.

The Austrian dataset was also pooled with a dataset from the UK development sample to test for invariance by language.

Outcome measures

Demographic (gender, age) and disease specific data (disease duration, MS phenotype, disease-modifying treatment) were retrieved from patients' files. The current EDSS was assessed by neurologists. Questionnaire data were collected twice within a 14–21-day period (test, retest).

The original USE-MS has been shown to be reliable and valid for assessing self-efficacy in PwMS (5). Scoring is achieved by summing up all 12 items while items 5, 7–9 and 11 are reversed scored. The USE-MS involves a 4-point Likert scale (0 = strongly disagree to 3 = strongly agree). A higher summary score signifies stronger self-efficacy beliefs in people.

Validated questionnaires used to evaluate external construct validity of the USE-MS-G were recommended by governmental or patient organisations (11, 12). These included the German language versions of the General Self-Efficacy Scale (GSE) (13), Resilience Scale (RS-13) (14), Multiple Sclerosis International Quality of Life (MusiQoL) (15), Hospital Anxiety Depression Scale (HADS) (16, 17) and Neurological Fatigue Index Multiple Sclerosis (NFI-MS) (18). Scoring and psychometric properties of these scales are described in detail in the study protocol (10).

Sample size

Phase 1

Patients were recruited until saturation was achieved, indicating that no further information could be obtained from additional interviews.

Phase 2

For Rasch analysis, a sample size of 243 participants has been shown to provide accurate estimates of item and person locations irrespective of the scale targeting (19). Moreover, with polytomous items, ≥ 10 observations per category are recommended (20). It is further relevant to collect a wide range of responses across the latent trait under consideration, i.e., self-efficacy (19).

Translation, face and content validity and cultural adaption

In Phase 1, following guidelines for the cross-cultural adaptation of patient-reported outcomes (21, 22) and its enhanced version from the University of Leeds, UK, a forward-backward translation process was conducted by 6 bilingual translators, 3 native in German, 3 in English. This comprised a synthesis of translations and expert committee consensus. Pretesting (Test 1, T1) and face-to-face cognitive interviews regarding the questionnaire wording were carried out in male and female PwMS across the disability range. After 30 recorded interviews, saturation was achieved. Cross-cultural equivalence between the USE-MS and USE-MS-G was accomplished in the semantic, idiomatic, experiential and conceptual areas (21, 22). Qualitative content analysis of the verbatim interview transcriptions was performed (described in detail elsewhere (10)). During all stages of the iterative adaption process of the USE-MS-G, consensus was reached with the original scale developers (5).

Statistical analyses

External validity and test-retest reliability

Correlational analyses were performed between the USE-MS-G and other measures to determine convergent and discriminant construct validity. We hypothesised moderate to high positive correlations of the USE-MS-G with the GSE, RS-13 and MusiQol and moderate to high negative correlations with the HADS and NFI-MS. Spearman's Rank correlation coefficients of 0.3–0.49 were considered low, 0.5–0.69 moderate and ≥ 0.7 strong (23); they were calculated with their 95% confidence intervals (CI) and p-values corrected for multiple comparisons using a Bonferroni correction. Descriptive statistics and external validity estimates were performed IBM SPSS software (IBM SPSS Statistics; Version 26.0. Armonk, NY: IBM Corp.) or GraphPad Prism Version 8 (GraphPad Software, La Jolla, CA). Statistical significance was defined as two-tailed p-value < 0.05 .

Test–retest reliability was determined using Lin's concordance correlation coefficient (24) (r_c) between Test 2 (T2) and Test 3 (T3). The r_c (0–1; 95% CI) was used to estimate the amount of agreement between the test and retest USE-MS-G data. The Pearson correlation coefficient was calculated as a measure of precision and a Bias correction factor, C_b as a measure of accuracy (24). MedCalc software (<https://www.medcalc.org/>) was used to determine the r_c .

Internal validity: Rasch analysis

Rasch analysis uses the mathematical Rasch model to assess whether a summary score for a scale can be calculated with confidence (7). Internal construct validity of the USE-MS-G was determined by examining the deviations from model expectations, i.e. the way in which persons are expected to interact

with test items to produce linear measurement (7). The model expects that the probability of a person providing a certain answer to an item is a logistic function of the difference between the person 'ability' (perceived self-efficacy) and the item 'difficulty'. This is checked visually by inspection of item characteristic curves and numerically by the analysis of variance (ANOVA) fit statistics (uniform DIF; non-uniform DIF (25)). The USE-MS-G contains 4 response categories and hence, the polytomous Rasch model was chosen for the current study (26).

Using different chi square (χ^2) fit statistics, USE-MS-G data were tested against the model expectations of unidimensionality. That is, the 'ability' and 'difficulty' are required to relate to the same construct of self-efficacy (described in detail elsewhere (26)). Perfect values for the different fit statistics and unidimensionality are provided in Table 3. Using a residual item correlation matrix between all items the expectation of local independence was examined. Item residuals represent the difference between an item's expected and observed values, divided by its standard deviation for standardisation. Residual correlations of + 0.2 above the mean correlation of the total matrix indicate local dependence (27). This denotes a confounding factor inducing an association between items, or multidimensionality (28, 29). In the presence of item-dependency, two "super-items" can be created from alternative items and compared with each other running a robust conditional chi-square test of fit (28). The proportion of common to total variance retained in a bi-factor equivalent solution corresponds to the explained common variance (ECV) (30). For a unidimensional scale, the ECV should be > 0.9, indicating that > 90% of the variance is common and retained in the latent estimate (28).

Table 3
Model Fit of the USE-MS-G to the Rasch model

Analysis	Item residual		Person residual		(Cond.) ¹ Chi-Square ²	PSI ³	Alpha	Unidimensionality ⁴		
	Mean	SD	Mean	SD	Value (df)	p		% tests > 5%	95% CI	
12-item scale	0.05	2.19	-0.31	1.33	138.46 (48)	0.000	0.86	0.87	23.0	5-9.9
2 super-items	0.15	0.88	-0.59	1.01	29.24 (22)	0.138	0.85	0.86	4.2	1.8–6.6
Ideal values	0.00	1.00	0.00	1.00		> 0.01*	> 0.70	> 0.70	< 5.0	LCI < 5
Abbreviations: T1/2: testlet, or super-item 1/2; Cond.: conditional; df: degrees of freedom; PSI: person separation index; Alpha: Cronbach's alpha; SD: standard deviation; CI: confidence interval; LCI: lower bound of the 95% CI										
¹ Conditional Chi-Square: only applicable for the super-item solution; for the item-based solution, the Chi Square is shown										
² Chi-Square of T1: 3,392 (4), p = 0,494; Chi-Square of T2: 1,054 (4), p = 0,901. Perfect values are > 0.004* (Bonferroni adjusted)										
³ The PSI indicates the reliability and differentiation of strata										
⁴ Based on independent t-tests to compare person residuals which are positively and negatively loading on the first principal component										
*Bonferroni adjusted and variable with number of items										

The property of invariance means that all participants recognise the difficulty in identical items regardless of their self-efficacy (31). If certain groups of participants respond differently to items, e.g. males and females, the assumption of invariance is violated, called differential item functioning (DIF) (31). The USE-MS-G and original USE-MS data (N = 485) were pooled and tested for invariance by language (English; German) to equate the language versions. Every item was examined for absence of DIF by gender (male; female), age (quartile groups), disease duration (quartile groups), timepoint (test; retest) and centre (Innsbruck; Münster). Bonferroni adjustment was performed wherever appropriate for the number of tests undertaken. Reliability was evaluated by the Person Separation Index (PSI, range 0–1) and Cronbach's alpha, which should be ≥ 0.85 for individual use or 0.70 for group use (32). Scale precision was examined by the standard error of measurement (SEM) and minimum detectable change (MDC) based on a 95% CI (33). Rasch Analysis was conducted with RUMM2030 software (<http://www.rummlab.com.au/>) which was based upon the unrestricted or partial credit model (34).

Results

Phase 1

The prefinal USE-MS-G resulted from the forward-backward translation procedure. The 30 interviewees (23 females) were 26–75 years old and scored 0–8.0 on the EDSS (median 3.5). Their mean disease duration was 20.3 (standard deviation, SD 11.7) years. Cognitive debriefing analysis showed that item phrasing 2, 3, 5, 8 and 12 needed to be reworded. Discrepancies were greatest with items 3 and 5, which were finally solved after a second round of interviews and consensus of the expert committee with the original test developers.

Phase 2: Demographics, test-retest reliability and external validity

In total, 623 eligible patients were informed about the study, 309 agreed to participate (a response rate of 49.6%) of whom 290 completed the questionnaires twice (93.8%). Two participants had a relapse from MS, 11 participants could not be reached again, and 6 participants reported poor health. Population characteristics are presented in Table 1 (for disease modifying treatment, DMT definitions see (10)).

Table 1
Population characteristics of the validation sample

Parameters	N = 309
Gender ¹	72 (23.3) males
	237 (76.7) females
Age ²	50.2 ± 11.8 years
MS phenotype ¹	194 (62.8) relapsing-remitting
	36 (11.7) primary progressive
	79 (25.6) secondary progressive
Disease duration ²	18.21 ± 10.80 years
EDSS ³	3.0 (0–9)
EDSS groups ¹	205 (66.3) with an EDSS of 0–4.0
	86 (27.8) with an EDSS of 4.5–6.5
	18 (5.8) with an EDSS of 7.0–9.0
Disease modifying treatment (DMT) ¹	163 (53) patients received no DMT
	61 (19.7) received lowly effective DMT ⁴
	85 (27.5) received highly effective DMT ⁵
Abbreviations: EDSS: Expanded Disability Status Scale	
¹ Data are shown as count (percentage)	
² Data are presented as mean ± standard deviation	
³ Data are shown as median (range)	
⁴ low effective DMTs: interferon-b 1a and 1b, pegylated interferon-b 1a, glatiramer acetate, dimethyl fumarate, teriflunomide, azathioprin, intravenous immunoglobulins	
⁵ high effective DMTs: alemtuzumab, cladribine, fingolimod, natalizumab, ocrelizumab, cyclophosphamide, mitoxantrone, rituximab	

Regarding external construct validity, Spearman's Rank correlations (95% CI) of the USE-MG-G with the MusiQoL, RS-13, GSE, HADS and NFI-MS are presented in Table 2. Test-retest reliability was at least moderate, as indicated by an r_c of 0.92 (95% CI 0.89 to 0.93), a Pearson phi coefficient (ρ) of 0.92 and a Bias correction factor C_b of 0.99.

Table 2
External validity measures: correlations with other scales

USE-MS-G		
Musi-QoL	ADL	0.723 (0.664 to 0.774)***
	PWB	0.642 (0.569 to 0.705)***
	SPT	0.535 (0.447 to 0.612)***
	RFriends	0.410 (0.310 to 0.502)***
	RFamily	0.303 (0.195 to 0.403)***
	SSL	0.311 (0.202 to 0.413)***
	Cop	0.557 (0.473 to 0.632)***
	Reject	0.616 (0.539 to 0.682)***
	RHealth	0.437 (0.339 to 0.525)***
	Index ¹	0.771 (0.719 to 0.815)***
RS-13	Total	0.607 (0.530 to 0.675)***
GSE	Total	0.537 (0.450 to 0.614)***
HADS	Anx	-0.451 (-0.538 to -0.354)***
	Depr	-0.751 (-0.797 to -0.696)***
	Total	-0.679 (-0.737 to -0.612)***
NFI-MS	Phys	-0.708 (-0.761 to -0.646)***
	Cogn	-0.519 (-0.598 to -0.429)***
	Sum	-0.682 (-0.739 to -0.615)***
Abbreviations: Musi-QoL: Multiple Sclerosis International Quality of Life questionnaire; ADL: activities of daily living; PWB: psychological wellbeing; SPT: symptoms; RFriends: relationships with friends; RFamily: relationships with family; SSL: sentimental and sexual life; Cop: coping; Reject: rejection; RHealth: relationship with healthcare system; RS-13: Resilience Scale-13; GSE: General Self-Efficacy Scale; HADS: Hospital Anxiety Depression Scale; Anx: anxiety; Depr: depression; NFI-MS: Neurological Fatigue Index for MS		
¹ The MusiQoL index is computed only if all dimension scores are computed (non-missing)		
Spearman's' correlation coefficients are shown with 95% confidence intervals; ***correlation is significant at the < 0.001 level (2-tailed, p-values corrected for 18 comparisons)		

Phase 2: Rasch analysis

USE-MS-G test data and pooled UK and Austrian data were fitted to the Rasch model separately. The original USE-MS-G data showed misfit to the Rasch model, multidimensionality and DIF by age, gender, disease duration and centre. Absence of item independency given, the data were combined into two super-items (testlets) consisting of all alternative items and fit to the bi-factor model was tested. The class interval structure of the current 2 testlets was equal. The results demonstrated a latent correlation between the two item sets of -0.976. The ECV was 0.988, which confirms that a single score driven from the testlet items truly indicates self-efficacy of the respondents. Fit to the Rasch model, excellent reliability and unidimensionality were shown, as presented in detail in Table 3 and Fig. 1. The fit residuals were well below the limits of ± 2.5 , which indicates that there was a small calculated difference between the expected and observed values for persons and items.

Inspection of the item characteristic curves showed classic fit to the Rasch model for both subtests (super-items). All observed scores for the five class intervals at the different 'ability' levels of self-efficacy matched the expected scores of the model (see Additional File 1). Examination of the item characteristics curves for DIF by gender (male; female), age groups (18–40; 41–51; 52–58; 59+ years), disease duration groups (0–10; 11–16; 17–25; 26+ years), timepoint (test; retest) and centre (Innsbruck; Münster) confirmed absence of any DIF (see Additional File 2). All ANOVA results for DIF by person factors (uniform DIF) and for person factor by class interval interaction (non-uniform DIF) were statistically non-significant. The USE-MS-G and original USE-MS data (N = 485) were also pooled and tested for invariance by language (English; German) to equate the language versions. With the pooled dataset, there was also no DIF by language or centre in any of the analyses. Figure 2 presents the item characteristics curves with DIF analyses for the person factor language.

Perfect threshold ordering was displayed by the threshold map, with all scoring categories progressing in a logical order, as well as the category probability curves of the testlets (Fig. 3). Zero % of floor and 1% of ceiling effects (3/309 persons) were observed. The targeting of the scale was good, showing a slightly higher level of self-efficacy than the mean of the scale, which itself showed a near ideal distribution across the trait (Fig. 4). A standard error of measurement (SEM) of 0.439 was found, a minimum detectable change (MDC) of 4.56 points, as measured on the original scale range of 0–36 points. The smallest amount of change beyond the measurement error, expressed as MDC percentage, would be 12.7%. Given fit to the Rasch model, a transformation of the raw score to interval scaling is available (Table 4).

Table 4
Transformation of raw score to interval scale latent estimate for the
USE-MS-G

Raw score	Interval estimate	Raw score	Interval estimate
0	0.00	19	18.88
1	2.55	20	19.81
2	4.20	21	20.73
3	5.28	22	21.63
4	6.14	23	22.51
5	6.88	24	23.36
6	7.59	25	24.19
7	8.29	26	24.99
8	9.02	27	25.75
9	9.79	28	26.49
10	10.60	29	27.22
11	11.45	30	27.94
12	12.34	31	28.69
13	13.25	32	29.50
14	14.18	33	30.43
15	15.12	34	31.61
16	16.06	35	33.37
17	17.00	36	36.00
18	17.95		
The transformation remains valid provided there are no missing data			
Additional Files			
Additional file 1 Item characteristic curves			
Additional file 2 Differential item functioning analyses			

Discussion

The purpose of this study was to translate, cross-culturally adapt the prefinal German USE-MS-G to Austria and validate the final USE-MS-G in PwMS across a wide range of disability. Forward-backward

translation and pretesting according to guidelines are critical procedures of a scale's cross-cultural adaptation, as it is expected to reflect the latent trait under investigation (22). Therefore, PwMS with different levels of disability were selected for Phase 1 of the study to gain insight into their perceptions of the prefinal USE-MS-G wording. However, even if great care is taken with this process, psychometric properties comparable to those of the original scale are not guaranteed. Therefore, fit of the German USE-MS to the Rasch model was examined in a well-powered sample of PwMS, representative of the MS population at large (35). Additionally, traditional correlational analyses of the USE-MS-G with relevant scales were performed. Results showed moderate to strong positive correlations with the RS-13, GSE and MusiQoL and hence corroborated the hypotheses of this study. The moderate relationship with the generic GSE stressed the importance of a disease-specific German language self-efficacy scale for PwMS. Symptoms of a chronic disease like MS suggest that affected people may hold beliefs regarding their ability to perform relevant tasks and reach desired goals diverging from those of healthy people. Further analyses showed moderate to strong negative correlations with the HADS and the NFI-MS, confirming the discriminant construct being assessed. Excellent reliability was shown by a Cronbach's alpha of 0.86 and PSI of 0.85, indicating that ≥ 3 self-efficacy levels can be differentiated by the USE-MS-G (34).

The study population covered the full range of MS disability levels and phenotypes. Fit to the Rasch model was shown for the USE-MS-G, good targeting and unidimensionality. This justifies the use of a summary score and transformation of the ordinal raw score into an interval score, which is suitable for parametric analysis. Absence of any DIF was observed and invariance by language for the pooled dataset. This means that the USE-MS-G is equivalent to the English original version and appropriate to measure self-efficacy in people with mild to severe MS.

The scale failed to show model fit at the item-based level. There was some individual item misfit with some over and under discriminating items. Items 5, 7, 8 and 12 showed DIF by language, and for 5 and 12 the English items were harder than German, but for items 7 and 8 the opposite was the case. So, at the level of the test, the DIF cancelled out and this was confirmed by the absence of DIF using the super-item, or bi-factor, solution. Further, low levels of item dependency were observed that advocated employing a bifactor approach. The bifactor solution with alternative items allocated to the testlets mirrors the scale's use as a domain score in clinical practice. Excellent fit to the model was shown, and merely 0.0050% of the variance was lost (1 minus the A value in the summary statistics window).

Good precision of the USE-MS-G was demonstrated, as expressed by a low SEM. Based on that, the minimum detectable difference was 12.7% i.e., change scores of less than 4.6 points are less than the measurement error in the scale. No (0%) floor effects and 1% ceiling effects were seen, with only 3 people scoring at the top end of the scale. This indicates that the USE-MS-G is able to discriminate all levels of self-efficacy in PwMS and sensitive to changes exceeding 4.6 points also at both ends of the spectrum.

To our knowledge, the English USE-MS is the most rigorously developed and tested scale for assessing self-efficacy in PwMS and accordingly, it was chosen for this study. The validity and reliability of the

German USE-MS-G has been demonstrated in this study and the scale is available for use in clinical and research practice. Assessing self-efficacy may be useful to enable an individualised and comprehensive treatment in PwMS. The USE-MS-G is easy to use and can be completed by PwMS within 5–10 minutes. It can be accessed free of charge from tonic.measures@gmail.com.

Conclusions

To conclude, the USE-MS-G is a robust, valid and reliable scale to assess self-efficacy in PwMS. The translation and cross-cultural adaption to Austria were performed according to international guidelines (21) and with cognitive interviews of patients as integral parts to ensure their comprehensibility of the item wording. Fit to the Rasch model provided, in further studies the interval scale transformation table can be used with confidence.

Abbreviations

ANOVA

Analysis of variance; Cb: Bias correcting factor; χ^2 : Chi square test; CI: confidence interval; DIF: Differential item functioning; DMT: Disease modifying treatment; ECV: Explained common variance; EDSS: Expanded Disability Status Scale; GSE: General Self-Efficacy Scale; HADS: Hospital Anxiety Depression Scale; LSES: Liverpool Self-efficacy Scale; MDC: Minimum detectable change; MSSS: Multiple Sclerosis Self-Efficacy Scale; MusiQoL: Multiple Sclerosis International Quality of Life; NFI-MS: Neurological Fatigue Index Multiple Sclerosis; p: Pearson phi coefficient; PSI: Person Separation Index; PwMS: People with multiple sclerosis; r_c : Lin's concordance correlation coefficient; RS-13: Resilience Scale; SD: Standard deviation; SEM: Standard error of measurement; T1: Test 1 (Phase 1); T2: Test 2 (rest in Phase 2); T3: Test 3 (retest in Phase 2); USE-MS: Unidimensional Self-efficacy Scale for Multiple Sclerosis; USE-MS-G: German version of the Unidimensional Self-efficacy Scale for Multiple Sclerosis

Declarations

Acknowledgements

We would like to warmly acknowledge and thank all the participants of this study and the self-help group leaders who helped to inform their members about this study.

Ethical approval and consent to participate

Ethical approval was received from the Ethics Committee of the Medical University of Innsbruck (reference number EK1260/2018; 13.12.2018).

Information brochures and invitations for study participation were displayed in the MS-Clinic, the Clinic for Rehabilitation, the Austrian MS Society patient magazine and on their website; they were also forwarded to MS support groups. Upon agreement, severely disabled PwMS (Expanded Disability Status

Scale (EDSS) ≥ 8) were visited at home to facilitate their participation. Additionally, during their regular visits, PwMS were notified about the study by Clinic staff. All procedures followed the tenets of the Declaration of Helsinki and written informed consent was obtained from all participants.

Consent for publication

Not applicable.

Availability of data and materials

All relevant data are shown in the manuscript and its additional files as figures or tables. All data described in the manuscript, including all relevant raw data, are freely available to reviewers and any scientist wishing to use them for non-commercial purposes, without breaching participant confidentiality on reasonable request (barbara.seebacher@i-med.ac.at).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Austrian MS Research Society [no grant number].

Authors' contributions

BS designed the study, collected the data, analysed the data, interpreted the data and wrote the manuscript. RJ analysed the data, provided data from the original English dataset, interpreted the data and edited the manuscript. RE, LZ and SK collected the data and helped to draft the manuscript. FD, MR and CB interpreted the data and edited the manuscript. All authors read and approved the final manuscript.

References

1. Compston A, Confavreux C, Lassmann H, McDonald I, Miller D, Noseworthy J, et al. *McAlpine's multiple sclerosis*. 4th ed. ed. London: Elsevier; 2006.
2. Bandura A. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs: Prentice-Hall; 1986.
3. Finn A. The relationship between self-efficacy and health promoting behaviors, illness severity, mental health, and perceived quality of life in multiple sclerosis patients. *Modern Psychological Studies*. 2006;11(2, Article 3).
4. Rigby SA, Domenech C, Thornton EW, Tedman S, Young CA. Development and validation of a self-efficacy measure for people with multiple sclerosis: the Multiple Sclerosis Self-efficacy Scale. *Mult Scler*. 2003;9(1):73–81.

5. Young CA, Mills RJ, Woolmore J, Hawkins CP, Tennant A. The unidimensional self-efficacy scale for MS (USE-MS): developing a patient based and patient reported outcome. *Mult Scler*. 2012;18(9):1326–33.
6. Airlie J, Baker GA, Smith SJ, Young CA. Measuring the impact of multiple sclerosis on psychosocial functioning: the development of a new self-efficacy scale. *Clinical rehabilitation*. 2001;15(3):259–65.
7. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press; 1980.
8. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an Expanded Disability Status Scale (EDSS). *Neurology*. 1983;33(11):1444–52.
9. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol*. 2001;50(1):121–7.
10. Seebacher B, Mills RJ, Reindl M, Zamarian L, Kuisma R, Kircher S, et al. German translation, cultural adaption and validation of the unidimensional self-efficacy scale for multiple sclerosis: a study protocol. *BMJ Open*. 2019;9(8):e029565.
11. Potter K, Cohen ET, Allen DD, Bennett SE, Brandfass KG, Widener GL, et al. Outcome measures for individuals with multiple sclerosis: recommendations from the American Physical Therapy Association neurology section task force. *Phys Ther*. 2014;94(5):593–608.
12. Surveillance of Multiple sclerosis in adults: management (NICE guideline CG186) London, UK: National Institute for Health and Care Excellence (NICE); 2018 [Available from: <https://www.nice.org.uk/guidance/CG186/documents/surveillance-review-proposal>].
13. Schwarzer R, Jerusalem M. Generalized Self-Efficacy scale. In: Weinman J, Wright S, Johnston M, editors. *Measures in health psychology: A user's portfolio Causal and control beliefs*. Windsor: NFER-NELSON; 1995. pp. 35–7.
14. Leppert K, Koch B, Brähler E, Strauß B. Die Resilienzskala (RS) – Überprüfung der Langform RS-25 und einer Kurzform RS-13. *Klinische Diagnostik Evaluation*. 2008;2:226–43.
15. Simeoni M, Auquier P, Fernandez O, Flachenecker P, Stecchi S, Constantinescu C, et al. Validation of the Multiple Sclerosis International Quality of Life questionnaire. *Mult Scler*. 2008;14(2):219–30.
16. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta psychiatrica Scandinavica*. 1983;67(6):361–70.
17. Petermann F. Hospital Anxiety and Depression Scale, Deutsche Version (HADS-D). *Zeitschrift für Psychiatrie Psychologie Psychotherapie*. 2011;59(3):251–3.
18. Mills RJ, Young CA, Pallant JF, Tennant A. Rasch analysis of the Modified Fatigue Impact Scale (MFIS) in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2010;81(9):1049–51.
19. Linacre JM. Sample size and item calibration stability. *Rasch Measurement Transactions*. 1994;7(4):328.

20. Linacre J. Understanding Rasch measurement: Optimizing rating scale category effectiveness. *J Appl Meas.* 2002;3:85–106.
21. Beaton D, Bombardier C, Guillemin F, Ferraz M. Recommendations for the Cross-Cultural Adaptation of the DASH & QuickDASH Outcome Measures. Toronto: Institute for Work and Health; 2007.
22. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol.* 1993;46(12):1417–32.
23. Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences.* 5th ed. ed. Boston: Houghton Mifflin; 2003.
24. Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.* 1989;45(1):255–68.
25. Andrich D. *Rasch Models for Measurement.* Beverly Hills: SAGE; 1988.
26. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis rheumatism.* 2007;57(8):1358–62.
27. Lazarsfeld PF, Henry NW. *Latent structure analysis.* Boston: Houghton Mifflin; 1968.
28. Pomeroy IM, Tennant A, Mills RJ, Young CA, Group TOS. The WHOQOL-BREF: a modern psychometric evaluation of its internal construct validity in people with multiple sclerosis. *Qual Life Res.* 2020;29(7):1961–72.
29. Andrich D. A latent trait model for items with response dependencies: Implications for test construction and analysis. In: Embretson SE, editor. *Test design.* New York: Academic Press; 1985. pp. 245–75.
30. Andrich D. Components of variance of scales with a bifactor subscale structure from two calculations of α . *Educational Measurement: Issues Practice.* 2016;35(4):25–30.
31. Hagquist C, Andrich D. Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health Qual Life Outcomes.* 2017;15(1):181.
32. Mills RJ, Young CA, Pallant JF, Tennant A. Development of a patient reported outcome scale for fatigue in multiple sclerosis: The Neurological Fatigue Index (NFI-MS). *Health Qual Life Outcomes.* 2010;8:22.
33. van Kampen DA, Willems WJ, van Beers LWAH, Castelein RM, Scholtes VAB, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). *J Orthop Surg Res.* 2013;8(1):40.
34. Wright BD, Masters GN. *Rating scale analysis.* Chicago: MESA Press; 1982.
35. Pugliatti M, Rosati G, Carton H, Riise T, Drulovic J, Vecsei L, et al. The epidemiology of multiple sclerosis in Europe. *European Journal Of Neurology.* 2006;13(7):700–22.

Figures

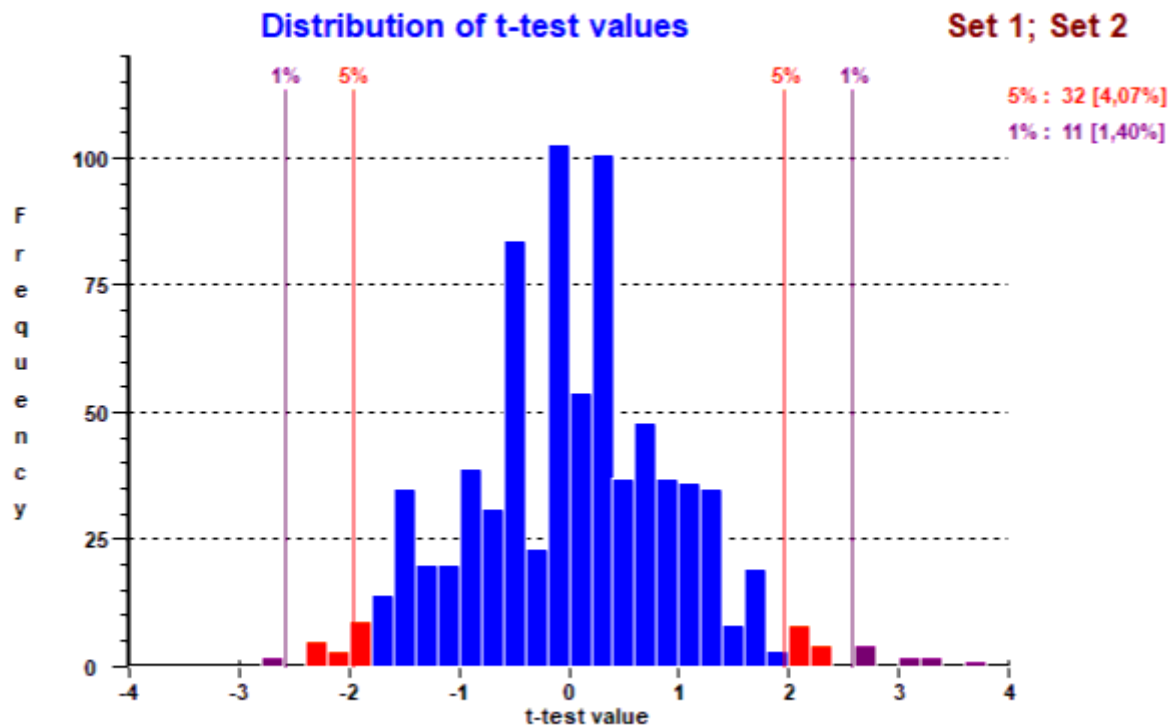
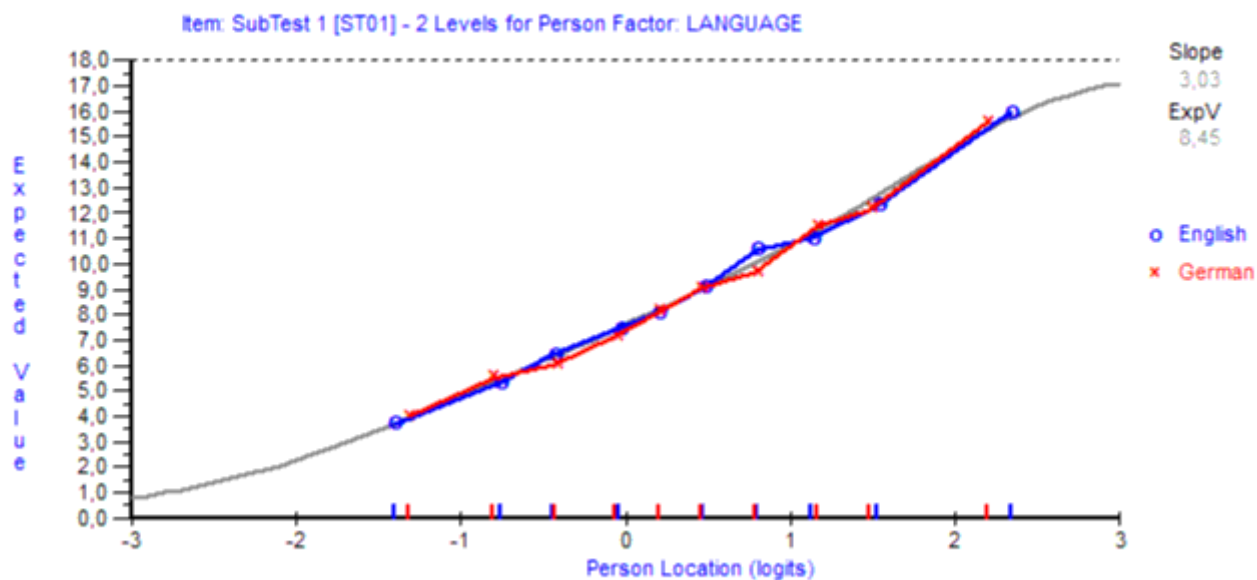
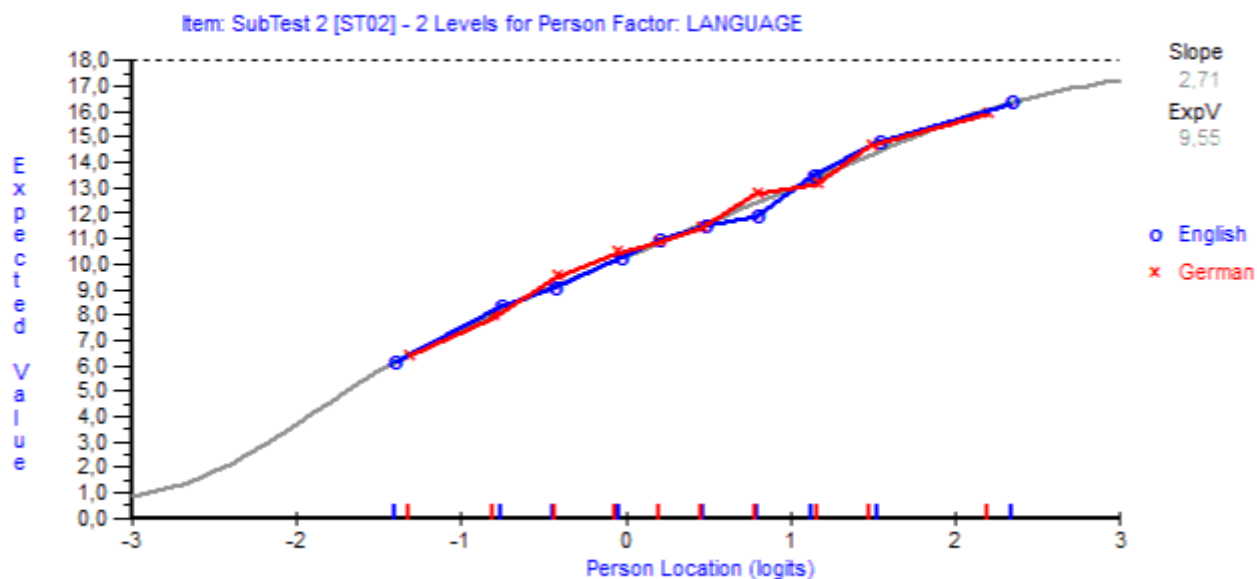


Figure 1

Equating t-test distribution plot Figure 1 shows that a majority of the respondents scored similarly on the two subsets of items (person residuals which are positively and negatively loading on the first principal component). On the 5% level, 32 persons (4.07%) scored significantly different and on the 1% level, 11 persons (1.40%) showed significantly different responses.

A**B****Figure 2**

DIF by language analyses results for the pooled UK and Austrian datasets ASubtest 1 works the same way for English and German speaking participants (no DIF by language) BSubtest 2 works the same way for English and German speaking participants (no DIF by language)

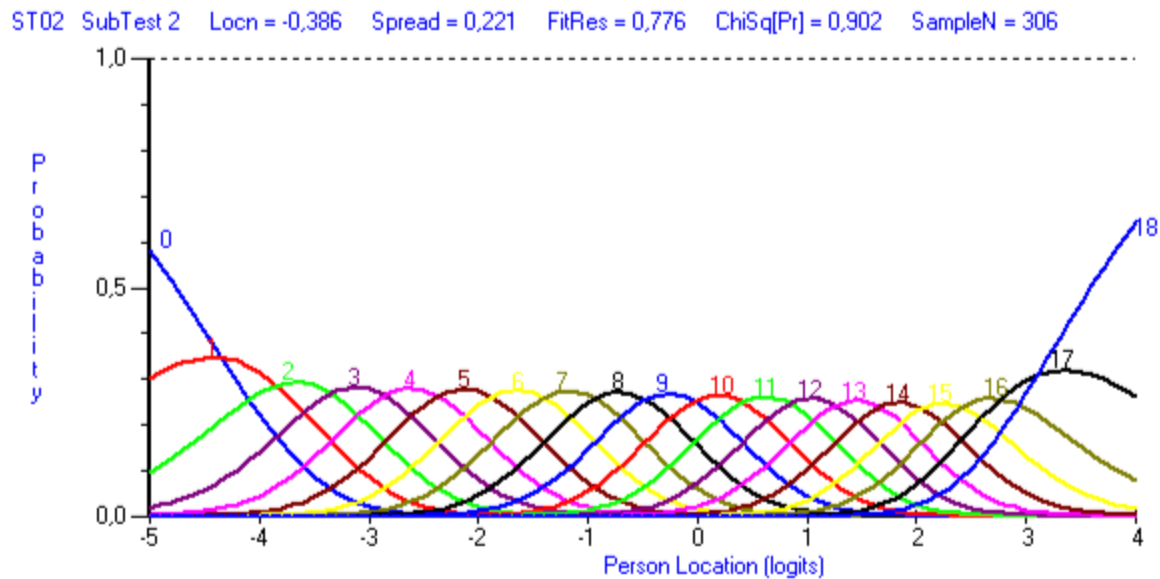


Figure 3

Category probability curve for subtest 2 The category probability curves for subtest2 show 19 ordered categories, where higher numbers represent higher levels of self-efficacy.

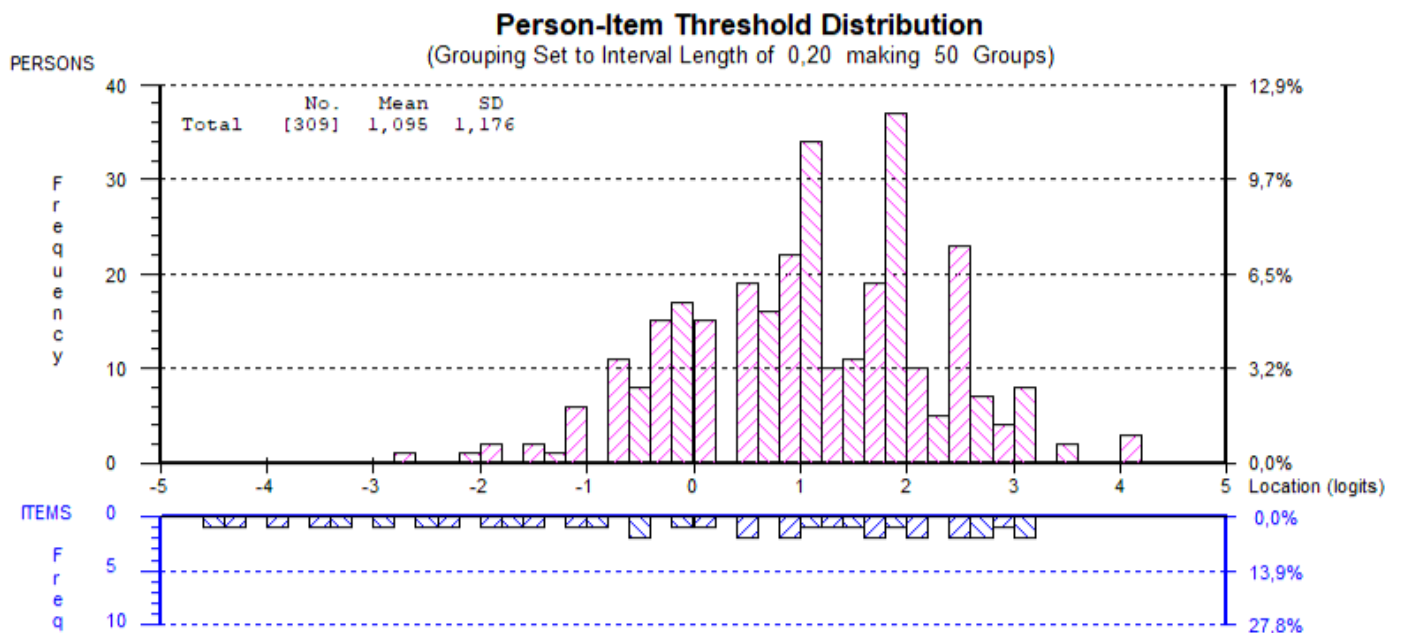


Figure 4

Person-item threshold distribution map Targeting of the scale is shown by the person-item threshold distribution map. Histograms on top of the plot represent the self-efficacy levels of the persons whereas those at the bottom show the difficulty of the items. On the x-axis, a logit scale is shown where the mean difficulty is set to 0 by default. On the y-axis of the top histogram, the probability of endorsing a standardised score is presented. The y-axis of the bottom histogram displays the probability of responding to a given score for a particular item.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile213082020.docx](#)
- [AdditionalFile113082020.docx](#)