BMJ Open

German translation, cultural adaption and validation of the unidimensional self-efficacy scale for multiple sclerosis: a study protocol

Barbara Seebacher,[1] Roger J Mills,[2] Markus Reindl,[1] Laura Zamarian,[1] Raija Kuisma,[3] Simone Kircher,[4] Christian Brenneis,[5,6] Rainer Ehling,[5,6] Florian Deisenhammer[1]

For numbered affiliations see end of article.

Correspondence to
Dr Barbara Seebacher;
barbara.seebacher@i-med.ac.at

ABSTRACT

Introduction Self-efficacy refers to individuals' confidence in their ability to perform relevant tasks to accomplish desired goals. This is independent of their actual abilities. In people with multiple sclerosis (MS), self-efficacy has been shown to powerfully influence motivation and health-related behaviour, such as adherence to prescribed treatment or physical activity. So far, a rigorously tested German language self-efficacy questionnaire for people with MS is missing.

Methods The purpose of this study is to translate the original Unidimensional Self-Efficacy Scale for Multiple Sclerosis (USE-MS) into German and to validate the German USE-MS (USE-MS-G). Based on Bandura's concept of self-efficacy and international guidelines for questionnaire development, the patient-led development of the pre-final German version will involve a forward–backward translation process, synthesis of translations, expert committee review and consensus with the original test developers. At two centres in Tyrol, Austria, content and face validity and cultural adaption for Austria will be established using face-to-face semistructured cognitive interviews of 30 people with MS (PwMS). A further 292 PwMS with minimal to severe disability will be tested at two timepoints to validate the USE-MS-G.

Results Mixed methods analyses will be applied. Interviews will be transcribed and analysed employing qualitative content analysis. External validity will be explored using Spearman's Rank correlation coefficients of the USE-MS-G with the 13-item Resilience Scale, General Self-Efficacy Scale, Multiple Sclerosis International Quality of Life questionnaire, Hospital Anxiety and Depression Scale and MS-specific Neurological Fatigue Index. Test–retest reliability, internal consistency and floor and ceiling effects will be evaluated. Internal validity will be examined using Rasch analysis.

Ethics and dissemination Ethical approval was received from the Ethics Committee of the Medical University of Innsbruck, Austria (reference number EK1260/2018; 13.12.2018). Results from this study will be disseminated to the participants and MS Societies, and to clinicians and researchers through peer-reviewed publications and conferences.

Strengths and limitations of this study

► This study protocol describes the German translation of the original English language Unidimensional Self-efficacy Scale for Multiple Sclerosis (USE-MS), on permission of the scale developers and applying international recommendations.

► Consistent with the conceptual framework of the English USE-MS, Bandura's concept of self-efficacy will be adhered to.

► Employing a patient-led process in phase 1, 30 people with MS (PwMS) will be interviewed about the pre-final German USE-MS, to establish face and content validity and cultural adaption for PwMS in Austria.

► In phase 2, the German USE-MS will be validated in a larger sample of 292 PwMS.

► Applying classical test theory and Rasch analysis approaches, internal and external validity, internal consistency and test–retest reliability will be explored.

Study registration ISRCTN Registry; trial ID ISRCTN14843579; prospectively registered on 02. 01. 2019; http://www.isrctn.com/ISRCTN14843579

INTRODUCTION

Multiple sclerosis (MS) is one of the most common neurological diseases in young adults worldwide, with increasing prevalence.[1] MS is characterised by a wide variety of symptoms and different disease courses.[2] Despite the development of novel disease modifying drugs and neurorehabilitation strategies, the unpredictability of the disease with psychological distress, losses in social contact and quality of life (QoL) are concerning for people with MS (PwMS). However, individuals' self-knowledge can modulate their approach to day-to-day activities. According to Bandura's social cognitive

theory, psychosocial functioning is regulated by reciprocal interactions between behaviour, personal factors and environmental conditions.[3] Self-regulation and intrinsic motivation enable individuals to set and pursue their own goals, observe and evaluate themselves in relation to attained goals.[4] Bandura defined self-efficacy as individuals' beliefs regarding their capability to perform significant tasks, to achieve goals that are meaningful for their daily lives.[3] Self-efficacy beliefs considerably influence people's feelings, thoughts and motivation[5] while, notably, being independent of their physical performance.[5] Such a concept appears important for people with disabilities because it may shape their motivation to initiate and adhere to treatment, particularly when facing side effects.

Perceived self-efficacy influences health-related behaviour such as adhering to medication[6] or engaging in physical activity in PwMS.[7] Health status evaluations of responses to rehabilitation and steroid treatment after an MS relapse can be predicted by self-efficacy levels.[8] Also, higher self-efficacy levels are associated with better long-term perceived cognitive functioning[9] and QoL.[10 11] PwMS who report higher perceived self-efficacy also state lower levels of fatigue, depression and anxiety.[12] Recent evidence has provided insight into the importance of self-management and intrinsic motivation for motor learning.[13] Recognising the relevance of self-efficacy especially for people with disabilities, valid and reliable measurement tools are still needed for its assessment. Three generic self-efficacy scales were found in the literature.[7 14–16] However, generic questionnaires may not adequately cover the construct of self-efficacy in a chronic neurological disease like MS. The initial impact of a diagnosis of MS, in addition to the manifold symptoms and necessity of managing a progressive disease may affect individuals' self-efficacy perceptions. Studies demonstrated that the capability to effectively solve problems, consistent with higher self-efficacy levels, is strongly associated with PwMS' psychological adaptation to their disability,[17] supporting the choice of a disease-specific over a generic self-efficacy questionnaire. MS-specific self-efficacy scales include the Liverpool Self-efficacy Scale (LSES),[18] Multiple Sclerosis Self-Efficacy Scale (MSSS),[19] MS Self-Efficacy Scale,[20] Unidimensional Self-Efficacy Scale for Multiple Sclerosis (USE-MS)[21] and University of Washington Self-Efficacy Scale for people with disabilities.[22]

Following current guidelines, patients should be involved in the translation and development process of disease-specific questionnaires, to ensure the scale reflects their experiences.[23] LSES and MSSS development used in-depth patient interviews while the USE-MS consists of items from both the LSES and MSSS. Bandura's concept of self-efficacy is reflected in the wording of all three questionnaires. The USE-MS study sample was the largest thereof (n=303), and only the USE-MS was exposed to Rasch[24 25] analysis assessing internal construct validity, in addition to conventional external construct

validity and reliability testing. Fit to the Rasch model was demonstrated, and good external validity and reliability.[21] Consequently, the USE-MS appears to be appropriate for use in clinical practice and research. However, so far no validated German language version of the USE-MS is available. The purpose of this study will therefore be to translate the USE-MS into German and validate the German language version in a larger sample of PwMS.

METHODS
Study aims
The first aim of this patient-led study is to translate the original English USE-MS, developed by Young et al (2012) into German, based on international guidelines.

The second aim is to establish face and content validity and cultural adaption of the German version for PwMS in Austria, using individual semistructured cognitive interviews.

The third aim is to evaluate internal and external validity, internal consistency and test–retest reliability of the German USE-MS (USE-MS-G), using classical test theory and Rasch analysis .

Study design
This will be a bi-centre prospective cross-sectional translation and validation study with repeated measures, consisting of phase 1 and phase 2. The StandardProtocol Items: Recommendations for Interventional Trials (SPIRIT) 2013 and SPIRIT-PRO Extension checklist for study protocols[26] is presented in online supplementary file 1.

Study setting and timeline
Locations will be the outpatient MS-Clinic of the Clinical Department of Neurology, Medical University of Innsbruck, Austria and Department of Neurology, Clinic for Rehabilitation Münster, Austria.

The expected overall study duration is 33 months, from February 1, 2019 to October 31, 2021.

Participants and recruitment
A random cross-sectional cohort of patients with clinically definite MS will be recruited from the two centres. Adult (≥18 years) people of any ethnicity and with any MS phenotype according to the McDonald's criteria[27–29] version valid at the time of diagnosis will be included in the study. Their disability status score on the Expanded Disability Status Scale (EDSS)[30] may range from 0 (no disability) to 9.0 (severe disability). Patients will be included if they are able to speak and understand German language. Exclusion criteria are concomitant diseases which may affect subjective self-efficacy ratings (eg, malignant diseases, other neurological or psychiatric disorders), a relapse of MS within the last 2 months or any medication change within 4 weeks prior to the study. A relapse between testing 2 and 3 would necessitate the exclusion of the participant.

The study will be advertised in the MS-Clinic, the Rehabilitation Centre and on the Austrian MS Society website. Further interested PwMS will be examined for eligibility by neurologists at the two study locations. Severely disabled PwMS (EDSS ≥8) will be offered home visits to enable their participation. Written informed consent will be obtained by the first author (BS) who is not involved in the treatment of the patients. Participants may withdraw from the study at any time and for any reasons without prejudice. Outpatient participants will be reimbursed for travel expenses only.

Patient and public involvement

In phase 1, patients will be lay members of the expert committee to consolidate all the translations and back translations of the USE-MS. Their role regarding the item and response option wording and sentence structure will be crucial, as the final questionnaire should be understood by PwMS. Patients will also be involved using face-to-face cognitive interviewing, to gain insight into their views about the clarity of the wording, meaning and completeness of the questions of the pre-final USE-MS-G. The Austrian MS (recruitment) and MS Research Societies (funding) will be involved in this study, with whom the findings will be shared as soon as available (patient magazine, meetings). The findings will also be disseminated to the UK MS Society and MS Trust.

Sample size
Phase 1

Patients will be recruited until saturation is achieved. Saturation is a standard term in qualitative methodology to signify the point when the analysis of data from new participants reveals no further emergent qualitative themes. Saturation is typically achieved after 10–30 people have been interviewed but is determined by the nature of the analysis and the participants themselves.[31]

Phase 2

Rasch analysis sample size requirements are predicated on the degree of precision required for estimating item and person difficulties. Regardless of targeting, one can be 99% confident that a sample size of 243 participants is adequately large to obtain a (high) precision of ±0.5 log odd units (logits). Good targeting provided, a sample size of 108 people would be sufficient.[21 32] Using the formula $n=n/(1-(z/100))$ where n is the calculated number of participants and z the expected attrition rate of 15%–20%, a total sample size of 286–304 participants will be aimed at in this study.

Outcomes and data collection

Assessments used in this study were developed using patient involvement and/or recommended by governmental or patient organisations (online supplementary file 2). Study outcomes and methods for their assessment are presented in figure 1. Participant characteristics and assessments used at all timepoints are shown in table 1.

At recruitment, disability will be assessed by neurologists (FD, CB or RE) using the EDSS, ranging from 0 to 10, with higher scores representing higher levels of disability.[30] Although psychometric validation studies criticised its low responsiveness to changes, the EDSS has no floor or ceiling effects,[33] has been shown to be valid and reliable[34] and is therefore recommended for use in clinical studies.[35]

Excellent internal and external validity and reliability of the original USE-MS has been shown.[21] Scoring of the USE-MS draws results from all 12 items while items 5, 7, 8, 9 and 11 are reversed scored. Higher numbers represent stronger self-efficacy beliefs in participants.[21] The USE-MS includes a 4-point Likert scale (0=strongly disagree to 3=strongly agree).

To assess external construct validity, the following questionnaires will be administered:

The validated German version[36] of the 10-item General Self-Efficacy Scale (GSE)[15] is a self-administered four-point Likert scale with a summary score ranging from 'not at all true' to 'exactly true'. The total GSE score ranges between 10 and 40, higher scores signifying greater self-efficacy. Psychometric testing demonstrated high internal consistency, moderate concurrent validity and unidimensionality.[15]

The validated German version[37] of the 13-item Resilience Scale (RS-13),[38] based on the 25-item RS[39] will be used. RS-13 item scores from a seven-point Likert scale are added up, indicating low (13–66 points), moderate (67–72 points) or high (73–91 points) resilience.[38] The German RS-13 showed high internal consistency and moderate test–retest reliability. Confirmatory factor analysis indicated an acceptable model fit.[38]

The validated German version[40] of the 31-item Multiple Sclerosis International Quality of Life (MusiQol) questionnaire[41] will be employed. Response options use a 6-point Likert scale, from 1='never/not at all' to 5='always/very much' and 6='not applicable'. Negatively worded item scores are reversed, and for each participant mean scores for each dimension of the item scores are calculated. All nine dimension scores are linearly transformed to a 0–100 scale, their mean representing the global index score, 0 indicating the worst level of health-related QoL and 100 the best. Psychometric testing showed satisfactory internal and external validity and acceptable reliability for all MusiQol dimensions.[41]

The validated German version[42] of the 14-item Hospital Anxiety Depression Scale (HADS)[43] will be used. The HADS is a self-report questionnaire with a four-point Likert scale and a 42-point maximum, higher scores representing higher levels of anxiety or depression. Items 2, 4, 7, 9, 12 and 14 are reversed scored, odd items are added to score the anxiety subscale (0–21 points) and even items are added to generate the depression subscale (0–21 points). Testing of the German version demonstrated good internal consistency and acceptable test–retest reliability.[42] The two-factor structure of the scale was confirmed.[42]
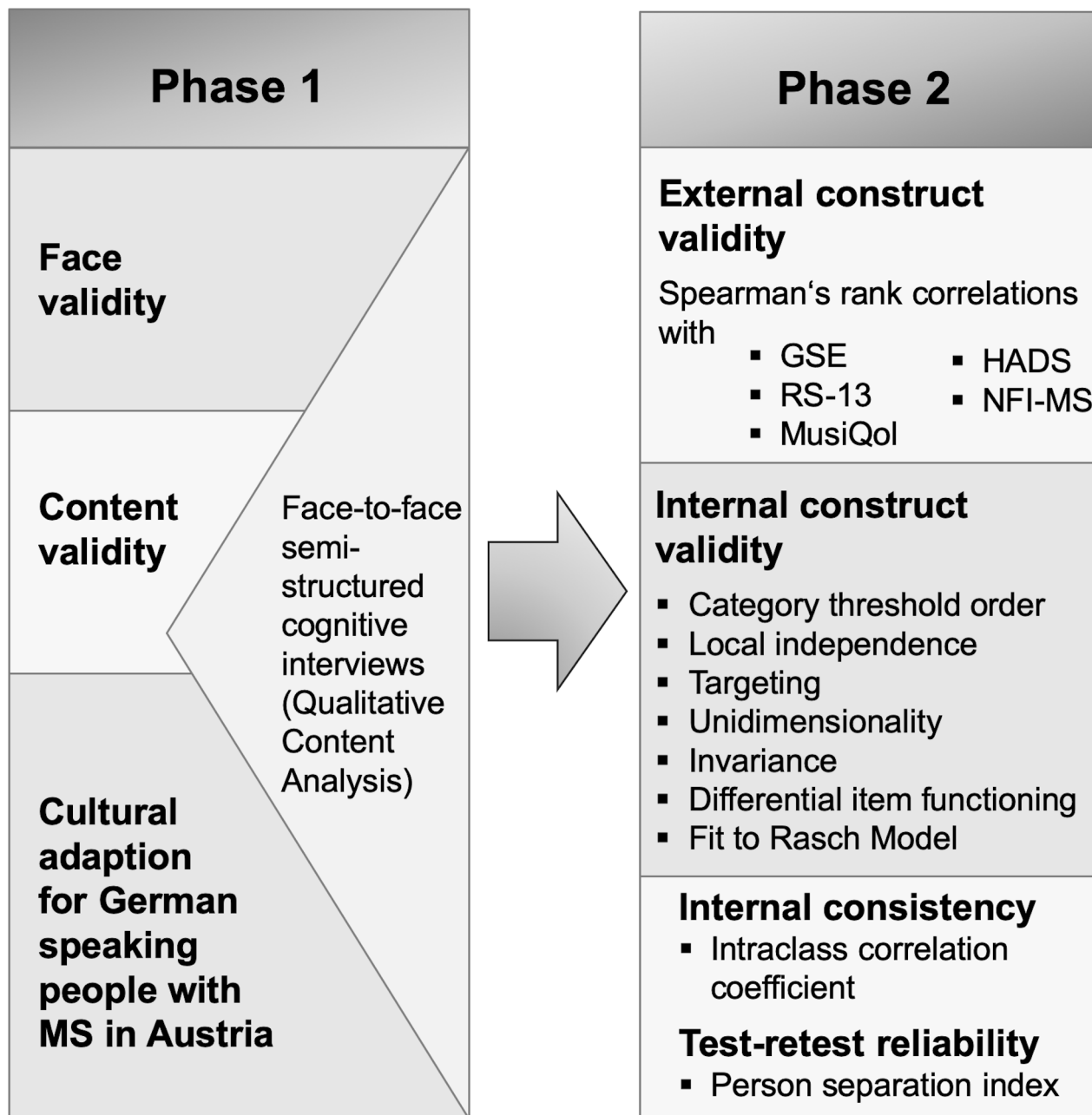
Figure 1 Study outcomes and their assessment. GSE, General Self-Efficacy Scale; HADS, Hospital Anxiety and Depression Scale; MusiQol, Multiple Sclerosis International Quality of Life questionnaire; NFI-MS, Neurological Fatigue Index; RS-13, Resilience Scale, short version.

The validated German version[44] of the 23-item Neurological Fatigue Index (NFI-MS) will be used.[45] Four factors of the NFI-MS were confirmed by principal component analysis (PCA) and explained 62% of the variance. The four subscales and total scale showed acceptable responsiveness,[46] good test–retest reliability, moderate convergent validity and fit to Rasch model expectations.[45] Items are scored on a 4-point Likert scale from 0='strongly disagree' to 3='strongly agree'. For scoring, the following item values are added: 1–8='physical subscale'; 9–12='cognitive subscale'; 13–18='relief by diurnal sleep or rest subscale'; 19–23='abnormal nocturnal sleep and sleepiness subscale'; and 1–7, 9, 11–12 ='physical and cognitive summary score'.[45]

Assessments will be performed by trained physiotherapists holding a Master's (SK) and PhD degree (BS) and a clinical neuropsychologist (LZ). The number of participants who decline to participate or drop out will be recorded, together with reasons (ConsolidatedStandards of Reporting Trials flow chart). Any health problems will be recorded.

Phase 1: data will be collected at one timepoint (Testing 1, T1), with an expected duration of 45-60 min.

Phase 2: for the test–retest reliability assessment, data will be collected at two time-points and will last 60–90 min: Testing 2 (T2) and Testing 3 (T3), 14–21 days after T2.[45 47]

Table 1 Participant characteristics and assessments used in this study

Participant characteristics and assessments (assessments will be collected in a random order to avoid order effect)	Phase 1	Phase 2	
	T1	T2	T3
Participant identifier (ID)	X	X	X
Age	X	X	
Gender	X	X	
MS phenotype*	X	X	
Disease duration	X	X	
EDSS†	X	X	
DMT‡	X	X	X
(Pre-final) German version of Unidimensional Self-Efficacy Scale for Multiple Sclerosis	X	X	X
Qualitative cognitive interview	X		
Resilience Scale, short version		X	X
General Self-Efficacy Scale		X	X
Multiple Sclerosis International Quality of Life questionnaire		X	X
Hospital Anxiety and Depression Scale		X	X
Neurological Fatigue Index		X	X

*Relapsing-remitting; primary progressive; secondary progressive multiple sclerosis.[95]
†EDSS groups: 0–4.0; 4.5–6.5; 7.0–7.5; 8.0–9.0.[30]
‡(a) No DMTs; (b) low effective DMTs: interferon-b 1a and 1b, pegylated interferon-b 1a, glatiramer acetate, dimethyl fumarate, teriflunomide, azathioprin, intravenous immunoglobulins; (c) high effective DMTs: alemtuzumab, cladribine, fingolimod, natalizumab, ocrelizumab, cyclophosphamide, mitoxantrone, rituximab.[96 97]
DMT, disease modifying treatment; EDSS, Expanded Disability Status Scale.

Data management

With regard to confidentiality, the Austrian and Tyrolean Data Protection Acts will be adhered to. Double data entry and range checks for data values will be used. For qualitative content analysis (QCA), double coding of the data set will be performed. Only the research team will have access to the data. All personal data will be codified by a participant ID. Data and files will be saved on a password protected computer, will not be transferred via emails and will be only used for the purposes for which they were collected. Participants will be informed about their right to disclosure for their own data even if these data lack clinical utility. Codified data will be kept for 15 years following completion of the study. Blank data collection forms can be requested from the corresponding author.

Study procedures

This study will follow the Beaton *et al* guidelines for the cross-cultural adaptation of patient-reported outcomes[48] and its enhanced version from the University of Leeds, UK.

Phase 1

Stage 1: Forward translation of the items, response options, instructions and scoring information into German will be performed by three independent translators; translator 1 is a medical professional and informed about self-efficacy, while translators 2 and 3 have no medical knowledge and are 'naïve' to self-efficacy. Translators are bilingual German native speakers and will create a written report for all translations (T1, T2 and T3), which will then be compared, to distinguish any wording differences or ambiguities.[49]

Stage 2: will be a synthesis of T1-3 into T-123. Involving a fourth, unbiased person, the three versions will be discussed with the translators and any discrepancies solved by consensus. A revised questionnaire and comprehensive report will be produced.[48]

Stage 3: Backward translation of T-123 into English will be done by three bilingual English native speakers who are blind to the original version. Translators are 'naïve' to self-efficacy and medicine, to minimise bias.[49] Vague wording, obvious inconsistencies or theoretical errors in the translations shall be detected. A report for each version, TB1, TB2 and TB3, will be written by the translators. To maximise comprehension, language will be used which can be understood by a 12-year old,[50 51] indicated by a Flesch reading-ease score of 80–90.[52] The German Flesch value=$180 - ASL - (58,5*ASW)$, where ASL=average sentence length and ASW=average number of syllables per word.[52]

Stage 4: Considering written documentations, an Expert Committee will review and integrate all versions of the questionnaire, involving instructions and scoring documentation, and develop the pre-final version of the USE-MS-G. The Expert Committee will consist of three neurologists, two physiotherapists, a neuropsychologist, a methodologist, two language professionals, the translators, three lay PwMS and the translation synthesis recorder. The Expert Committee will be in close contact with the original USE-MS developers. A written report of the consensus process will be created. Decision-making will be based on guidelines to accomplish cross-cultural equivalence between the original and German versions in four areas,[49] shown in figure 2.

Stage 5: Pretesting of the pre-final USE-MS-G will be performed in 30 PwMS, involving completion of the scale and face-to-face cognitive interviews. Cognitive
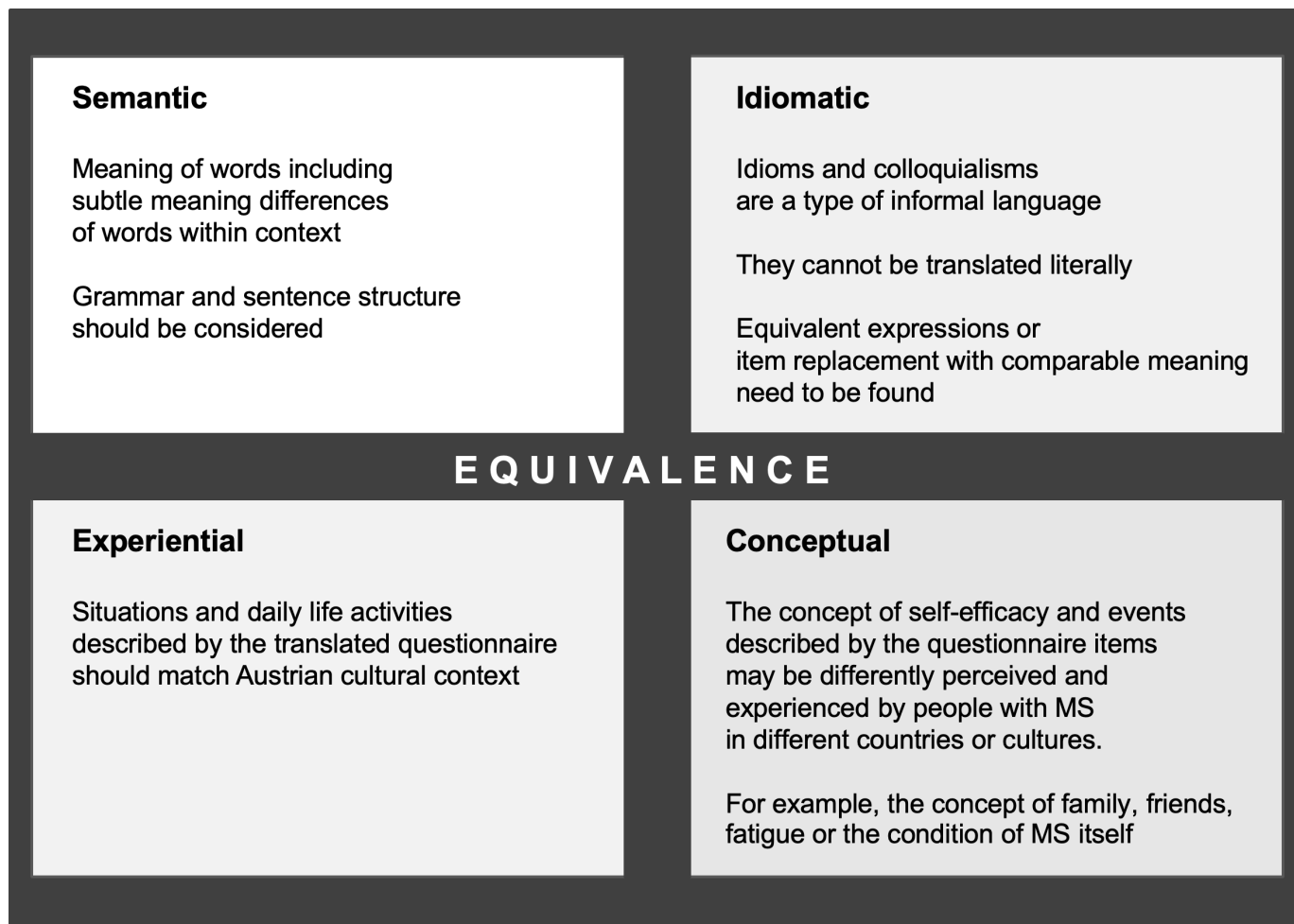
Figure 2 Cross-cultural equivalence areas to be achieved between original and German Unidimensional Self-Efficacy Scale for Multiple Sclerosis; adapted from Guillemin et al.[49] MS, multiple sclerosis.

interviewing will be used to evaluate whether survey questions are easily comprehended, response categories match natural responses, and if people are motivated to respond truthfully and accurately.[53–55] Leading questions will be avoided to minimise bias. Enquiries for comprehension and meaning will be used, and repetition of content by patients.[53 55] Probing will be applied to explore cognitive processes such as memory, underlying reasons for certain responses and overall level of difficulty or confidence.[54] Verbal probes, following Willis' model, will be used immediately after the questions[56]: (a) standardised, anticipated probes: scripted; (b) standardised, conditional probes: scripted, but will be used only if activated by certain participant behaviours such as hesitation[57]; (c) non-standardised, spontaneous probes: flexible, at researcher's digression and (d) non-standardised, emergent probes: applied in reaction to participant behaviour.[58] The interview guide is presented in table 2. Recording and field notes will be used, reviewed for inconsistencies or gaps shortly before the end of the interview.

An overview of study procedures is presented in .figure 3

Phase 2
The USE-MS-G will be validated in a larger sample of 292 PwMS who will complete the above described questionnaires at T1 and T2.

Data analyses
Mixed methods data analyses will be used.

Phase 1: qualitative analyses
Interviews will be transcribed and analysed using QCA[59 60] using QDA MINER LITE software (Provalis Research, Montreal, Canada) and adhering to the Consolidated Criteria for Reporting Qualitative Research.[61] Analysis steps will be performed as follows[62–65]:
► Data organisation based on the research question.
► Identification of recurring ideas, concepts, themes and words.
► Development of a coding frame (requirements: unidimensionality, mutual exclusivity of subcategories within dimensions, exhaustiveness of subcategories and saturation, where each subcategory is used at least once).

Table 2 Questions used for semistructured interview

Participants will be given sufficient time to complete the pre-final German USE-MS.

1.	Having read the questions in the questionnaire, what are your thoughts about them?
2.	Would you please repeat this question in your own words?
3.	What do you think this question is asking?
4.	What do you think about that particular question?
5.	What do you think about the wording of this question, in terms of its clarity?
6.	How easy or hard was this to answer?
7.	How sure are you of your answer?
8.	Could you talk me through your answers in more detail?
9.	What were you thinking of when you answered this question?
10.	Do you have any other comments?
11.	If responses from participants are somewhat unclear, the interviewer asks, 'Why so?'
12.	Should a participant hesitate, the interviewer conveys, 'You spent some time answering that question—what were you thinking about?'

Adapted from Reducing Survey Error through Research on the Cognitive and Decision Processes in Surveys[54] and Olsson *et al*.[98]
USE-MS, Unidimensional Self-Efficacy Scale for Multiple Sclerosis.

► Selection of relevant material, structuring, marking and segmentation of text sections, based on Bandura's concept of self-efficacy and the original USE-MS, to identify main and subcategories.
► Definition, naming and characterisation of categories and decision rules, to enable consistent assignment of data segments.
► Illustration of categories and subcategories using citations.
► Creation of a data matrix, followed by quantitative data analysis (descriptive statistics, eg, frequencies).
► Report.
► Rigour and credibility will be maximised by Cypress,[66] Noble[67] and Smith and Validity.[68]
► Systematic and consistent approach throughout the analysis.
► Revision and expansion of the coding frame.
► Double coding of the whole dataset by two independent researchers (10–14 days after initial coding).
► Checking for researcher effects (reflexivity).[69]

Phase 2: quantitative analyses

Descriptive statistics and reliability estimates will be performed using IBM SPSS software, release V.25.0 (IBM Corporation, Armonk, NY, USA). Rasch Analysis will be conducted with RUMM2030 software.[70] Statistical significance is defined as two-tailed p value<0.05.

Missing data will be treated as follows:

1. Missing data should be avoided by checking questionnaires for missing item responses and asking participants for completion.
2. Rasch analysis calculates an estimate from all available data and does not require a complete data set.[71]

Test–retest reliability

Test–retest reliability will be evaluated using Lin's concordance correlation coefficient (r_c) between T2 and T3 (0–1).[72 73] R_cs will be calculated with their 95% CIs. Values of <0 will be considered to indicate poor, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial and 0.81–1 almost perfect agreement.[74] The data will be racked for the analysis of the concordance correlation coefficient, and stacked for differential item functioning (DIF) by timepoint.

External validity

It is hypothesised that scores on the German USE-MS will demonstrate moderate to high positive correlations with scales assessing conceptually similar constructs (convergent validity; with the GSE, RS-13 and MusiQol) and moderate to high negative correlations with scales measuring divergent constructs (divergent validity; with the HADS and NFI-MS); Spearman's Rank correlation coefficients of 0.3–0.49 being considered low, 0.5–0.69 moderate and ≥0.7 strong.[75]

Internal validity: Rasch analysis

Rasch analysis[25] assumes the probability of a person endorsing an item is a logistic function of the difference between the 'person ability' (perceived self-efficacy) and the 'item difficulty' (level of self-efficacy) expressed.[24] Item characteristic curves, arranged on the log-odds units (logit) scale, will be used to visualise the probability of a person's correct response in relation to the item difficulty.[76]

The polytomous Rasch model will be chosen for this study, suitable for scales with multiple response categories for their items.[77] A significant likelihood ratio test signifying inconsistent distance between response category thresholds would require the use of Masters' unrestricted (partial credit) model,[78] otherwise Andrich's rating scale model.[79] Category thresholds are located centrally between two adjacent categories where either response is equally likely.[77 80] The four-point USE-MS includes three thresholds.

Ordered item category thresholds

Category probability curves will be inspected, checking regular distribution and monotonic advance of measures across categories.[80]

Targeting

Targeting refers to the degree to which the scale captures the full range of self-efficacy. Inspecting person-item threshold distribution maps, the mean location score for the respondents will be compared with the default items zero value. A well-targeted scale is centred around zero
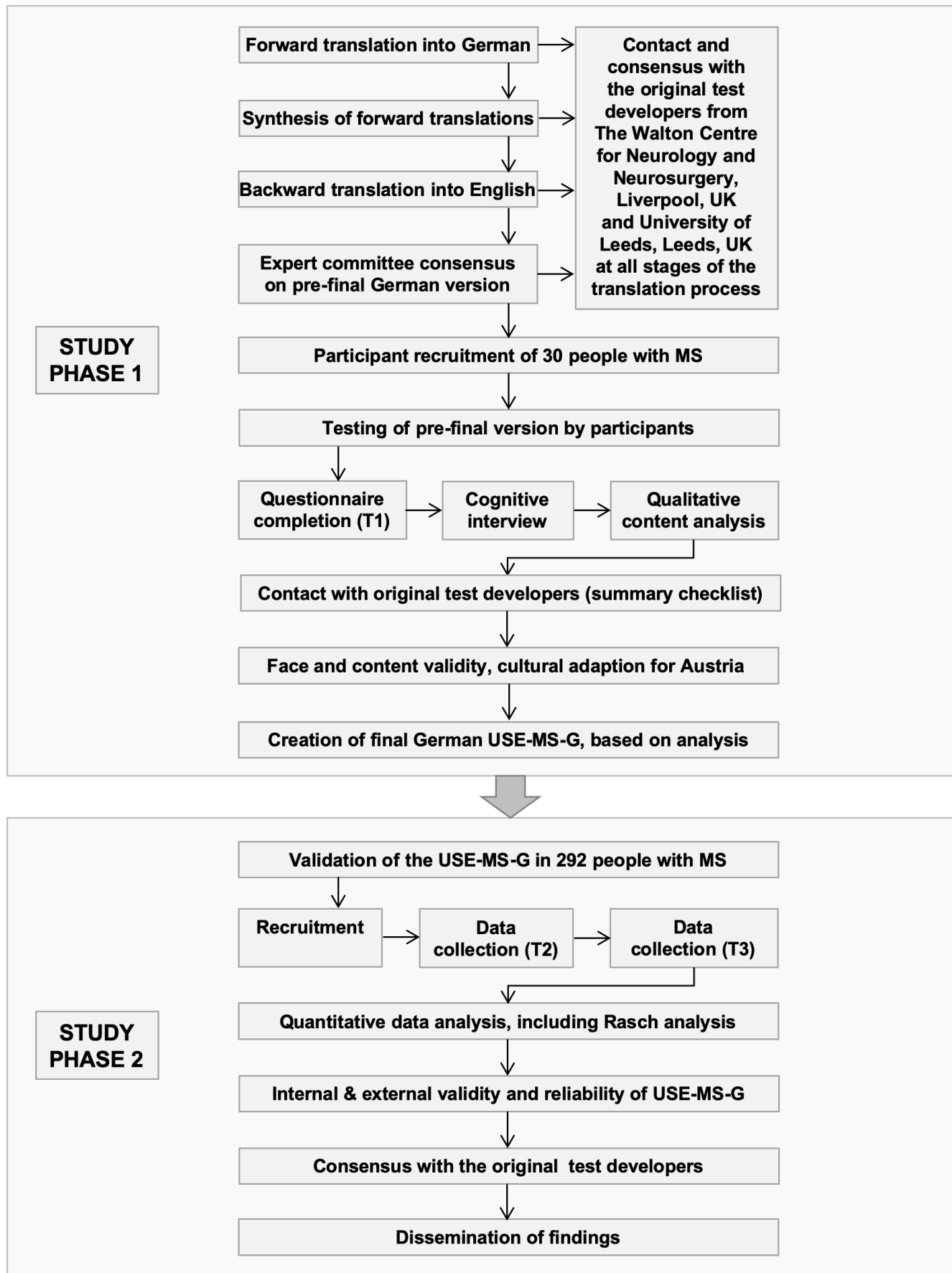
Figure 3 Flowchart of the study procedures. MS, multiple sclerosis; T1 (2; 3), testing 1 (2; 3); USE-MS-G, German Unidimensional Self-Efficacy Scale for Multiple Sclerosis.

logits (±0.5 logits), corresponding to the scale's item of mean difficulty.[81]

The proportion of floor and ceiling effects will be monitored, considered noteworthy if >5%.[82]

Local independence

Local independence means there should be no associations between the items. Inspection of the correlation matrix of item standardised residuals should show Pearson's correlations of <0.2 above the mean value of the matrix as a whole.

Unidimensionality

Unidimensionality as a Rasch model requirement allows a summary score measurement of a single construct. Using a PCA of the residuals, positively and negatively loadings of the first component will be identified, generating two subsets and separate person estimates. Independent t-tests will explore significant differences.[83] If less than 5% of t-tests are significant or the lower bound of the binomial CI overlaps 5%, unidimensionality is supported.[84 85]

Fit to the Rasch model

Different fit statistics will seek to determine if the assumption of a probabilistic ordering of items is satisfied:

1. Summary χ^2 interaction statistics and individual item χ^2 statistics are expected to be non-significant (Bonferroni-adjusted p values for the number of items).[45]
2. Individual person and item fit residuals are expected to be between ±2.5 (99% CI).[86]
3. Person and summary item fit residuals reflect perfect model fit if their mean and SD are close to 0 and 1, respectively.[87 88]

Reliability

Reliability is indicated by the person separation index (range: 0–1)[89] and Cronbach's alpha (missing data excluded), which should be ≥0.85 for individual use or 0.70 for group use.[45 90]

Invariance, DIF and differential test functioning

Invariance means that all persons completing a questionnaire, regardless of their ability (or self-efficacy), recognise the difficulty in identical items.[89] Any likelihood of differently scored items between the groups violates the assumption of invariance, called DIF.[91 92] The data will be pooled with a dataset from the UK development sample and tested for invariance by language to equate the language versions. Absence of DIF will be tested in gender (female; male), age (quartile groups), disease duration (quartile groups), language (English, German), timepoint (retest) and centre and indicated by a non-significant analysis of variance of the residuals (5% alpha with Bonferroni correction) where the group is the main factor.[92 93] Any observed DIF will be examined to know whether it cancels out at the test level.[91] If there are many items displaying DIF by language, differential test functioning will be performed.

If model fit is not achieved, an iterative stepwise procedure will be initiated, involving strategies for combining response categories, stepwise deletion of the worst fitting item, testlet (superitem) construction and adjusting for DIF as appropriate.[94]

Ethics approval, permissions and dissemination plan

Due to the absence of an intervention, no insurance policy is required for this study and no harm to participants is expected.

Permission to translate into German and validate the original USE-MS[21] was provided by the test developers who hold the copyright for the USE-MS-G.

Author affiliations
[1]Clinical Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria
[2]Department of Neurology, Walton Centre NHS Foundation Trust, Liverpool, UK
[3]School of Health Sciences, University of Brighton, Eastbourne, UK
[4]Clinical Department of Neurology, University Hospital of Innsbruck, Innsbruck, Austria
[5]Department of Neurology, Clinic for Rehabilitation Münster, Münster, Austria
[6]Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster, Münster, Austria

REFERENCES

1. Pugliatti M, Rosati G, Carton H, *et al*. The epidemiology of multiple sclerosis in Europe. *Eur J Neurol* 2006;13:700–22.
2. CompstonA, ConfavreuxC, LassmannH. *McAlpine's multiple sclerosis*. London: Elsevier, 2006.
3. Bandura A. *Social foundations of thought and action: a social cognitive theory*. Englewood Cliffs: Prentice-Hall, 1986.
4. Bandura A, Schunk DH. Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *J Pers Soc Psychol* 1981;41:586–98.
5. Bandura A. Self-efficacy. In: *Encyclopedia of human behavior*. New York: Academic Press, 1994: 71–81.

6. Jongen PJ, Lemmens WA, Hoogervorst EL, *et al*. Glatiramer acetate treatment persistence - but not adherence - in multiple sclerosis patients is predicted by health-related quality of life and self-efficacy: a prospective web-based patient-centred study (CAIR study). *Health Qual Life Outcomes* 2017;15:50.
7. Casey B, Uszynski M, Hayes S, *et al*. Do multiple sclerosis symptoms moderate the relationship between self-efficacy and physical activity in people with multiple sclerosis? *Rehabil Psychol* 2018;63:104–10.
8. Riazi A, Thompson AJ, Hobart JC. Self-Efficacy predicts self-reported health status in multiple sclerosis. *Mult Scler* 2004;10:61–6.
9. Hughes AJ, Beier M, Hartoonian N, *et al*. Self-Efficacy as a longitudinal predictor of perceived cognitive impairment in individuals with multiple sclerosis. *Arch Phys Med Rehabil* 2015;96:913–9.
10. Motl RW, Snook EM, Activity P. Physical activity, self-efficacy, and quality of life in multiple sclerosis. *Ann Behav Med* 2008;35:111–5.
11. Stuifbergen AK, Seraphine A, Roberts G. An explanatory model of health promotion and quality of life in chronic disabling conditions. *Nurs Res* 2000;49:122–9.
12. Motl RW, McAuley E, Snook EM, *et al*. Physical activity and quality of life in multiple sclerosis: intermediary roles of disability, fatigue, mood, pain, self-efficacy and social support. *Psychol Health Med* 2009;14:111–24.
13. Wulf G, Lewthwaite R. Optimizing performance through intrinsic motivation and attention for learning: the optimal theory of motor learning. *Psychon Bull Rev* 2016;23:1382–414.
14. French MA, Moore MF, Pohlig R, *et al*. Self-Efficacy mediates the relationship between Balance/Walking performance, activity, and participation after stroke. *Top Stroke Rehabil* 2016;23:77–83.
15. Schwarzer R, Jerusalem M. Generalized Self-Efficacy scale. In: Weinman J, Wright S, Johnston M, eds. *Measures in health psychology: A user's portfolio Causal and control beliefs*. Windsor, UK: NFER-NELSON, 1995: 35–7.
16. Motl RW, Balto JM, Ensari I, *et al*. Self-Efficacy and walking performance in persons with multiple sclerosis. *J Neurol Phys Ther* 2017;41:114–8.
17. Schmitt MM, Goverover Y, Deluca J, *et al*. Self-Efficacy as a predictor of self-reported physical, cognitive, and social functioning in multiple sclerosis. *Rehabil Psychol* 2014;59:27–34.
18. Airlie J, Baker GA, Smith SJ, *et al*. Measuring the impact of multiple sclerosis on psychosocial functioning: the development of a new self-efficacy scale. *Clin Rehabil* 2001;15:259–65.
19. Rigby SA, Domenech C, Thornton EW, *et al*. Development and validation of a self-efficacy measure for people with multiple sclerosis: the multiple sclerosis self-efficacy scale. *Mult Scler* 2003;9:73–81.
20. Schwartz CE, Coulthard-Morris L, Zeng Q, *et al*. Measuring self-efficacy in people with multiple scleoris: a validation study. *Arch Phys Med Rehabil* 1996;77:394–8.
21. Young CA, Mills RJ, Woolmore J, *et al*. The unidimensional self-efficacy scale for MS (USE-MS): developing a patient based and patient reported outcome. *Mult Scler* 2012;18:1326–33.
22. Amtmann D, Bamer AM, Cook KF, *et al*. University of Washington self-efficacy scale: a new self-efficacy scale for people with disabilities. *Arch Phys Med Rehabil* 2012;93:1757–65.
23. Tsang S, Royse C, Terkawi A. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi J Anaesth* 2017;11:80–S89.
24. Rasch G. *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press, 1980.
25. Rasch G. *Probabilistic models for some intelligence and attainment tests (revised and expanded ED*. Chicago: The University of Chicago Press, 1960.
26. Calvert M, Kyte D, Mercieca-Bebber R, *et al*. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols. *JAMA* 2018;319:483–94.
27. Thompson AJ, Banwell BL, Barkhof F, *et al*. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology* 2018;17:162–73.10.1016/S1474-4422(17)30470-2
28. Polman CH, Reingold SC, Banwell B, *et al*. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol* 2011;69:292–302.
29. McDonald WI, Compston A, Edan G, *et al*. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001;50:121–7.
30. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;33:1444–52.
31. Saunders B, Sim J, Kingstone T, *et al*. Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant* 2018;52:1893–907.
32. Linacre JM. Sample size and item calibration stability. *Rasch Measurement Transactions* 1994;7.
33. Hobart J, Freeman J, Thompson A. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. *Brain* 2000;123:1027–40.
34. Sharrack B, Hughes RAC, Soudain S, *et al*. The psychometric properties of clinical rating scales used in multiple sclerosis. *Brain* 1999;122:141–59.
35. Meyer-Moock S, Feng Y-S, Maeurer M, *et al*. Systematic literature review and validity evaluation of the expanded disability status scale (EDSS) and the multiple sclerosis functional composite (MSFC) in patients with multiple sclerosis. *BMC Neurol* 2014;14:58.
36. Schwarzer R, Jerusalem ME. *Skalen Zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation Der psychometrischen Verfahren Im Rahmen Der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin, 1999.
37. Schumacher J, Leppert K, Gunzelmann T, *et al*. Die Resilienzskala – ein Fragebogen Zur Erfassung Der psychischen Widerstandsfähigkeit ALS Personmerkmal. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie* 2004;53:16–39.
38. Leppert K, Koch B, Brähler E, *et al*. Die Resilienzskala (RS) – Überprüfung Der Langform RS-25 und einer Kurzform RS-13. *Klinische Diagnostik und Evaluation* 2008;2:226–43.
39. Wagnild GM, Young HM. Development and psychometric evaluation of the resilience scale. *J Nurs Meas* 1993;1:165–78.
40. Flachenecker P, Vogel U, Simeoni MC, *et al*. MusiQol: international questionnaire investigating quality of life in multiple sclerosis: validation results for the German subpopulation in an international comparison]. *Nervenarzt* 2011;82:1281–9.
41. Simeoni MC, Auquier P, Fernandez O, *et al*. Validation of the multiple sclerosis International quality of life questionnaire. *Mult Scler* 2008;14:219–30.
42. Petermann F, Anxiety H, Scale D, *et al*. HADS-D). *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie* 2011;59:251–3.
43. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
44. NFI-MS Neurologischer Fragebogen zur Müdigkeit. *NFI-MS Austria/ German - Version of 30 Sep 13 - Mapi ID7555 / NFI-MS_AU10_deu-ATdoc*, 2010.
45. Mills RJ, Young CA, Pallant JF, *et al*. 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.
46. Mills RJ, Calabresi M, Tennant A, *et al*. Perceived changes and minimum clinically important difference of the neurological fatigue index for multiple sclerosis (NFI-MS). *Mult Scler* 2013;19:502–5.
47. Marx RG, Menezes A, Horovitz L, *et al*. A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol* 2003;56:730–5.
48. Beaton D, Bombardier C, Guillemin F, *et al*. *Recommendations for the Cross-Cultural Adaptation of the DASH & QuickDASH Outcome Measures*. Toronto, Canada: Institute for Work and Health, 2007.
49. 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;46:1417–32.
50. Brislin RW, Lonner WJ, Thorndike RM. *Questionnaire wording and translation. Cross-cultural research methods*. New York: John Wiley, 1973: 32–58.
51. Webster K, Cella D, Yost K. The functional assessment of chronic illness therapy (FACIT) measurement system: properties, applications, and interpretation. *Health Qual Life Outcomes* 2003;1:79.
52. Flesch R. *How to write plain English Canterbury, UK: University of Canterbury. archived from the original on July 12, 2016*, 2016.
53. Willis G, Interviewing C. *Questionnaire design: a training manual. cognitive methods staff working paper series no 7*. Hyattsville, MD, USA: Office of Research and Methodology, National Center for Health Statistics, 1994.
54. Reducing Survey Error through Research on the Cognitive and Decision Processes in Surveys. *Meeting of the American statistical association*. Baltimore, Maryland, 1999: 8–12.
55. Beatty PC, Willis GB. Research synthesis: the practice of cognitive interviewing. *Public Opin Q* 2007;71:287–311.
56. Oksenberg L, Cannell CF, Kalton G. New strategies for Pretesting survey questions. *Journal of Official Statistics* 1991;7:349–65.
57. Conrad F, Blair J. Data Quality in Cognitive Interviews: The Case for Verbal Reports. In: Presser S, Rothgeb JM, Couper MP, *et al.*, eds. *Methods for testing and evaluating survey questionnaires*. Hoboken, NJ: John Wiley and Sons, 2004.
58. Willis G. *Cognitive interviewing: a tool for improving questionnaire design. thousand oaks*. CA: Sage, 2005.

59. Berelson B. *Content analysis in communication research*. Glencoe: Free Press, 1952.
60. Lasswell HD. *Power and personality. New York*. NY: W: W. Norton & Company, 1948.
61. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19:349–57.
62. Schreier M. *Qualitative content analysis in practice*. London, UK: SAGE Publications, 2012.
63. Bryman A, Burgess RGE. *Analyzing qualitative data*. Abingdon, Oxon: : Routledge, 2007.
64. Denzin NK, Lincoln YSE. *The SAGE Handbook of qualitative research*. Thousand Oaks, CA: SAGE Publications, 20185th ed.
65. BrymanA. *Social research methods*. New York, NY: Oxford University Press, 2012.
66. Cypress BS. Rigor or reliability and validity in qualitative research: perspectives, strategies, reconceptualization, and recommendations. *Dimens Crit Care Nurs* 2017;36:253–63.
67. Noble H, Smith J. Issues of validity and reliability in qualitative research. *Evid Based Nurs* 2015;18:34–5.
68. Validity LL. Reliability, and generalizability in qualitative research. *J Family Med Prim Care* 2015;4:324–7.
69. Miles HB, Huberman AM. *Qualitative data analysis*. Thousand Oaks, CA: Sage, 1994.
70. Andrich D, Lyne A, Sheridon B, *et al*. *RUMM 2030*. Perth: RUMM Laboratory, 2009.
71. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009;13:1–177. iii, ix-x.
72. LI-K L. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255–68.
73. LI-K L. Corrections. *Biometrics* 2000;56:324–5.
74. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
75. Hinkle DE, Wiersma W, Jurs SG. *Applied statistics for the behavioral sciences*. 5th ed. Boston: Houghton Mifflin, 2003.
76. Andrich D. *Rasch models for measurement*. Beverly Hills: SAGE, 1988.
77. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? when should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007;57:1358–62.
78. Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982;47:149–74.
79. AndrichD. A rating formulation for ordered response categories. *Psychometrika* 1978;43:561.
80. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002;3:85–106.
81. Khadka J, Pesudovs K, McAlinden C, *et al*. Reengineering the glaucoma quality of life-15 questionnaire with Rasch analysis. *Invest. Ophthalmol. Vis. Sci.* 2011;52:6971–7.
82. Fisher WPJ. Rating scale instrument quality criteria. *Rasch Measurement Transactions* 2007;21.
83. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002;3:205–31.
84. Tennant A, Pallant JF, matters! U. A tale of two Smiths?). *Rasch Measurement Transactions* 2006;20:1048–51.
85. Kersten P, White PJ, Tennant A. Is the pain visual analogue scale linear and responsive to change? an exploration using Rasch analysis. *PLoS One* 2014;9:e99485.
86. Andrich D. *Rasch models for measurement series: quantitative applications in the social sciences No. 68*. London: Sage, 1988: 68–86.
87. Twiss J, McKenna SP, Graham J, *et al*. Applying Rasch analysis to evaluate measurement equivalence of different administration formats of the activity limitation scale of the Cambridge pulmonary hypertension outcome review (camphor). *Health Qual Life Outcomes* 2016;14:57.
88. Milinis K, Tennant A, A. Young C, *et al*. Rasch analysis of SF-Qualiveen in multiple sclerosis. *Neurourol Urodyn* 2017;36:1161–6.
89. BondTG, FoxCM. *Applying the Rasch model: fundamental measurement in the human sciences*. London: Lawrence Erlbaum Associates, 2007.
90. Linacre JM. Data variance explained by Rasch measures. *Rasch Measurement Transactions* 2006;20:1045–54.
91. Tennant A, Pallant JF. Dif matters: a practical approach to test if differential item functioning makes a difference. *Rasch Measurement Transactions* 2007;20:1082–4.
92. HollandPW, WainerH. *Differential item functioning*. Hillsdale, N.J: Lawrence Erlbaum, 1993.
93. Grimby G. Useful reporting of DIF. *Rasch Meas Transactions* 1998;12.
94. Siegert R, Jackson D, Tennant A, *et al*. Factor analysis and Rasch analysis of the Zarit burden interview for acquired brain injury carer research. *J Rehabil Med* 2010;42:302–9.
95. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National multiple sclerosis Society (USA) Advisory Committee on clinical trials of new agents in multiple sclerosis. *Neurology* 1996;46:907–11.
96. Montalban X, Gold R, Thompson AJ, *et al*. ECTRIMS/EAN guideline on the pharmacological treatment of people with multiple sclerosis. *Mult Scler* 2018;24:96–120.
97. Diener H-C, Weimar CH. Diagnose und Therapie der Multiplen Sklerose,Gültigkeit der Leitlinie nach Überprüfung durch das Leitliniensekretariat verängert bis 29.2.2017. In: Diener H-C, Weimar C, eds. *Leitlinien für Diagnostik und Therapie in der Neurologie*. Stuttgart: Kommission "Leitlinien" der Deutschen Gesellschaft für Neurologie, Thieme Verlag 2012/2017.
98. Olsson J-E, Ekblad S, Bertilson BC, *et al*. Swedish adaptation of the general medical Council's multisource feedback questionnaires: a qualitative study. *Int J Med Educ* 2018;9:161–9.