PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	GERMAN TRANSLATION, CULTURAL ADAPTION AND VALIDATION OF THE UNIDIMENSIONAL SELF-EFFICACY SCALE FOR MULTIPLE SCLEROSIS: STUDY PROTOCOL
AUTHORS	Seebacher, Barbara; Mills, Roger J; Reindl, Markus; Zamarian, Laura; Kuisma, Raija; Kircher, Simone; Brenneis, Christian; Ehling, Rainer; Deisenhammer, Florian

VERSION 1 – REVIEW

REVIEWER	Ali Dehghani Dr. Ali Dehghani Jahrom University of Medical Siences. IRAN
REVIEW RETURNED	12-Feb-2019

GENERAL COMMENTS	The reviewer also provided a marked copy with additional comments. Please contact the publisher for full details.

REVIEWER	Hilda Mulligan Senior Lecturer, School of Physiotherapy, Otago University, New Zealand
REVIEW RETURNED	11-Mar-2019

GENERAL COMMENTS	The manuscript contains a descriptive title that identifies the study's intent, design, and population of interest. There is an excellent yet succinct introduction leading to justification for the purpose of the study. The three aims of the study are clearly articulated. There is a very clear and reproducible description of the study design and study procedures. This details the study setting, timeline, inclusion and exclusion criteria for study participants, recruitment approaches (and how the research team will include people with severe multiple sclerosis who may not wish or be able to visit the study centre for data collection), and ethical considerations and approval. There is explanation for sample size calculations, where and how data will be collected, stored and retrieved for analysis, and the processes via which collection of good quality data will be promoted. Analysis of both qualitative and quantitative data is exhaustively described and justified. The manuscript contains evidence of data monitoring via an Expert committee which includes health professionals, a methodologist, translators, language professionals and lay people with multiple sclerosis. There is a statement that the funders of the study have no role in the study design or preparation of the manuscript. The study has a clear plan for dissemination of the findings to trial participants and to other stakeholders. There is

	clear indication of authors' contributions to the study, and where further information can be sourced by others if relevant. I commend this research team on a well presented, descriptive and detailed study protocol for a German translation, cultural adaption and validation of the Unidimensional Self-efficacy Scale for Multiple Sclerosis. This research project will enable valid measurement of self-efficacy for people with multiple sclerosis in the German population.

REVIEWER	AS University Italy
REVIEW RETURNED	13-Apr-2019

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.

REVIEWER	Peter Watson University of Cambridge UK
REVIEW RETURNED	29-Apr-2019

GENERAL COMMENTS	German translation, cultural adaption and validation of the unidimensional self-efficacy scale for multiple sclerosis: study protocol. bmjopen-2019-029565 I would suggest analysing this data using Factor analyses rather than the Rasch analysis suggested. Loadings obtained from Factor analysis could be used to assess both internal consistency and validity. I don't see from this paper how Rasch analysis would do either of these. I also don't follow the power calculations. Page 3, line 44. Since the aim here is to assess internal consistency (line 42) I wondered why the usual approach of a factor analysis quoting a Cronbach's alpha which is used for reliability could not be used. Could not a correlation with a 'gold standard' measure e.g. scores from the English language version be used to assess the internal validity (line 44)? In other words to see what the German version of the questionnaire is measuring. One could for example use a congruence coefficient which compares pairs of factor loadings between different language versions of the same test. See e.g. Hofstee et al. (1997) amongst others. I don't see the point of using a Rasch analysis. Page 10. It is not clear to me how the sample size for the referenced power calculation (lines 27-34) has been derived. The phrase used here of "detect a problem within a questionnaire" is vague, I also don't know what is meant by "saturation" in line 34. Please clarify this term and what effect size or estimate of precision you are using to derive the sample size required.

	Page 10, lines 39-48. I don't think expressing a precision in log odds units is very helpful since results are expressed in lay terms weither as odds ratios or proportions. Could you not work out a sample size for a precision based upon a particular proportion? I find 1/(1-0.8) x 243 = 304 people > 292 as quoted on line 46 needed for the study correcting for attrition of 20% with a sample size of 243. Reference Hofstee, W. K. B., Kiers, H. A. L., De Raad, B. D., Goldberg, L. R. and Ostendorf, F. (1997) A Comparison of Big-Five structures of personality traits in Dutch, English, and German (1997) European Journal of Personality 11(1):15-31

REVIEWER	Stanley Hum McGill University Canada
REVIEW RETURNED	01-Jun-2019

GENERAL COMMENTS	5. Is it appropriate to have the author/neurologist consent patients? Conflict of interest. 7. o)verall, statistics are appropriate. minor points in comments below. 13. page numbers to not match in supp. material (spriit). I will not comment on the qualitative analysis. Rasch analysis: With two available timepoints, authors should consider stacking and racking the data for analysis. Ordered thresholds: Protocol states that ordered thresholds will be checked but did not state how it plans to handle disordered thresholds. Instructions from the editor was to prevent unnecessary duplication of work and enable collaboration. Considering this is a “translation” of the English USE-MS and one of the authors was part of the development of the English version (using Rasch analysis); if the English dataset is available an important DIF analysis by language would provide important information. In todays Open science initiatives where pooling data can increase sample size and lead to opportunities for secondary analysis of data that have the potential of lead to novel results. If there is an interest to do this, the German translated data should match the personal factors for DIF analysis used in the English Rasch analysis. DIF analysis: Other Rasch analysis of self-efficacy showed DIF by education, work status. Why are education and work status not included as personal factors to check for DIF? There are two centers and this should be checked for DIF.

	Everyone checks for DIF how about Differential Test Functioning (DTF) which is often more important. Although I generally agree with the statement that “generic questionnaires may not adequately cover the construct of self-efficacy in a chronic neurological disease like MS”; this can be assessed. A Rasch analysis of all items from self-efficacy scales would be appropriate. Will the data be made available as part of the Open science initiative? Methods: Protocol states personal data will be anonymised by a participant ID. Is the data truly anonymised or more accurately codified.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Ali Dehghani
Institution and Country: Dr. Ali Dehghani - Jahrom University of Medical Siences. IRAN Please state any competing interests or state ‘None declared’: None declared.

Please leave your comments for the authors below Refer to the upload file

Reviewer: 2
Reviewer Name: Hilda Mulligan
Institution and Country: Senior Lecturer, School of Physiotherapy, Otago University, New Zealand
Please state any competing interests or state ‘None declared’: None declared

Dear Dr Mulligan, thank you very much for your comments.

Please leave your comments for the authors below. The manuscript contains a descriptive title that identifies the study’s intent, design, and population of interest. There is an excellent yet succinct introduction leading to justification for the purpose of the study. The three aims of the study are clearly articulated.
There is a very clear and reproducible description of the study design and study procedures. This details the study setting, timeline, inclusion and exclusion criteria for study participants, recruitment approaches (and how the research team will include people with severe multiple sclerosis who may not wish or be able to visit the study centre for data collection), and ethical considerations and approval. There is explanation for sample size calculations, where and how data will be collected, stored and retrieved for analysis, and the processes via which collection of good quality data will be promoted. Analysis of both qualitative and quantitative data is exhaustively described and justified.
The manuscript contains evidence of data monitoring via an Expert committee which includes health professionals, a methodologist, translators, language professionals and lay people with multiple sclerosis. There is a statement that the funders of the study have no role in the study design or preparation of the manuscript. The study has a clear plan for dissemination of the findings to trial participants and to other stakeholders. There is clear indication of authors’ contributions to the study, and where further information can be sourced by others if relevant.

I commend this research team on a well presented, descriptive and detailed study protocol for a German translation, cultural adaption and validation of the Unidimensional Self-efficacy Scale for Multiple Sclerosis. This research project will enable valid measurement of self-efficacy for people with multiple sclerosis in the German population.

Reviewer: 3
Reviewer Name: A Signori
Institution and Country: University Italy. Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Reviewer: 4
Reviewer Name: Peter Watson
Institution and Country: University of Cambridge UK Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Dear Dr Watson, thank you for your comments and the opportunity to improve the quality of our article. Please find our point-to-point responses as follows.

I would suggest analysing this data using Factor analyses rather than the Rasch analysis suggested. Loadings obtained from Factor analysis could be used to assess both internal consistency and validity. I don't see from this paper how Rasch analysis would do either of these. I also don't follow the power calculations.

Page 3, line 44. Since the aim here is to assess internal consistency (line 42) I wondered why the usual approach of a factor analysis quoting a Cronbach's alpha which is used for reliability could not be used. Could not a correlation with a 'gold standard' measure e.g. scores from the English language version be used to assess the internal validity (line 44)? In other words to see what the German version of the questionnaire is measuring. One could for example use a congruence coefficient which compares pairs of factor loadings between different language versions of the same test. See e.g. Hofstee et al. (1997) amongst others. I don't see the point of using a Rasch analysis.

Rasch analysis has superseded classical test theory as the standard and preferred method of determining construct validity of a scale6. It is the only way to convert ordinal scale scores into interval level data suitable for parametric analysis7 8. It facilitates a more robust method of assessing unidimensionality than factor analysis and the testing of invariance in item difficulty across groups of respondents (differential item functioning) is superior to comparison of factor loadings9 10. We now intend to test invariance by language against an existing data set from the original UK development of the scale.

Page 10. It is not clear to me how the sample size for the referenced power calculation (lines 27-34) has been derived. The phrase used here of "detect a problem within a questionnaire" is vague, I also don't know what is meant by "saturation" in line 34. Please clarify this term and what effect size or estimate of precision you are using to derive the sample size required.

We agree with you and removed the sample size calculation, due to the fact that Phase 1 is a qualitative study. In addition, we explained the qualitative research term “saturation”. The section now reads:
“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 to 30 people have been interviewed but is determined by the nature of the analysis and the participants themselves11.”

Page 10, lines 39-48. I don't think expressing a precision in log odds units is very helpful since results are expressed in lay terms weither as odds ratios or proportions. Could you not work out a sample size for a precision based upon a particular proportion? I find 1/(1-0.8) x 243 = 304 people > 292 as quoted on line 46 needed for the study correcting for attrition of 20% with a sample size of 243.

The logit is the basic unit of Rasch analysis; both item difficulty and person location are determined in logits upon the same linear metric if the data fit the Rasch model. The minimum sample size of 243 is a standard figure in the Rasch literature to produce a high degree of precision in the logit estimates. For the calculation of the total sample size including an expected attrition rate, we performed a re-analysis. Given a correct result, there seemed to be an error in your formula. Using the procedure of dividing the number (n) by 1 minus the dropout rate we arrived at the adjusted sample size (N). i.e., N = n / (1-(z/100)). For our study and based on a 20% attrition rate, this is 243/(1-(20/100))=303,75 (i.e. 304 participants). However, as this is a cross-sectional study with only two data collection points within a short period of time and based on comparable studies12 13, we suggest that the attrition rate may be much lower than 20%.
Hence, we corrected the description of the sample size calculation concerning the attrition rate in our manuscript to:
“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.”

Reference

Hofstee, W. K. B., Kiers, H. A. L., De Raad, B. D., Goldberg, L. R. and Ostendorf, F. (1997) A Comparison of Big-Five structures of personality traits in Dutch, English, and German (1997) European Journal of Personality 11(1):15-31

Reviewer: 5
Reviewer Name: Stanley Hum
Institution and Country: McGill University - Canada Please state any competing interests or state ‘None declared’: none declared

Dear Dr Hum, thank you for your comments and the opportunity to improve the quality of our article. Please find our point-to-point responses as follows.

Please leave your comments for the authors below 5. Is it appropriate to have the author/neurologist consent patients? Conflict of interest.

We agree with you and have therefore changed the description of the informed “Written informed consent will be obtained by the first author (BS) who is not involved in the treatment of the patients.”

7. overall, statistics are appropriate. minor points in comments below.
13. page numbers to not match in supp. material (spriit).

We edited the page numbers in the SPIRIT Checklist so that they now match the pages of our manuscript.

I will not comment on the qualitative analysis.

Rasch analysis:
With two available timepoints, authors should consider stacking and racking the data for analysis.

Yes, the data will be racked for the analysis of the concordance correlation coefficient, and stacked for DIF by time point. We have included this information in the Test-Retest Reliability Section of the manuscript.

Ordered thresholds: Protocol states that ordered thresholds will be checked but did not state how it plans to handle disordered thresholds.

We will use standard techniques of collapsing response categories to deal with disordered thresholds, once any local dependence has been rectified. This is briefly mentioned starting page 19 line 59 (unrevised manuscript).

Instructions from the editor was to prevent unnecessary duplication of work and enable collaboration. Considering this is a “translation” of the English USE-MS and one of the authors was part of the development of the English version (using Rasch analysis); if the English dataset is available an important DIF analysis by language would provide important information. In todays Open science initiatives where pooling data can increase sample size and lead to opportunities for secondary analysis of data that have the potential of lead to novel results. If there is an interest to do this, the German translated data should match the personal factors for DIF analysis used in the English Rasch analysis.

Yes, we agree, pooling with a data set from the UK development sample and testing for invariance by language would be the ideal method for equating the language versions and is now part of the protocol:
“The data will be pooled with a data set 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), time point (retest) and centre and indicated by a non-significant ANOVA of the residuals (5% alpha with Bonferroni correction) where the group is the main factor14 15.”

DIF analysis:
Other Rasch analysis of self-efficacy showed DIF by education, work status.
Why are education and work status not included as personal factors to check for DIF?
There are two centers and this should be checked for DIF.

Factors for DIF need to be independent of the ‘causal pathway’ of the latent trait16 and so in this we decided to avoid factors which might directly affect self-efficacy and stick to factors which should be truly independent such as age and sex. Our protocol has now been amended to also include DIF by time point (retest), language and centre.

Everyone checks for DIF how about Differential Test Functioning (DTF) which is often more important.

If there are many items displaying DIF by language then we can assess DTF. This is also part of the protocol now: “If there are many items displaying DIF by language, Differential Test Functioning (DTF) will be performed.”

Although I generally agree with the statement that “generic questionnaires may not adequately cover the construct of self-efficacy in a chronic neurological disease like MS”; this can be assessed. A Rasch analysis of all items from self-efficacy scales would be appropriate.

Whilst some scales purport to measure the same latent trait, different methods of development in different subject populations may mean that amalgamation for Rasch analysis is not valid but such an analysis is beyond the scope of this study.

Will the data be made available as part of the Open science initiative?

We will be happy to consider submitting our data to the Open Science initiative once we have completed future analyses related to this data set. We have added this information to the Data Sharing Statement.

Methods:
Protocol states personal data will be anonymised by a participant ID. Is the data truly anonymised or more accurately codified.

We agree with you that the data will be codified and have changed the term in our manuscript.

VERSION 2 – REVIEW

REVIEWER	Stanley Hum Montreal Neurological Institute and Hospital Brain Health Outcomes Platform 3801 University St. Montreal, Quebec Canada H3A 2B4
REVIEW RETURNED	21-Jul-2019

GENERAL COMMENTS	Suggestion: I seems the data may be made available under a controlled data access model. If that is the case and if it has not already been done, the informed consent form should reflects the requirements of shared data under the controlled data access model such that participants do not need to be reconsented to share data with researcher not affiliated with the research group conducting the study.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 5

Reviewer Name: Stanley Hum

Institution and Country: Montreal Neurological Institute and Hospital, Brain Health Outcomes Platform, 3801 University St. Montreal, Quebec, Canada, H3A 2B4

Please state any competing interests or state ‘None declared’: None declared

Suggestion: I seems the data may be made available under a controlled data access model. If that is the case and if it has not already been done, the informed consent form should reflects the requirements of shared data under the controlled data access model such that participants do not need to be reconsented to share data with researcher not affiliated with the research group conducting the study.

Dear Dr Hum, thank you for your comment.

The informed consent form to be used already includes the consent to controlled data sharing. We added the following sentence to our data sharing statement:

“The informed consent form will include the consent to controlled data sharing.”