

Psychometrische Prüfung des deutschsprachigen „Neurologischen Fragebogens zur Müdigkeit bei Multipler Sklerose (NFI-MS-G)“ bei Rehabilitanden mit Multipler Sklerose

Psychometric Evaluation of the ‘German Neurological Fatigue Index for Multiple Sclerosis (NFI-MS-G)’ in a Sample of Rehabilitation Patients with Multiple Sclerosis



Autoren

Barbara Seebacher^{1,2}, Mike C. Horton³, Markus Reindl¹, Christian Brenneis^{2,4}, Rainer Ehling^{2,4}, Florian Deisenhammer¹, Roger J. Mills⁵

Institute

- 1 Universitätsklinik für Neurologie, Medizinische Universität Innsbruck, Österreich
- 2 Karl Landsteiner Institut für Interdisziplinäre Rehabilitationsforschung, Reha Zentrum Münster, Österreich
- 3 Psychometric Laboratory for Health Sciences, Faculty of Medicine and Health, University of Leeds, UK
- 4 Department für Neurologie, Reha Zentrum Münster, Österreich
- 5 Department of Neurology, Walton Centre NHS Foundation Trust, Liverpool, UK

Schlüsselwörter

Multiple Sklerose, Fatigue, Instrumente zur Selbsteinschätzung, Österreich, Validierung.

Key words

Multiple Sclerosis, Fatigue, Patient Reported Outcome Measures, Austria, Validation Studies as Topic.

online publiziert 14.12.2022

Bibliografie

Rehabilitation 2023; 62: 31–39

DOI 10.1055/a-1903-4483

ISSN 0034-3536

© 2022. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Korrespondenzadresse

Dr. Seebacher Barbara
Universitätsklinik für Neurologie
Medizinische Universität Innsbruck
Anichstrasse 35
6020 Innsbruck
Österreich
barbara.seebacher@i-med.ac.at

Zusätzliches Material finden Sie unter <https://doi.org/10.1055/a-1903-4483>.

ZUSAMMENFASSUNG

Ziel Ziel dieser Studie war es, ein Instrument zur Selbsteinschätzung von Fatigue für Multiple-Sklerose(MS)-Patientinnen und -Patienten zu erstellen, welches das Konstrukt der Fatigue umfassend abbildet und anhand der Annahmen des Rasch-Modells entwickelt wurde. Der Neurological Fatigue Index – Multiple Sclerosis (NFI-MS) basiert sowohl auf einem medizinischen als auch von MS-Betroffenen beschriebenen Bezugsrahmen von Fatigue-Symptomen und wurde stringent validiert. Daher wurde in dieser Studie die deutsche Fassung des NFI-MS, der Neurologische Fragebogen zur Müdigkeit bei Multipler Sklerose (NFI-MS-G), bestehend aus einer physischen und kognitiven Subskala, validiert.

Methodik In dieser bizenrischen Studie komplettierten 309 MS-Betroffene im Rahmen ihrer ambulanten Rehabilitation oder ≥ 2 Monate vor bzw. nach ihrer stationären Rehabilitation zu zwei Testzeitpunkten innerhalb von 14–21 Tagen den NFI-MS-G sowie weitere Fragebögen. Die Validierung erfolgte anhand von Korrelationen mit etablierten Fragebögen und der Rasch-Analyse. Zusätzlich wurden psychometrische Gütekriterien der Known-Groups-Validität, internen Konsistenz, Test-Retest-Reliabilität, Messgenauigkeit und Lesbarkeit evaluiert. Abschließend wurden die englische NFI-MS und deutsche NFI-

MS-G miteinander verglichen, um zu überprüfen, ob die Sprachversionen einander entsprechen.

Ergebnisse Die NFI-MS-G zeigte eine gute interne Konstruktvalidität, konvergente und Known-Groups-Validität und interne Konsistenz (Cronbachs Alpha 0,84–0,93). Die physische Subskala zeigte geringgradige Abhängigkeiten zwischen Items 1 und 7, 2 und 3 und 4 bis 6, die mittels einer Kombination betroffener Items zu Testlets behandelt werden konnten. Die physische und kognitive Subskala zeigten sich eindimensional, nicht jedoch die Summenskala. Anstelle der Summenskala wurde ein 2-Domänen-Subtest erstellt, der ein Fatigue-Konstrukt höherer Ordnung misst. Des Weiteren wurde eine gute Test-Retest-Reliabilität (Konkordanzkorrelations-Koeffizient nach Lin 0,86–0,90), geringe Boden- und Deckeneffekte, leichte Lesbarkeit und Invarianz für Gruppen unterschiedlichen Geschlechts, Alters, Krankheitsdauer, Testzeitpunkt und Zentrum festgestellt.

Schlussfolgerung Die deutsche Fassung des NFI-MS bildet das Konstrukt der Fatigue umfassend ab und besitzt eine adäquate psychometrische Güte. Die deutsche Fassung unterscheidet sich von der englischen Originalversion hinsichtlich mangelnder Eindimensionalität der Summenskala und geringgradiger lokaler Abhängigkeiten der physischen Subskala, die mittels einer Testlet-Analyse ausgeglichen werden konnten.

ABSTRACT

Purpose The purpose of this study was to provide a patient-reported outcome measure for people with multiple sclerosis (MS) comprehensively reflecting the construct of fatigue and developed upon the assumptions of the Rasch model. The Neurological Fatigue Index – Multiple Sclerosis (NFI-MS) is based on both a medical and patient-described symptom framework

of fatigue and has been validated. Therefore, in this study the German version of the NFI-MS (NFI-MS-G) consisting of a physical and cognitive subscale and a summary scale was validated.

Method In this bi-centre-study, 309 people with MS undergoing outpatient rehabilitation or being ≥ 2 months before or after their inpatient rehabilitation completed the German NFI-MS-G twice within 14–21 days together with other questionnaires. Correlation with established questionnaires and Rasch analysis were used for its validation. Additionally, psychometric properties of known-groups validity, internal consistency, test-retest reliability, measurement precision and readability were tested. Finally, the English NFI-MS and German NFI-MS-G were compared with each other to equate the language versions.

Results The NFI-MS-G showed good internal construct validity, convergent and known-groups validity and internal consistency (Cronbach's alpha 0.84–0.93). The physical subscale showed minor local dependencies between items 1 and 7, 2 and 3 and 4 to 6, that could be treated by combining the respective items to testlets. Unidimensionality was found for the physical and cognitive subscales but not for the summary scale. Replacing the summary scale, a 2-domains subtest measuring the higher-order construct of fatigue was created. Good test-retest reliability (Lin's concordance correlation coefficient of 0.86–0.90) and low floor and ceiling effects were demonstrated. The NFI-MS-G was found easily readable and invariant across groups of gender, age, disease duration, timepoint and centre.

Conclusion The German version of the NFI-MS comprehensively represents the construct of fatigue and has adequate psychometric properties. The German version differs from the English original version with respect to a lack of unidimensionality of the summary scale and minor local dependencies of the physical subscale that could be canceled out using a testlet analysis.

Einleitung

Die Multiple Sklerose (MS) ist eine entzündlich-demyelinisierende Erkrankung des Zentralen Nervensystems, die häufig zu progredienter Behinderung und Beeinträchtigung der Lebensqualität (LQ) führt [1]. Mindestens 60 % der MS-Patientinnen und MS-Patienten sind von Fatigue betroffen, einer ausgeprägten Müdigkeit, die ein initiales Symptom der Erkrankung darstellen kann [2]. Die Ätiologie der Fatigue konnte noch nicht zufriedenstellend geklärt werden und ist sehr wahrscheinlich multifaktoriell [1, 2]. Grundsätzlich kann zwischen kognitiver und physischer bzw. motorischer/muskulärer Fatigue unterschieden werden [2]. Der MS Council definierte Fatigue als einen subjektiven Mangel an physischer und/oder mentaler Energie, der von der betroffenen Person oder den Angehörigen als Beeinträchtigung der üblichen oder gewünschten Aktivität empfunden wird [3]. Die muskuläre Fatigue ist durch eine vorübergehende Abnahme muskulärer Kapazität beziehungsweise mangelnde Aufrechterhaltung des erforderlichen Kraftniveaus bei körperlicher Aktivität definiert [4]. Fatigue ist ein Krankheitssymptom, das die Teilhabe auf allen Ebenen ganz wesentlich beeinflusst, vor allem die Teilhabe am Arbeitsleben. Daher ist es rele-

vant, die Fatigue von anderen MS-Symptomen differenzieren zu können.

Unter Berücksichtigung der Komplexität der Fatigue bedarf es geeigneter Instrumente zur Selbsteinschätzung durch Patientinnen und Patienten, sogenannter Patient-Reported Outcome Measures (PROMs). Die Ergebnisse von PROMs dienen dazu, die klinische Entscheidungsfindung in der Rehabilitation zu unterstützen. Daher sollte der Summenscore eines Fragebogens die vollständige Information über den funktionellen Status eines Patienten enthalten. Zur Messung der MS-Fatigue werden bereits zahlreiche PROMs im klinischen Alltag und in der Forschung eingesetzt (vgl. die Übersichtsarbeiten [5, 6]). Ungeachtet der Vielzahl der Skalen und Häufigkeit ihrer Verwendung zeigten Studien gewisse Mängel ihrer psychometrischen Eigenschaften [5, 6]. Der Neurological Fatigue Index – Multiple Sclerosis (NFI-MS) ist eine MS-spezifische Fatigueskala, deren Items auf Basis von Patienten-Interviews generiert wurden und die anhand der Annahmen des Rasch-Modells entwickelt wurde [7]. Im Zuge mehrerer Studien wurde auf der Grundlage großer bevölkerungsrepräsentativen MS-Patienten-Stichproben die ursprüngliche 23-Item NFI-MS-Skala weiterentwickelt. Die psycho-

metrische Evaluierung erfolgte jeweils mittels einer Faktorenanalyse und Rasch-Analyse, sodass eine valide, reliable und änderungs-sensitive 12-Item-Skala mit einer jeweils eindimensionalen physischen und kognitiven Subskala resultierte [8]. Der Vorteil von Rasch-Modell-basierten Skalen ist, dass die Personenkennwerte intervallskaliert sind und das Antwortverhalten der Patienten durch eine einzige latente, das heißt nicht direkt messbare Merkmalsdimension bedingt ist [9]. Das Rasch-Theoriemodell ist ein probabilistisches Modell. Es überprüft die Wahrscheinlichkeit, mit der eine Person mit einer bestimmten ‚Fähigkeit‘, bezogen auf ein latentes Merkmal (Fatigue), eine spezifische Antwortkategorie eines Items mit einer definierten, messbaren „Schwierigkeit“ (dem ausgedrückten Ausmaß an Fatigue) auf einer Skala bestätigt [10]. Das bedeutet, dass – im Fall einer validen Skala (bzw. Subskala) – die Items einer bestimmten Schwierigkeit für alle Patienten gleich funktionieren und die Skala strikt eindimensional ist [9, 10].

Der originale NFI-MS Fragebogen stellt somit ein geeignetes Instrument zur Erfassung der Fatigue im klinischen Setting und in der Forschung dar. Seine deutsche Fassung „Neurologischer Fragebogen zur Müdigkeit bei Multipler Sklerose“ (NFI-MS-G) [11] (©The University of Leeds 2010) wurde bisher noch in keiner prospektiven Studie validiert. Daher war das primäre Ziel dieser Studie, die psychometrischen Kriterien des NFI-MS-G in einer größeren Stichprobe von MS-Patienten mittels Korrelationen mit etablierten Skalen und einer Rasch-Analyse zu evaluieren. Ein weiteres Ziel war, die Skala hinsichtlich ihrer Reliabilität und Invarianz bezüglich Gruppen unterschiedlichen Alters, Geschlechts, weiterer Personenmerkmale und englischer/deutscher Sprache zu untersuchen. Ziel war somit festzustellen, ob die englische NFI-MS und deutsche NFI-MS-G einander entsprechen.

Material und Methoden

Studiendesign und Studienpopulation

Diese Längsschnittstudie wurde prospektiv vom 12.2.2019 bis 15.06.2020 durchgeführt. Die Studie war Teil einer anderen Validierungsstudie [12] und wurde gemäß den Richtlinien der Deklaration von Helsinki (1964) durchgeführt. Studienzentren waren die Universitätsklinik für Neurologie, Medizinische Universität Innsbruck und das Reha Zentrum Münster, Österreich. Die Studie wurde durch die Ethikkommission der Medizinischen Universität Innsbruck genehmigt (EK Nr.: 1260/2018; 13.12.2018). Die Studie wurde prospektiv am 2.1.2019 im ISRCTN-Register registriert (ISRCTN14843579). Alle Patienten wurden ausführlich über den Studienzweck und die Studierendurchführung informiert und erteilten ihr schriftliches Einverständnis zur Teilnahme an der Studie und Publikation der anonymisierten Daten.

Es wurde eine Gelegenheitsstichprobe von Rehabilitanden mit einer gesicherten MS nach den McDonald-Kriterien [13] mit einem Expanded Disability Status Score (EDSS) [14] von 0 (keine Behinderung) bis 9,0 (schwere Behinderung) erhoben. Eine genauere Darstellung des Rekrutierungsvorgangs findet sich in [12]. Patienten wurden eingeschlossen, sofern sie deutschsprachig und ≥ 18 Jahre alt waren. Patienten wurden ausgeschlossen, wenn sie eine Komorbidität aufwiesen, die einen Einfluss auf die Fatigue haben könnte (z. B. maligne, andere neurologische oder psychiatrische Erkrankun-

gen), einen MS-Schub innerhalb 8 Wochen oder eine Veränderung ihrer Medikation innerhalb von 4 Wochen vor Studienbeginn hatten.

Die demografischen (Alter, Geschlecht) und krankheitsspezifischen Daten wie MS-Phänotyp, Krankheitsdauer, Medikation (Disease Modifying Treatment, DMT), klassifiziert in kein DMT, moderat und hoch effektive DMT [15] der Patienten wurden den Krankenakten entnommen. Der EDSS wurde von Neurologen beurteilt.

Zu validierendes Instrument

Der NFI-MS-G-Fragebogen wird von Patienten subjektiv beurteilt und ausgefüllt, und seine Durchführung benötigt etwa 5–10 Minuten. Die physische (Items 1–8) und kognitive (Items 9–12) Subskala wird jeweils durch Aufsummieren der Itemrohwerte berechnet und diese werden als eindimensional betrachtet [7]. Gemäß den Ergebnissen der NFI-MS-Originalstudie kann zudem eine Summenskala durch Addition der Items 1–7, 9 und 11–12 berechnet werden. Abstufungen der Antwortkategorien erfolgen mittels einer 4-Punkte-Likert-Skala (von 0 = „stimme überhaupt nicht zu“ bis 3 = „stimme vollkommen zu“). Höhere Skalenwerte bedeuten eine stärker ausgeprägte Fatigue.

Von den Testentwicklern (University of Leeds und The Walton Centre, Liverpool, UK) [7] wurde die schriftliche Genehmigung zur Validierung des NFI-MS-G eingeholt. Die deutschsprachige NFI-MS (Fragebogen zur Müdigkeit, NFI-MS-G) wurde von Mapi Research Trust (Lyon, Frankreich) übersetzt (Version vom 30.9.2013 – Mapi.ID7555/NFI-MS_AU1.0_deu-AT.doc) und kann über diesen bezogen werden. Zur Testung der Invarianz hinsichtlich der Sprache wurden der österreichische (N = 309) und englische (N = 318; Validierungsstichprobe aus Mills et al. (2010) [7]) Datensatz kombiniert. Die Datenerhebung wurde zu zwei Messzeitpunkten (T1, T2) innerhalb eines 14–21-Tage-Zeitraums durchgeführt.

Lesbarkeit

Der Lesbarkeitsindex LIX nach Björnson wurde berechnet: $LIX = (\text{Anzahl der Wörter}/\text{Anzahl der Sätze}) + (\text{Anzahl langer Wörter}/\text{Anzahl der Wörter})$. Lange Wörter sind als solche mit > 6 Buchstaben definiert [16]. Indexwerte > 60 deuten auf einen sehr schwer lesbaren Text hin (Fachliteratur), 50–60 auf einen schweren (Sachliteratur), 40–50 auf einen mittelschweren (Belletristik) und < 40 auf einen leichten bis sehr leichten Text (Kinder- und Jugendliteratur) hin.

Assessmentinstrumente zur Überprüfung der konvergenten Validität

Zur Untersuchung der konvergenten Validität des NFI-MG-G wurden validierte deutschsprachige Assessmentinstrumente verwendet. Diese inkludierten die deutsche Fassung der Hospital Anxiety Depression Scale (HADS-D) [17], den Internationalen Fragebogen zur Erfassung der LQ bei Multipler Sklerose (MusiQol) [18] und die Resilienzskala (RS-11) [19]. Diese Assessmentinstrumente wurden gewählt, da Angst, Depression und Resilienz einen signifikanten Einfluss auf die Fatigue bei Patienten mit MS haben [6, 20], während diese die LQ beeinflusst [20]. Basierend darauf wurden die im Folgenden dargestellten Hypothesen für die Evaluierung der konvergenten Validität aufgestellt. Um einen Reihenfolgeeffekt zu vermeiden, wurden die Fragebögen in einer zufälligen Reihenfolge von den Teilnehmern ausgefüllt.

Die HADS-D ist ein etablierter Fragebogen zur Erfassung von Angst (0–21 Punkte) und Depression (0–21 Punkte), die anhand einer 4-Punkte-Likert-Skala von Patienten beurteilt wird. Höhere Werte deuten auf eine stärker ausgeprägte Symptomatik hin. Die HADS-D weist sehr gute psychometrische Eigenschaften auf [17]. Mehrere Studien wiesen auf einen Zusammenhang zwischen Depression und Fatigue hin [6, 20]. Daher wurde für die vorliegende Studie eine moderate positive Korrelation der HADS-D und dem NFI-MS-G erwartet.

Der international empfohlene MusiQoL dient der Erfassung von gesundheitsbezogener LQ bei MS-Patienten. Die Antwortkategorien der Items rangieren auf einer Likert-Skala von 1 („niemals/keineswegs“) bis 5 („immer/sehr viel“). Die originale und die deutsche Fassung zeigten eine für alle 9 Subskalen und den Indexwert eine zufriedenstellende konvergente und diskriminante Validität sowie eine akzeptable Reliabilität [18]. MS-bezogene Fatigue beeinträchtigt die LQ [20], weshalb für die aktuelle Studie eine geringe bis moderate negative Korrelation zwischen MusiQoL und NFI-MS-G erwartet wurde.

Die Resilienzskala (RS-11) misst Resilienz, die psychische Widerstandskraft, welche es ermöglicht, schwierige Lebenssituationen schadlos zu überstehen. Die Beantwortung des RS-11 erfolgt anhand einer 7-Punkte-Likert-Skala (von „ich stimme nicht zu“ bis „ich stimme völlig zu“), wobei höhere Werte eine stärker ausgeprägte Resilienz repräsentieren. Für die RS-11 wurde eine gute Validität und Reliabilität nachgewiesen [19]. Im klinischen Alltag wird die Resilienz aufgrund des meist schubhaft-chronischen Verlaufs der MS als ein wichtiger psychischer Schutzfaktor betrachtet. Für die vorliegende Studie wurden moderate negative Korrelationen zwischen der RS-11 und dem NFI-MS-G erwartet.

Stichprobengröße

Basierend auf den Empfehlungen relevanter Studien [21] war das Ziel, mindestens 250 MS-Patienten mit einem breiten Spektrum an Fatigue zu rekrutieren.

Datenanalyse

Interne Konstruktvalidität

Eine Rasch-Analyse wurde zur Validierung des NFI-MS-G eingesetzt. Die Modellannahmen des polytomen Rasch-Modells (wie Eindimensionalität und lokale stochastische Unabhängigkeit zwischen Items) wurden überprüft [10], um die für eine intervallskalierte Messung notwendigen Annahmen zu testen. Das bedeutet, dass die Skalenwerte sich auf einer Logit-Skala befinden (Logit = natürlicher Logarithmus einer Chance) [22].

Mittels des Likelihood Ratio Test wurde geprüft, ob unter den polytomen Rasch-Modellen das komplexere Partial Credit Modell (PCM) [23] besser zu den Daten passt als das sparsamere Rating Scale Modell [24]. Das Ergebnis des Tests war statistisch signifikant [$p = 0,001956$], daher wurde für alle folgenden Berechnungen das PCM verwendet. Die Prüfung der Eindimensionalität (ED) und lokalen Unabhängigkeit (LU) erfolgte über die Prüfung der Residuen. Für ED wurden mehrere t-Tests durchgeführt, wobei die untere Grenze des 95%-Konfidenzintervall der Anzahl signifikanter t-Tests 5% nicht überschreiten sollte [25] (Details siehe Online-Supplement-1). Für LU wurde die Korrelationsmatrix aller Item-Residuen berechnet (Pearson-Korrelation). Korrelationen von $>0,2$ über der

mittleren Korrelation der Gesamtmatrix werden als Abweichungen lokaler Unabhängigkeit interpretiert [26]. Zur Analyse von Differential Item Functioning (DIF), d. h. der Abhängigkeit der Itemparameter von der Zugehörigkeit zu einer bestimmten Gruppe, wurden grafische Modelltests sowie Varianzanalysen der Residuen durchgeführt [27, 28]. Nicht-signifikante Ergebnisse deuten bei Abwesenheit von DIF hin. Für folgende Gruppen wurden DIF-Analysen durchgeführt: (Geschlecht [weiblich, männlich], Alter [Quartilgruppen], Krankheitsdauer [Quartilgruppen], Testzeitpunkt [Test 1/2], Zentrum [Innsbruck, Münster] und Sprache [Englisch, Deutsch]). Falls ED oder LU verletzt waren, wurden Items zur Super-Items zusammengefasst und eine Subtest-Analyse durchgeführt. Diese entspricht der Modellierung der LU durch ein restringiertes Bifaktor-Modell (Testlet-Modell) (siehe **Online-Tab. 1–4** und **Online-Dateien 1–2**). Da es sich um eine bereits vorhandene Skala handelt und in dieser Studie die psychometrischen Eigenschaften in einer österreichischen Stichprobe getestet werden, wurde erwartet, dass dieselben Subskalen der NFI-MG-G dieselbe Passung zum Rasch-Modell zeigen, wie für die NFI-MS berichtet. Es wurde erwartet, dass Patienten unterschiedlicher Gruppen (Geschlecht, Alter, Krankheitsdauer und Zentrum) und Sprache die Schwierigkeit derselben Items in gleicher Weise wahrnehmen (Invarianz). Die Rasch-Analyse wurde mit RUMM2030 Software (RUMM Laboratory, Perth, Australia) durchgeführt.

Konvergente und Known-Groups-Validität

Zur Berechnung der konvergenten Validität zwischen dem NSI-MS-G und anderen relevanten PROMs wurden Rangkorrelationskoeffizienten nach Spearman (r_s) berechnet. A priori wurden moderate ($r_s = 0,5–0,69$) [29]) positive Korrelationen der NFI-MS-G-Skalen mit der HADS und geringe ($r_s = -0,3$ bis $-0,49$) [29] bis moderate negative Korrelationen mit dem MusiQoL und der RS-11 erwartet.

Zur Überprüfung der Known-Groups-Validität wurden Subgruppen des Geschlechts (Frauen, Männer), Krankheitsverlaufs (schubhaft, progredient) und Behinderungsgrads (EDSS 0–4 und 4,5–9) mittels Mann-Whitney-U-Tests und medianen Hodges-Lehmann-Unterschieden für unabhängige Stichproben (95% Konfidenzintervall, KI) miteinander verglichen. Zur Überprüfung des Einflusses des Alters auf die Fatigue anhand der NFI-MS-G wurden r_s (95% KI) mit zweiseitigem p-Wert berechnet. Es wurden keine signifikanten Unterschiede der Fatigue zwischen Frauen und Männern erwartet, aber höhere Fatiguewerte bei Teilnehmern mit einem progredienten Krankheitsverlauf, stärkerer Behinderung und höheren Alters.

Die deskriptive Statistik (absolute und relative Häufigkeiten, Mittelwert (Minimum, Maximum) und Median (Spannweite) für nominale, metrische und ordinale Daten) und Berechnungen der konvergenten sowie Known-Groups-Validität wurden mit IBM SPSS Software (IBM SPSS Statistics; Version 26.0. Armonk, NY: IBM Corp.) oder GraphPad Prism Version 8 (GraphPad Software, La Jolla, CA) durchgeführt. Die statistische Signifikanz wurde als zweiseitiger p-Wert $<0,05$ definiert.

Reliabilität

Die interne Konsistenz der NFI-MS-G wurde mittels Cronbachs Alpha und des Person Separation Index (PSI) evaluiert. Die Test-Retest-Reliabilität als Maß der Übereinstimmung zwischen T1 und T2 wurde anhand des Konkordanz-Koeffizienten nach Lin (r_c) mit

einem 95 % KI [30] berechnet. Außerdem wurden die medianen Werte der NFI-MS-G-Subskalen und die Summenskala zu T1 und T2 kalkuliert und ihre Unterschiede mittels Hodges-Lehmann-Unterschieden für verbundene Stichproben im Zuge eines Wilcoxon-Vorzeichen-Rang-Tests untersucht. Der Pearson-Korrelationskoeffizient wurde als Messung der Präzision berechnet, und ein Bias-Korrekturfaktor, C_b als eine Messung der Genauigkeit [30]. MedCalc Software (MedCalc Software Ltd, Ostend, Belgien) wurde zur Bestimmung des r_c verwendet.

Der Standardmessfehler (SEM) und der auf einem 95 % KI basierende Minimum Detectable Change (MDC) zur Messung der Präzision des NFI-MS-G sowie Boden- und Deckeneffekte wurden berechnet (für Details siehe **Online-Tab. 4**).

Ergebnisse

Studienpopulation

Insgesamt wurden 623 potenzielle Patienten im Rahmen ihrer ambulanten oder stationären Rehabilitation über die Studie informiert, von denen 309 Patienten (49,6 %) einer Studienteilnahme zustimmten. 290 Teilnehmer (93,8 %) füllten den NFI-MS ein zweites Mal aus, und 19 Teilnehmer schieden aus der Studie aus (MS-Schub, N = 2; nicht mehr erreichbar, N = 11; schlechter Gesundheitszustand, N = 6). Die Charakteristika der Studienpopulation sind in **Tab. 1** dargestellt.

Tab. 1 Charakteristika der Studienpopulation.

Parameter	N = 309
Geschlecht ¹	Männer: 72 (23,3 %)
	Frauen: 237 (76,7 %)
Alter ²	50,2 (19,9-86,4) Jahre
Krankheitsdauer ²	18,2 (0,1-47,2) Jahre
Krankheitsverlauf ¹	Schubhaft: 194 (62,8 %)
	Primär progredient: 36 (11,7 %)
	Sekundär progredient: 79 (25,6 %)
EDSS ³	3,0 (0-9)
EDSS Gruppen ¹	EDSS * Punkte 0-4,0: 205 (66,3 %)
	EDSS * Punkte 4,5-6,5: 86 (27,8 %)
	EDSS * Punkte 7-9: 18 (5,8 %)
Krankheitsmodifizierende Therapie (Disease modifying treatment, DMT) ¹	Kein DMT: 163 (53 %)
	Moderat effektives DMT * * : 61 (19,7 %)
	Hoch effektives DMT * * : 85 (27,5 %)

¹Häufigkeiten (Prozentsatz); ²Mittelwert (Minimum-Maximum); ³Median (Minimum-Maximum); * EDSS = Expanded Disability Status Scale; * * Moderat effektive DMTs: Interferon-b 1a und 1b, pegyliertes Interferon-b 1a, Glatiramerazetat, Dimethylfumarat, Teriflunomid, Azathioprin, intravenöse Immunglobuline; (c) hoch effektive DMTs: Alemtuzumab, Cladribin, Fingolimod, Natalizumab, Ocrelizumab, Cyclophosphamid, Mitoxantron, Rituximab.

Lesbarkeit

Der LIX nach Björnson lag bei 37,3 Punkten, was auf eine geringe Komplexität der Items hinweist. Die Lesbarkeit des NFI-MS-G entspricht damit jener von Jugendliteratur.

Interne Konstruktvalidität

Die NFI-MS-G Daten und der kombinierte Datensatz aus NFI-MS-Originaldaten und NFI-MS-G-Daten wurden separat zur Schätzung der Modellparameter verwendet, sodass eine Passung des Modells erfolgen konnte. Die Analyse erfolgte für alle individuellen Skalenelemente und als Summenstatistik.

Die physische und kognitive Komponente zeigten eine gute Passung zum Modell, Eindimensionalität, geordnete Schwellenwerte sowie nur ein vernachlässigbares DIF für das Geschlecht bzw. die Krankheitsdauer bei jeweils einem Item der kognitiven Komponente (siehe **Online-Tab. 1-2**). Die Invarianz der Skala, gleichbedeutend mit einem Fehlen von DIF, wurde anhand nicht signifikanter Varianzanalyse (ANOVA) Fit-Statistiken der Residuen aus einer Interaktion von Personenfaktoren und Klassenintervallen gezeigt. Eine visuelle Inspektion der Item Characteristic Curves (ICC) bestätigte diese Diagnose (siehe Online-Datei 1 für Details). Nach der Kombination von Items der physischen Komponente zu Superitems wurden keine weiteren lokalen Abhängigkeiten mehr nachgewiesen. Die NFI-MS-G-Summenstatistik zeigte ein geringgradiges „Modell-Misfit“ bezogen auf die Eindimensionalität der physischen Subskala, da das untere Limit des 95 %-igen KI bezogen auf die Anzahl signifikanter t-Tests 5 % überstieg (**Tab. 2**). Außerdem wurde eine geringgradige lokale Abhängigkeit beobachtet, welche mittels Kombination einzelner Items zu Subtests adressiert wurden.

Somit gelten die Annahmen des PCM als erfüllt. Die detaillierten Resultate der Modellanpassung und ideale Werte sind in **Tab. 2** dargestellt. Ein Pfaddiagramm der angewandten Modellstruktur wird in **Abb. 1** gezeigt.

Aufgrund der beschriebenen geringgradigen lokalen Abhängigkeiten im deutschen, aber nicht im englischen Datensatz war eine Überprüfung der Invarianz hinsichtlich der Sprache nicht möglich.

Konvergente und Known-Groups-Validität

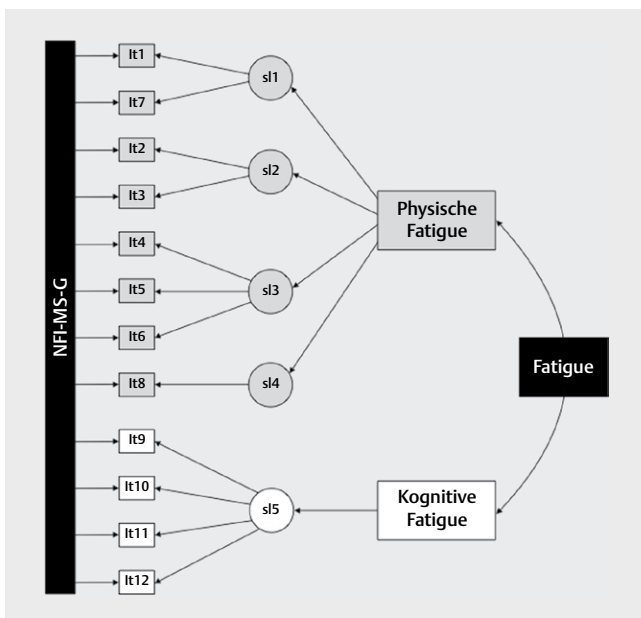
Die Rangkorrelationskoeffizienten nach Spearman zur Messung der konvergenten Validität zeigten geringe bis moderate Korrelationen der NFI-MS-G-Skalen mit der HADS-D, RS-11 und dem MusiQoL. In Übereinstimmung mit den Studienhypothesen war das Verhältnis mit der HADS-D positiv und mit dem MusiQoL und der RS-11 negativ (**Online-Tab. 5**).

Es wurden keine statistisch signifikanten Unterschiede hinsichtlich Fatigue zwischen Frauen und Männern gefunden. Die physische Fatigue war bei Patienten mit progredientem Krankheitsverlauf und einer schwereren Behinderung signifikant höher. Bezüglich der kognitiven Fatigue gab es keine signifikanten Unterschiede zwischen den MS-Verläufen, aber höhere Werte bei schwerer behinderten Patienten. Die physische und Gesamt-Fatigue stieg mit dem Alter signifikant, während dieser Effekt bei der kognitiven Fatigue geringer war. Die Unterschiede zwischen den genannten Gruppen sind in **Tab. 3** dargestellt.

► **Tab. 2** Modellanpassung an das Rasch-Modell und Reliabilität der NFI-MS-G.

Analyse	Standardisierte Fit Residuen Items		Standardisierte Fit Residuen Personen		Chi-Quadrat-Test ¹		PSI ²	Alpha	Eindimensionalität ³	Extremwerte ⁴	
	MW	STD	MW	STD	Wert (df)	p				% Tests > 5% (95% KI)	Bodeneffekte
NFI-MS-G											
12-Item-Skala											
Physisch	-0,13	1,01	-0,49	1,40	29,70 (32)	0,58	0,90	0,92	6,3 (3,7–8,8)	4,9	1,9
Kognitiv	0,13	0,57	-0,35	0,99	14,27 (16)	0,58	0,80	0,85	2,3 (-0,2–4,7)	7,1	1,6
Summens.	-0,07	1,47	-0,50	1,65	46,19 (40)	0,23	0,91	0,93	11,4 (8,9–13,8)	2,9	0,3
Subtest der Physischen Komponente – Kombination von Items 1 & 7, 2 & 3 und 4 & 5 & 6 (Item 8 wurde als individuelles Item behalten)											
Physisch	0,26	1,37	-0,38	1,00	14,32 (16)	0,58	0,86	0,84	4,9 (2,3–7,4)	4,9	1,9
Subtest-Analyse der beiden Superitems der Physischen und Kognitiven Komponente										R	A
2 Super-items	-0,03	2,98	-0,5	0,93	2,46 (8)	0,96	0,80	0,72	2,3 (0–4,8)	0,96	0,87
Gesamtanalyse der Superitems der Physischen (Items 1 & 7, 2 & 3 und 4 & 5 & 6 & Item 8) und der Kognitiven Komponente (Items 9–12)										R	A
Summens.	0,37	1,25	-0,39	1,03		0,511	0,86	0,87	2,3 (-0,2–4,7)	0,81	0,94
Idealwerte	0,00	1,00	0,00	1,00		>0,05*	>0,85	>0,70	UKI < 5		

¹Der Chi-Quadrat-Test wird für die Item-Merkmal-Interaktion verwendet; ²Der PSI gibt die Reliabilität und Differenzierung von Strata (Untergruppen) an; ³Basierend auf t-Tests für unabhängige Stichproben, zum Vergleich der auf der ersten Hauptkomponente positiv und negativ geladener Personenresiduale (mit 95% KI); ⁴Boden- und Deckeneffekte sind Prozentsätze; * Bonferroni-adjustiert und abhängig von der Anzahl der Items; A: gemeinsame Varianz der physischen und kognitiven Komponente; Alpha: Cronbachs Alpha; df: Freiheitsgrade, Degrees of Freedom; Kond.: konditional; KI: Konfidenzintervall; MW = Mittelwert; PSI: Person Separation Index; R: latente Korrelation zwischen der physischen und kognitiven Komponente; STD = Standardabweichung; UKI: untere Grenze des 95% KI.

► **Abb. 1** Pfaddiagramm des angewandten Testlet-Modells.

Reliabilität

Es wurde eine sehr gute interne Konsistenz der beiden NFI-MS-G-Skalen festgestellt. Die Cronbachs-Alpha-Werte lagen zwischen 0,84 und 0,93 und der Personen-Separations-Index (PSI) bei 0,81–0,91, wobei die kognitive Subskala geringgradig unterhalb der angestrebten 0,85 für den PSI lag. Ein Konkordanzkorrelations-Koeffizient nach Lin zwischen 0,86 bis 0,90 wies auf eine ausgezeichnete Test-Retest-Reliabilität hin.

Die Analyse zeigte eine hohe Messgenauigkeit mit Boden- und Deckeneffekten, abgesehen von Bodeneffekten der kognitiven Subskala von 7,1%. Der SEM für die physische und kognitive Subskala lag bei 0,59 und 0,82. Der MDC betrug 16,7% bzw. 4/24 Punkte für die physische und 23,5% bzw. 2,8/12 Punkte für die kognitive Subskala. Für die Summenskala wurde ein SEM von 0,52 und ein MDC von 13,8% bzw. 4,1/30 Punkten festgestellt. Weitere Details zur NFI-MS-G-Reliabilität werden in den **Online-Tab. 4 und 6** dargestellt.

Aufgrund der erfüllten Voraussetzung der Modellanpassung an das Rasch-Modell ist eine Transformationstabelle der Rohwerte in intervallskalierte Werte verfügbar (**Online-Tab. 7**). Diese linearen Werte sind nur bei Vorliegen eines kompletten Datensatzes erhältlich.

► **Tab. 3** Known-Groups-Validität des NFI-MS-G.

NFI-MS-G	Physische Subskala	Kognitive Subskala	Summenskala
Männer/Frauen ¹	0 [-2 bis 1]	0 [-1 bis 0]	-1 [-3 bis 1]
Schubhafter/progredienter MS-Verlauf ¹	-3 [-4 bis -1] ***	-1 [-1 bis 0]	-3 [-5 bis -1] ***
Geringe/mittlere bis schwere Behinderung ¹	-4 [-5 bis -3] ***	-1 [-2 bis 0] *	-4 [-6 bis -3] ***
Alter ²	0,25 [0,14 bis 0,35] ***	0,13 [0,02 bis 0,24] *	0,22 [0,11 bis 0,33] ***

¹Mediane Hodges-Lehmann-Unterschiede für unabhängige Stichproben (95 % Konfidenzintervall), berechnet mittels Mann-Whitney-U-Tests; es wurde untersucht, ob die Unterschiede zwischen den Gruppen (z. B. Männer/Frauen) signifikant sind; ²Rangkoeffizienten nach Spearman (95 % Konfidenzintervall); n.s. = nicht signifikant; * p < 0,05, ** p < 0,01 oder *** p < 0,001 (2-seitig, ¹p-Werte korrigiert für 4 Vergleiche); Expanded Disability Status Scale, EDSS 0–4 und 4,5–9.

Diskussion

In dieser Studie wurde das Rasch-Theoriemodell zur Validierung der deutschen Fassung des NFI-MS in einer größeren Stichprobe MS-Betroffener angewandt. Für die deutschsprachige MS-Population aus Österreich wurden ähnliche Ein- und Ausschlusskriterien gewählt wie in den ursprünglichen Entwicklungs- und Validierungsstudien. Die österreichische Stichprobe repräsentierte hinsichtlich des Krankheitsverlaufs die MS-Population Europas [31] und ähnelte hinsichtlich der Größe (N = 309) jener der englischen Entwicklungs- (N = 317) und Validierungsstichproben (N = 318) [7]. Rund 50 % der geeigneten Patienten stimmten einer Studienteilnahme zu und entspricht damit der Zustimmungsrate von 51,9 % bei der englischen NFI-MS-Validierungsstudie.

Die Studienergebnisse bestätigten die Hypothesen bezüglich der Eindimensionalität und Invarianz der physischen NFI-MS-G-Subskala für Gruppen unterschiedlichen Geschlechts, Alters, Krankheitsdauer und Testzentren. Ähnliche Ergebnisse zeigte die kognitive Subskala. Die Hypothese hinsichtlich der Eindimensionalität der Summenskala wurde widerlegt. Allerdings wurden für die physische Subskala geringfügige lokale Abhängigkeiten beobachtet, primär zwischen Item 1 („Ich bin schnell müde geworden“) und Item 7 („Normalerweise bin ich an den meisten Tagen müde geworden“). Aufgrund dieser ähnlichen Formulierungen wurden die abhängigen Items jeweils zusammengefasst und Item 8 als eigenständiges Item beibehalten.

Die Summenskala sollte ursprünglich einen einzelnen Gesamtwert zur Darstellung des übergeordneten Konstrukts der Fatigue bilden, der sich aus Items der physischen und kognitiven Komponente zusammensetzt. Diese zeigte sich innerhalb der österreichischen Stichprobe nicht als eindimensional. Die Autoren schlagen daher vor, die physischen und kognitiven Komponentenwerte beizubehalten und für die österreichische Skala getrennt zu behandeln. Falls jedoch ein einziger Summenwert erforderlich ist, kann dieser der 2-Komponenten-Subtestanalyse entnommen werden (**Online-Tab. 7**), die den 12-Item-Rohwert auf eine beschränkte gemeinsame Metrik umparametrisiert. Die Anpassung an das Rasch-Modell rechtfertigt die Transformation der Rohwerte in intervallskalierte Werte, die für eine parametrische Analyse geeignet sind.

Da ein Pooling des englischen und österreichischen Datensatzes zur Überprüfung der Invarianz bezüglich der Sprache nicht möglich war, müssen die beiden NFI-MS-Sprachfassungen individuell betrachtet werden.

Die Studienergebnisse bestätigten außerdem die Hypothesen hinsichtlich der Zusammenhänge des NFI-MS-G mit den HADS-D-Subskalen sowie den MusiQoL-Skalen und dem RS-11. Eine mittlere Test-Retest-Reliabilität und interne Konsistenz der physischen und kognitiven Subskalen wurden gefunden [32]. PSI-Werte von 0,90 und 0,81 für die physische und kognitive Subskala geben an, dass der NFI-MS-G zwischen ≥ 3 bzw. ≥ 4 Fatigue-Graden differenzieren kann [33].

Eine adäquate Präzision des NFI-MS-G konnte für die NFI-MS-G gezeigt werden. Der MDC für die physische Subskala betrug 16,7 % oder 4,0 von 24 Punkten. Der MDC für die kognitive Subskala betrug 23,5 % bzw. 2,8 von 12 Punkten, was relativ hoch ist. Für die Summenskala wurde ein MDC von 13,8 % bzw. 4,1 von 30 Punkten festgestellt. Änderungswerte, welche den MDC und damit die Messfehler übertreffen, weisen auf wahre Veränderungen hin. Geringe Boden- und zu vernachlässigende Deckeneffekte deuteten darauf hin, dass die NFI-MS-G zwischen unterschiedlichen Fatigue-Graden bei MS Patienten differenzieren kann und auch am jeweiligen Ende des Spektrums änderungssensitiv ist [34].

Eine Limitation dieser Studie ist die fehlende cross-kulturelle Adaptierung des NFI-MS-G für Österreich. Die deutschsprachige Fassung lag bereits vor, und der Lesbarkeitsindex wies auf eine sehr leichte Lesbarkeit hin, sodass auf Interviews mit MS-Patienten verzichtet wurde. Die 12 Items erschienen dem österreichischen Studienteam sehr gut verständlich. Vergleichbar mit weiteren Fatigue-Skalen [5] erwies sich die Durchführung des NFI-MS-G zudem als einfach und rasch. Dies betrachten wir gerade für MS-Betroffene mit Fatigue als eine wichtige Skaleneigenschaft. Eine weitere Limitation ist, dass zur Schätzung der Test-Retest-Reliabilität auch die Daten von Rehabilitanden in niederfrequenter ambulanter Rehabilitation inkludiert wurden. Auch wenn eine Prüfung der Fatigue-Werte im Test-Retest-Vergleich keine Veränderungen zeigte, könnte die Rehabilitation eine Fatigue-Verschlechterung verhindern haben.

Die vorliegenden Studienergebnisse bestätigen die gute Validität und Reliabilität der NFI-MS auch für ihre deutsche Fassung.

Angesichts der hohen Prävalenz der Fatigue in der MS-Population und dem Bedarf an akkuraten Messungen dieses herausfordernden Symptoms erscheint die NFI-MS-G als eine relevante Option zur Verwendung in der klinischen Praxis und Forschung. Darüber hinaus kann diese Skala für das Monitoring des MS-Krankheitsverlaufs und der sozialmedizinischen Folgen der Erkrankung eingesetzt werden. Dazu liegen bereits Studien mit der englischen NFI-MS-Originalskala vor (vgl. [35] für einen Review).

KERNBOTSCHAFT

Der Neurological Fatigue Index – Multiple Sclerosis (NFI-MS) wurde auf Basis von Interviews MS-Betroffener und den Methoden der Rasch-Analyse und Faktorenanalyse entwickelt und validiert. In der vorliegenden Studie wurde seine deutsche Fassung NFI-MS-G bezüglich ihrer Lesbarkeit überprüft und die Gütekriterien der Validität und Reliabilität untersucht. Unter Berücksichtigung der genannten Limitationen hinsichtlich des 2-Domänen-Subtests zeigten die Ergebnisse gute psychometrische Eigenschaften des NFI-MS-G, der damit für die klinische Praxis und Forschung verfügbar ist und kostenfrei von den Autoren bezogen werden kann. Die NFI-MS- und NFI-MS-G-Datensätze sollten nur nach Anpassungen der Skalenwerte bezüglich lokaler Abhängigkeiten der deutschen Fassung kombiniert werden.

Danksagung

Wir danken den Teilnehmern dieser Studie herzlich.

Förderung

Diese Studie wurde von der Österreichischen MS Forschungsgesellschaft unterstützt [keine Fördernummer].

Interessenkonflikt

Die Autorinnen/Autoren geben an, dass kein Interessenkonflikt besteht.

Literatur

- [1] Compston A, Confavreux C, Lassmann H et al. *McAlpine's multiple sclerosis*. 4th ed. Aufl. London: Elsevier; 2006
- [2] Olbert E, Seiser A, Struhal W. Fatigue bei Multipler Sklerose. *psychopraxis neuropraxis* 2019; 22: 52–57. doi:10.1007/s00739-019-0548-9
- [3] Guidelines MSCP. *Fatigue and multiple sclerosis: evidence-based management strategies for fatigue in multiple sclerosis*. Washington D.C.: Paralyzed Veterans of America; 1998
- [4] Abd-Elfattah HM, Abdelazeim FH, Elshennawy S. Physical and cognitive consequences of fatigue: A review. *Journal of Advanced Research* 2015; 6: 351–358. doi:10.1016/j.jare.2015.01.011
- [5] Elbers RG, Rietberg MB, van Wegen EE et al. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation* 2012; 21: 925–944. doi:10.1007/s11136-011-0009-2
- [6] Flachenecker P, Kumpfel T, Kallmann B et al. Fatigue in multiple sclerosis: a comparison of different rating scales and correlation to clinical parameters. *Multiple Sclerosis Journal* 2002; 8: 523–526. doi:10.1191/1352458502ms8390a
- [7] 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 and quality of life outcomes* 2010; 8: 22. doi:10.1186/1477-7525-8-22
- [8] 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: 502–505. doi:10.1177/1352458512457840
- [9] Wirtz M, Böcker M. Properties and benefits of applying the Rasch model in clinical diagnostics. *Die Rehabilitation* 2007; 46: 238–245. doi:10.1055/s-2007-971068
- [10] Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press; 1980
- [11] Krupp LB. Fatigue und M.S Im Internet: <https://www.msif.org/wp-content/uploads/2014/09/MS-in-focus-19-Fatigue-German.pdf> Stand: 18.10.2019
- [12] Seebacher B, Mills RJ, Reindl M et al. German translation, cultural adaptation and validation of the unidimensional self-efficacy scale for multiple sclerosis. *BMC Neurol* 2021; 21: 163. doi:10.1186/s12883-021-02183-y
- [13] 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: 121–127
- [14] Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an Expanded Disability Status Scale (EDSS). *Neurology* 1983; 33: 1444–1452
- [15] Wiendl H, Gold R, Berger T et al. Multiple Sclerosis Therapy Consensus Group (MSTCG): position statement on disease-modifying therapies for multiple sclerosis (white paper). *Therapeutic advances in neurological disorders* 2021; 14: 17562864211039648. doi:10.1177/17562864211039648
- [16] Lenhard W, Lenhard A. Berechnung des Lesbarkeitsindex LIX nach Björnson. In: *Bibergau: Psychometrica*. 2014-2017. doi:10.13140/RG.2.1.1512.3447
- [17] Petermann F. Hospital Anxiety and Depression Scale, Deutsche Version (HADS-D). *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie* 2011; 59: 251–253. doi:10.1024/1661-4747/a000077
- [18] 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. *Der Nervenarzt* 2011; 82: 1281–1289. doi:10.1007/s00115-011-3276-9
- [19] 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
- [20] Braley TJ, Chervin RD. Fatigue in multiple sclerosis: mechanisms, evaluation, and treatment. *Sleep* 2010; 33: 1061–1067
- [21] Linacre JM. Sample size and item calibration stability. *Rasch Measurement Transactions* 1994; 7: 328
- [22] Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007; 46: 1–18

- [23] Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982; 47: 149–174. doi:10.1007/BF02296272
- [24] Andrich D. A rating formulation for ordered response categories. *Psychometrika* 1978; 43: 561–573. doi:10.1007/BF02293814
- [25] 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 and rheumatism* 2007; 57: 1358–1362. doi:10.1002/art.23108
- [26] Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Appl Psychol Meas* 2017; 41: 178–194. doi:10.1177/0146621616677520
- [27] Holland PW, Wainer H. *Differential Item Functioning* Hillsdale, N. J. Lawrence Erlbaum; 1993
- [28] Andrich D. *Rasch models for measurement series: quantitative applications in the social sciences* no. 68. London: Sage; 1988
- [29] Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*. 5th ed. Aufl. Boston: Houghton Mifflin; 2003
- [30] Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; 45: 255–268
- [31] Kobelt G, Thompson A, Berg J et al. New insights into the burden and costs of multiple sclerosis in Europe. *Multiple Sclerosis Journal* 2017; 23: 1123–1136. doi:10.1177/1352458517694432
- [32] Bühner M. *Einführung in die Test- und Fragebogenkonstruktion*. 2. aktualisierte. Aufl. München: Pearson Studium – Psychologie; 2006
- [33] Wright BD, Masters GN. *Rating scale analysis*. Chicago, Illinois: MESA Press; 1982
- [34] Rammstedt B, Kemper C, Schupp J. Standardisierte Kurzskalen zur Erfassung psychologischer Merkmale in Umfragen. *Methoden, Daten, Analysen* 2013; 7: 145–152. doi:10.12758/mda.2013.009
- [35] Oliva Ramirez A, Keenan A, Kalau O et al. Prevalence and burden of multiple sclerosis-related fatigue: a systematic literature review. *BMC Neurology* 2021; 21: 468. doi:10.1186/s12883-021-02396-1