



Published in final edited form as:

Rehabil Psychol. 2012 May ; 57(2): 159–166. doi:10.1037/a0027890.

Comparison of the psychometric properties of two fatigue scales in multiple sclerosis

Dagmar Amtmann,

University of Washington, Rehabilitation Medicine, Box 354237, 4907 25th Ave NE, Seattle, WA 98105, Phone: 206 543-4741, Fax: 206 685-3244. dagmara@u.washington.edu

Alyssa M. Bamer,

University of Washington, Rehabilitation Medicine, Box 356490, Seattle, WA 98195, Phone: 303 953-8085, Fax: 206 685-3244. adigiaco@u.washington.edu

Vanessa Noonan,

University of Washington, Rehabilitation Medicine, Box 356490, Seattle, WA 98195, Phone: 604-707-2126 Fax: 604-707-2121. Vanessa.Noonan@vch.ca

Nina Lang,

University of Washington, Rehabilitation Medicine, Box 356490, Seattle, WA 98195, Phone: 206-221-2414, Fax: 206 685-3244. ninaclaire@gmail.com

Jiseon Kim, and

University of Washington, Rehabilitation Medicine, Box 356490, Seattle, WA 98195, Phone: 512 299-5991. jiseonk@u.washington.edu

Karon F. Cook

Northwestern University, Feinberg School of Medicine, 710 N. Lake Shore Dr. Suite 729, Chicago, IL 60611, Phone: 713 291-3918. karon.cook@northwestern.edu

Abstract

Objective—To compare psychometric functioning of the Fatigue Severity Scale (FSS) and the Modified Fatigue Impact Scale (MFIS) in a community sample of persons with multiple sclerosis (MS).

Research Method—A self-report survey including the FSS, MFIS, demographic and other health measures was completed by 1271 individuals with MS. Analyses evaluated the reliability and validity of the scales, assessed their dimensional structures, and estimated levels of floor and ceiling effects. Item response theory (IRT) was used to evaluate the precision of the MFIS and FSS at different levels of fatigue.

Results—Participants had a mean score on the FSS of 5.1 and of 44.2 on the MFIS. Cronbach's alpha values for FSS and MFIS were all 0.93 or greater. Known-groups and discriminant validity of MFIS and FSS scores were supported by the analyses. The MFIS had low floor and ceiling effects, while the FSS had low floor and moderate ceiling effects. Unidimensionality was supported for both scales. IRT analyses indicate the FSS is less precise in measuring both low and high levels of fatigue compared to the MFIS.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/rep

Conclusions—Researchers and clinicians interested in measuring physical aspects of fatigue in samples whose fatigue ranges from mild to moderate can choose either instrument. For those interested in measuring both physical and cognitive aspects of fatigue, and whose sample is expected to have higher levels of fatigue, the MFIS is a better choice even though it is longer. IRT analyses suggest both scales could be shortened without a significant loss of precision.

Keywords

Fatigue Severity Scale; Modified Fatigue Impact Scale; multiple sclerosis; psychometrics; Item Response Theory

Introduction

Fatigue has been defined as a “subjective lack of physical and/or mental energy that is perceived by the individual or caregiver to interfere with usual and desired activities,” (Multiple Sclerosis Council for Clinical Practice Guidelines [MSCCPG], 1998) and is among the most common and disabling symptoms reported by patients with multiple sclerosis (MS). Estimates of the prevalence of fatigue in persons with multiple sclerosis (MS) range from 53 to 92% (Branas, Jordan, Fry-Smith, Burls, & Hyde, 2000). A literature review of indexed studies measuring fatigue in MS between 2005 and 2010 identified 32 different self-report measures of fatigue. Of these, the most commonly used scales were the Fatigue Severity Scale (FSS) (Krupp, LaRocca, Muir-Nash, & Steinberg, 1989) and the Modified Fatigue Impact Scale (MFIS) (MSCCPG, 1998). The purpose of this study was to compare psychometric functioning of the FSS and MFIS in a sample of persons with MS. Specifically, we evaluated the reliability and validity of the scales, examined their dimensional structures, and assessed levels of floor and ceiling effects. In addition to the traditional psychometric analyses, we used an item response theory approach to evaluate the precision of the MFIS and FSS at different levels of fatigue (e.g., high versus low fatigue), a question that cannot be answered within the framework of Classical Test Theory (CTT).

Methods

Participants

Research participants were recruited through the Greater Washington chapter of the USA National Multiple Sclerosis Society (NMSS), which serves 23 counties in Washington State. Letters of invitation were sent to 7806 persons on the NMSS mailing list. Of the 1629 who responded to invitation letters, 1597 met eligibility criteria of being at least 18 years of age and reported having been diagnosed with MS. Eligible individuals were either mailed a self-report paper survey (n=1368) or directed to an online version of the same survey (n=229). Of these, 1271 individuals responded (80%). A short, anonymous demographics survey was sent to non-responders of the mailing list to assess possible recruitment bias. Responses received from 1046 non-responders indicated that 13% did not have MS despite being listed as persons with MS on the mailing list, and 34% did not recall receiving the initial survey invitation. Overall, the 1271 individuals who completed the study survey were similar on demographic variables to the non-responders except they were more educated (84% reported some college or more education compared to 72% of non-responders; $\chi^2=30.7, p<0.001$), slightly younger (53% of responders were 51 or older compared to 71% of non-responders; $\chi^2=70.3, p<0.001$), and had shorter mean disease duration (M=13, SD=10) than non-responders (M= 17, SD=12) [$\chi^2(2041)=8.00, p<0.0001$]. The study was approved by the Human Subjects Division at the University of Washington, and all participants provided written informed consent.

Instruments

Fatigue Measures—The FSS is comprised of nine items scored from 1 to 7 (1 = completely disagree, 7 = completely agree) (Krupp, et al., 1989). Scale scores are the mean of item scores, with lower scores indicating less fatigue. Respondents to the FSS are asked to consider the past week when choosing their answers. Items were chosen based on their ability to identify common features of fatigue in patients with MS and systemic lupus erythematosus. The item content of the FSS primarily focuses on the characteristics of fatigue (e.g., exercise brings on my fatigue, fatigue interferes with certain duties and responsibilities). All but one of the FSS items target physical aspects of fatigue. The remaining item is related to cognitive aspects of fatigue (motivation is lower when fatigued).

The MFIS is a shortened version of the Fatigue Impact Scale (Fisk et al., 1994) that contains 21 of the original 40 items. The measure was developed from interviews of individuals living with MS about how fatigue affects their daily activities and other life areas (Guidelines, 1998). In 1998, the Multiple Sclerosis Council for Clinical Practice Guidelines recommended the MFIS for use in clinical practice and research. The 21 items of the MFIS are scored from 0 to 4 (0 = Never, 1 = Rarely, 2 = Sometimes, 3 = Often, 4 = Almost always), and respondents are asked to consider their experiences with fatigue during the past four weeks. It can also be divided into three subscales: cognitive (10 items), physical (9 items) and psychosocial fatigue (2 items). The MFIS item content emphasizes symptoms of fatigue (e.g., muscles felt weak, needed to rest more often, clumsy).

Other Measures—The survey included measures of pain interference [Brief Pain Inventory (BPI)](Cleland & Ryan, 1994), pain severity and impact [Pain Impact Questionnaire (PIQ-6)](Becker, Schwartz, Saris-Baglama, Kosinski, & Bjorner, 2007), anxiety [7-item Hospital Anxiety and Depression Scale (HADS)](Zigmond & Snaith, 1983) and depression [Patient Health Questionnaire (PHQ-9)](Kroenke, Spitzer, & Williams, 2001). A question about MS clinical course [self-report graphical image questionnaire] (Bamer, Cetin, Amtmann, Bowen, & Johnson, 2007) was included as was the mobility subscale of the self-report version of the Expanded Disability Status Scale (EDSS)](Bowen, Gibbons, Gianas, & Kraft, 2001). Responses to the mobility subscale were used to categorize individuals into three groups: minimal (< 4.0), intermediate (4.5-6.5), and advanced (> 7.0) mobility impairment (Bowen, et al., 2001). Questions asking about demographics (e.g. age, gender), socioeconomic variables (e.g. employment) and the duration of MS also were included.

Analyses

Floor and Ceiling Effects—To evaluate floor and ceiling effects, the percentage of respondents with the highest and lowest scale scores were calculated for each of the fatigue scales.

Reliability Analyses—Classical reliability analyses included estimation of internal consistency (Cronbach's alpha) and item-to-scale correlations for the FSS, for the total MFIS score, and for two of the MFIS subscales (MFIS-Cognitive, MFIS-Physical). Because the MFIS-Psychosocial subscale only contains two items, these statistics were not calculated for it. Cronbach's alpha has an optimal range (0.7 to 0.9) of internal consistency or item homogeneity, but values over 0.9 indicate item redundancy.(Boyle, 1991) Item-to-scale correlations greater than 0.40 are typically interpreted as evidence of scale reliability. (Everitt, 2002, p. 208) Correlations were estimated using Spearman rank correlation coefficients.

Validity Analyses—The construct validity of the FSS and MFIS was assessed by examining associations among different measures. We hypothesized that scores on the FSS and MFIS would be moderately to highly correlated. The highest correlation was expected between FSS scores and MFIS-physical scores, because both target physical aspects of fatigue. Weaker associations were expected between fatigue scores and scores on measures assessing other health domains, including pain (BPI, PIQ-6), anxiety (HADS), and depression (PHQ-9). Correlations were estimated using Spearman rank correlation coefficients.

Known-groups validity was assessed by evaluating whether scores distinguished among subgroups that theoretically should differ in mean scores. The MS literature supports the hypothesis that individuals with greater mobility impairment experience greater fatigue. (Hadjimichael, 2008) One way analysis of variance (ANOVA) was used to test the hypothesis that scores for the fatigue scales would be different based on EDSS categories, with higher fatigue reported by participants with higher (i.e., worse) EDSS scores.

Dimensional Structure—To evaluate the dimensional structure of FSS and MFIS responses, we used confirmatory factor analysis (CFA) to fit a model in which all items loaded on a single factor (unidimensional model). Unidimensionality is inherently assumed when summary scores are obtained using all items of the scale. These analyses were conducted using MPLUS 5.21. (Muthen & Muthen, 2009) Fit of the unidimensional model was evaluated based on the comparative fit index (CFI). (Hu & Bentler, 1999) A CFI of 0.90 or greater has been suggested as a criterion for acceptable fit. (Hu & Bentler, 1999)

When the criterion for acceptable fit with a unidimensional model is not met, an alternative factor model is McDonald's bifactor model. (Reise, Morizot, & Hays, 2007) In this model, all items load on a single, general factor. In addition subsets of items are identified either empirically or theoretically that are expected to load on sub-dimensions (group factors). Because the MFIS has subscales, we fit a bifactor model in which all items loaded on general fatigue, cognitive items loaded on one group factor, and all other MFIS items loaded on a second group factor, i.e., the two psychosocial items were grouped with the physical items. The two psychosocial items were grouped with the physical items, because their content was more similar to items in that subdomain than to the psychosocial items, and this grouping was also supported by an exploratory factor analysis. Modeling the MFIS data using a bifactor model allowed us to estimate the amount of variance accounted for by the subscales compared to the variance accounted for by the overall fatigue factor.

The bifactor analyses also served a second purpose. One of the assumptions of Item Response Theory (IRT) analyses is unidimensionality. When factor loadings on the general factor are greater than 0.30 and the general factor accounts for more variance than do the group factors, then unidimensional IRT models can be reliably applied. (McDonald, 1981)

IRT Analyses—FSS and MFIS item responses were modeled using the graded response model (GRM) (Samejima, 1969), a model appropriate for items with more than two response options. Based on this model we calculated “information” for each scale and subscale. Information is the equivalent of reliability estimates in classical methods. Its chief advantage is that values are estimated for every level of the trait being measured. CTT reliability statistics generate a single value for an entire scale. This obscures the fact that a scale typically measures different levels of trait with different levels of precision. Scale information was plotted along with the distributions of MFIS and FSS scores. This graphical display provides a picture of a scales relative precision within the study sample, and we have included reference lines to indicate where the scales measure with reliability greater than 0.80 or 0.90.

Results

Descriptive Analyses

Of the 1271 individuals participating in the study, 80% (n=992) were women, most were either married or living with a significant other (n=867; 70.3%), and 36.2% (n=447) reported being employed 20 or more hours a week. Participants had a mean age of 50.7 (SD=11.6; range 18-88) and mean disease duration of 13.2 (SD=10.1; range 0-60) years. The most common type of MS reported was relapsing remitting (n=700; 58.5%). Based on the mobility subscale of the EDSS, severity of MS was categorized as minimal for 32.4% of the sample (EDSS 4.0), intermediate for 47.9% (EDSS 4.5 - 6.5) and advanced for 19.7% (EDSS 7.0). The sample was similar to MS community samples in published studies with the exception of our sample having a higher proportion of women (81% versus 64% (Kos et al., 2005) and 81% versus 70.4% (Mills, Young, Nicholas, Pallant, & Tennant, 2009)). Demographic information and disease characteristics are displayed in Table 1.

Participants had a mean score on the FSS of 5.1 (SD=1.5) and of 44.2 (SD=18.2) on the MFIS. Mean scores on all variables included in the analyses are listed in Table 2.

Floor and Ceiling Effects

We calculated the percentage of with the lowest (floor effect) or the highest (ceiling effect) possible scores on the FSS and MFIS measures (see Table 2). The FSS had low floor effects (0.9%), but higher ceiling effect (6.8%). The floor effects for MFIS-Total scores were comparable to those of the FSS (1.1%), but had a much smaller ceiling effects (0.7%) compared to the FSS. As expected with a two-item subscale, the MFIS-psychosocial subscale had the largest floor (7.4%) and ceiling (9.0%) effects.

Reliability Analyses

Cronbach's alpha values for the FSS scale, the MFIS subscales, and the MFIS total scores were all 0.93 or greater (see Table 3). This suggests some redundancy in item content. Item-to-scale correlations also were high for the FSS and for the MFIS scale and MFIS subscales.

Validity Analyses

The patterns of correlations between MFIS subscale scores and MFIS and FSS total scores were in the hypothesized direction and consistent with hypothesized magnitude. The FSS, which chiefly targets physical fatigue, had the highest correlation with the MFIS-physical ($\rho=0.77$) and the lowest correlation with MFIS-cognitive ($\rho=0.55$) (see Table 4).

Estimated associations between the MFIS and FSS scores and scores on other health constructs supported the discriminant validity of MFIS and FSS scores. Correlations between fatigue scores and HADS-anxiety scores were lower than correlations with scores on measures of pain and depression (see Table 4).

Known-groups validity also was supported. Individuals with less mobility impairment (EDSS 4.0) reported significantly less fatigue compared to those with more severe mobility impairment (EDSS 7.0). There were statistically significant differences among the EDSS groups for the FSS [F (2,1218)=118.9, $p<0.0001$], MFIS-cognitive [F (2,1224)=41.2, $p<0.0001$], MFIS-physical [F (2,1221)=234.2, $p<0.0001$], MFIS-psychosocial [F (2,1229)=130.9, $p<0.0001$], and MFIS-total [F (2,1215)=138.7, $p<0.0001$]. Post hoc comparisons found statistically significant differences between scores for participants with mild symptoms compared to participants with either moderate or severe symptoms (see Table 5). The MFIS-psychosocial scores were significantly different for respondents with moderate versus severe mobility impairment. All other fatigue scale and subscale

differences were non-significant between those with moderate versus severe mobility impairment (see Table 5).

Dimensionality

The CFA results supported the unidimensionality of the FSS but not the MFIS. The CFI for FSS was 0.97 [N=1241, Df=5, $\chi^2=13511$] well above the recommended threshold of .90. However, the CFI for MFIS was 0.84 [N=1240, Df=8, $\chi^2=15036$].

Because the unidimensional model did not fit the MFIS data well, a bifactor model was fitted in which all items loaded on a single general factor, the cognitive items loaded on a group factor, and the physical and psychosocial items loaded on a second group factor. The general factor accounted for much more of both the total variance (52%) and the common variance (70%) in scores. The cognitive and physical/psychosocial group factors accounted for 5% and 18% of the total variance, respectively; they accounted for 6% and 24% of the common variance, respectively.

IRT Analyses

After calibrating FSS and MFIS scores to separate graded response models, we calculated the amount of information provided by each scale and subscale. The resulting functions were plotted against FSS and MFIS total scores observed in the current sample. The scores are displayed in histograms below each graph in Figures 1a and 1b. Also included in the graph are reference lines for reliability estimates of 0.80 and 0.90. As the figure shows, the FSS provides substantial precision in measuring middle levels of fatigue, but was less precise in measuring both low and high levels of fatigue. We calculated the percentages of individuals who were measured with reliability less than each of the two reference reliability standards of 0.80 and 0.90. A total of 107 individuals (8.7%) were measured with <0.80 reliability. Most of these (n=96; 7.8%) were persons with high levels of fatigue (ceiling effect). A total of 189 (21.4%) were measured with <0.90 reliability; most of these (n=166; 13.6%) at the ceiling of the scale.

Figure 1b plots the information for the MFIS-Total score and all subscales. As the figure shows, compared to the FSS scores, the MFIS-Total score provided substantially more precision at the “tails” of the score distribution. As with the FSS, we evaluated percentages of individuals that were measured with reliability less than 0.80 and 0.90. A total of 13 individuals (1.1%) were measured with <0.80 reliability, all of which were at the high end of the scale. A total of 32 individuals (2.6%) were measured with <0.90 reliability, the majority of which were at the high end of the scale (n=22; 1.8%). MFIS scores were much less negatively skewed than FSS scores. A total of 164 (13.4%) subjects had FSS theta values greater than 1.0 (indicating very high levels of fatigue), but the FSS provides relatively small amounts of information at these levels of fatigue. In contrast, the MFIS measures with adequate precision at all levels of fatigue represented in the sample.

Discussion

The objective of this study was to use modern measurement methods to further examine the psychometric properties of two fatigue scales commonly used in MS research and to assist researchers with the selection of study instruments that best meet the needs of their study. Results suggest that researchers interested in measuring physical fatigue of samples whose fatigue ranges from mild to moderate can choose either instrument. For those interested in measuring both physical and mental fatigue and whose sample is expected to have high levels of fatigue we recommend using MFIS.

The mean FSS score in this study (5.1; SD=1.5) is similar to studies by Valko et al. (Valko, Bassetti, Bloch, Held, & Baumann, 2008) and Krupp et al. (Krupp, et al., 1989), which reported 4.7 (1.6) and 4.8 (1.3) respectively. Scores on the MFIS in this study also were similar to those obtained in other studies. The MFIS median value in the current study was 45.0. Other studies have reported median values ranging from 33.0 (Kos, et al., 2005; Tellez et al., 2005) to 45.0 (Kos, et al., 2005).

In the study sample, very few participants had scores at the floor or ceiling of either fatigue scale. Typically floor and ceiling effects are considered problematic when more than 15% of the sample has either the lowest or highest score possible (Terwee et al., 2007). Neither the FSS nor the MFIS had ceiling and floor effects of this magnitude in this study. The MFIS-psychosocial had the most subjects (9.0%) at the ceiling (worst possible score).

Reliability was evaluated using both CTT and IRT methods. The CTT analyses included estimation of internal consistency and item-to-scale correlations. Cronbach's alpha values for the FSS and MFIS were all greater than 0.85 suggesting redundancy and opportunity to shorten the scales. The item-to-scale correlations were all greater than the criterion of 0.40, providing evidence of item homogeneity for the MFIS subscales and MFIS total score. In addition to Cronbach's alpha, this study also used test information obtained from an IRT analysis to examine the precision of the MFIS and FSS along the whole fatigue continuum. MFIS appears to measure with greater precision than FSS at higher levels of fatigue.

Construct validity was assessed by comparing the associations among fatigue subscale scores and total scores (convergent validity) as well as with other health constructs (discriminant validity), including pain interference, anxiety, and depression. Correlations between FSS scores and both MFIS-physical and MFIS-cognitive scores were found to be similar to the values obtained by Tellez et al. (2005) in a previous study, such that there was a greater association between FSS and the MFIS-physical scores than between FSS and the MFIS-cognitive scores. Furthermore, results from this study are consistent with the finding of Tellez et al. (2005) that MFIS scores are more highly correlated with depression scores than are FSS scores. A similar pattern was observed in relation to scores for other health concepts, i.e., lower correlations were observed between FSS scores compared to the MFIS total and domain scores.

Known-groups validity was supported in this study by observing higher fatigue scores (higher levels of fatigue) in subjects with moderate to severe MS symptoms in all the FSS and MFIS scores. This finding is consistent with the MS fatigue literature. (Hadjimichael, 2008) The MFIS-psychosocial was the only domain where a significant difference was observed between participants reporting moderate (EDSS 4.5 – 6.5) and severe (EDSS 7.0) symptoms which is surprising, because there are only two items in this domain. Future studies should evaluate whether this result is replicated in other samples.

Testing the assumption of unidimensionality required for interpreting the summary score and fitting an IRT model also provides evidence related to construct validity. The degree of unidimensionality was found sufficient for fitting an IRT model for the FSS and MFIS. Previously reported analyses with MS samples both reported support for unidimensionality (Hagell et al., 2006; Mills, et al., 2009). Strict unidimensionality is desirable for applications of IRT, however it is well recognized that data from psychological measures are rarely (if ever) strictly unidimensional. In fact, to represent complex constructs adequately some multidimensionality may be necessary (Reise et al., 2010). The issue is what degree of multidimensionality and resulting parameter estimates distortions can be tolerated. Published studies suggest that IRT scores are fairly robust to dimensionality violations (Camilli et al., 1995; Dorans & Kingston, 1985)

In the current study, IRT was used to examine psychometric properties of scales. If the purpose was to develop IRT-based scoring for MFIS for instance for a computerized adaptive testing application, parameter bias caused by unmodeled multidimensionality might be of more concern. However, in this study context and because the bifactor analysis results supported sufficient unidimensionality for fitting an IRT model, the more multidimensional structure of MFIS compared to FSS is viewed more as a strength of the scale than a concern.

The results of this study highlight important differences between the FSS and MFIS. First, the FSS is shorter but measures primarily physical fatigue and does not measure with adequate precision at higher levels of fatigue which are often reported by individuals with MS. The FSS also utilizes a one week recall period, which one study suggests may be less accurate in measuring mean levels of fatigue than the four week recall period used by the MFIS (Broderick et al., 2008). MFIS is longer and measures both physical and cognitive fatigue with adequate precision along the whole continuum of fatigue commonly reported by people with MS. Therefore, in studies that include people with high levels of fatigue and where cognitive fatigue is of interest, it is preferable to administer MFIS even though it is longer than FSS.

Limitations and Future Directions

In considering the results from this study, it is important to consider its strengths and limitations. This study included a large sample ($n=1271$) of individuals with MS living in the community. Limitations of this study include the cross-sectional nature of the data that does not allow for evaluation of responsiveness, an important aspect of psychometric functioning. In addition, the response rate to the initial invitation was low, and no effort was made to recruit a sample representative of individuals living with MS in the United States; therefore, it would be helpful if the study was replicated with different MS samples.

Our analyses (both CTT and IRT) suggested some item redundancy in both scales. In addition, IRT methods can be used to assess bias in responses to the items (referred to as differential item functioning). Longitudinal studies that administer these measures at baseline and after an effective treatment could be used to evaluate the degree to which scores on the measures detect change over time. Additional work may be needed to establish how much change in summary scores constitutes a clinically meaningful change.

The application of modern measurement methods can be used to improve the psychometric properties of both fatigue scales. In addition, fatigue instruments developed using modern psychometric theory are now publicly available (Cella et al., 2010) that allow for computerized adaptive testing and development of short instruments targeted to certain populations or certain levels of fatigue. These instruments lower respondent burden while estimating scores with a high level of precision along the entire fatigue continuum.

Acknowledgments

The contents of this manuscript were developed under grants from the Department of Education, NIDRR grant numbers H133B031129 & H133B080025, and the National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institute of Health grant number 5U01AR052171. However, these contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

References

- Bamer AM, Cetin K, Amtmann D, Bowen JD, Johnson KL. Comparing a self report questionnaire with physician assessment for determining multiple sclerosis clinical disease course: a validation study. *Multiple Sclerosis*. 2007; 13(8):1033–1037. [PubMed: 17468438]

- Becker J, Schwartz C, Saris-Baglama RN, Kosinski M, Bjorner JB. Using Item Response Theory (IRT) for developing and evaluating the Pain Impact Questionnaire (PIQ-6TM). *Pain Medicine*. 2007; 8(S3):361–370.
- Bowen J, Gibbons L, Gianas A, Kraft GH. Self-administered Expanded Disability Status Scale with functional system scores correlates well with a physician-administered test. *Multiple Sclerosis*. 2001; 7(3):201–206. [PubMed: 11475445]
- Boyle GJ. Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*. 1991; 12:291–294.
- Branas P, Jordan R, Fry-Smith A, Burls A, Hyde C. Treatments for fatigue in multiple sclerosis: a rapid and systematic review. *Health Technology Assessment*. 2000; 4(27):1–61. [PubMed: 11074395]
- Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. *Pain*. 2008; 139(1):146–157. [PubMed: 18455312]
- Camilli G, Wang M-m, Fesq J. The effect of dimensionality on equating the Law School Admission Test. *Journal of Educational Measurement*. 1995; 32:79–96.
- Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Hays R, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010; 63(11):1179–1194. [PubMed: 20685078]
- Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. *Ann Acad Med Singapore*. 1994; 23(2):129–138. [PubMed: 8080219]
- Dorans NJ, Kingston NM. The Effects of Violations of Unidimensionality on the Estimation of Item and Ability Parameters and on Item Response Theory Equating of the GRE Verbal Scale. *Journal of Educational Measurement*. 1985; 22:249–62.
- Everitt, BS. *The Cambridge Dictionary of Statistics*. 2. Cambridge, UK: Cambridge University Press; 2002.
- Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. *Clin Infect Dis*. 1994; 18(Suppl 1):S79–83. [PubMed: 8148458]
- Hagell P, Hoglund A, Reimer J, Eriksson B, Knutsson I, Widner H. Measuring fatigue in Parkinson's disease: a psychometric study of two brief generic fatigue questionnaires. *J Pain Symptom Manage*. 2006; 32:420–432. [PubMed: 17085268]
- Hadjimichael O, Vollmer T, Oleen-Burkey M. Fatigue characteristics in multiple sclerosis: the North American Research Committee on Multiple Sclerosis (NARCOMS) survey. *Health Qual Life Outcomes*. 2008; 6:100. [PubMed: 19014588]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6:1–55.
- Kos D, Kerckhofs E, Carrea I, Verza R, Ramos M, Jansa J. Evaluation of the Modified Fatigue Impact Scale in four different European countries. *Multiple Sclerosis*. 2005; 11:76–80. [PubMed: 15732270]
- Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001; 16(9):606–613. [PubMed: 11556941]
- Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology*. 1989; 46:1121–1123. [PubMed: 2803071]
- McDonald RP. The dimensionality of test and items. *British Journal of Mathematical and Statistical Psychology*. 1981; 34:100–117.
- Mills R, Young C, Nicholas R, Pallant J, Tennant A. Rasch analysis of the Fatigue Severity Scale in multiple sclerosis. *Multiple Sclerosis*. 2009; 15:81–87. [PubMed: 18755824]
- Multiple Sclerosis Council for Clinical Practice Guidelines. *Fatigue and Multiple Sclerosis: Evidence-Based Management Strategies for Fatigue in Multiple Sclerosis*. Washington, DC: Paralyzed Veterans of America; 1998.

- Muthen, LK.; Muthen, BO. Mplus (Version 5.21) [Computer software]. Los Angeles, CA: Muthen & Muthen; 2009. <http://www.statmodel.com/>
- Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*. 2010; 92:544–559. [PubMed: 20954056]
- Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res*. 2007; 16(Suppl 1):19–31. [PubMed: 17479357]
- Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*. 1969; 17
- Tellez N, Rio J, Tintore M, Nos C, Galan I, Montalban X. Does the Modified Fatigue Impact Scale offer a more comprehensive assessment of fatigue in MS? *Multiple Sclerosis*. 2005; 11:198–202. [PubMed: 15794395]
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007; 60(1):34–42. [PubMed: 17161752]
- Valko PO, Bassetti CL, Bloch KE, Held U, Baumann CR. Validation of the fatigue severity scale in a Swiss cohort. *Sleep*. 2008; 31:1601–1607. [PubMed: 19014080]
- Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983; 67(6):361–370. [PubMed: 6880820]

Impact

- Although fatigue is one of the most common symptoms in MS, research on the psychometric properties of common fatigue scales in this population is lacking. This paper is the first to directly compare and evaluate the psychometric properties of the Fatigue Severity Scale and the Modified Fatigue Impact Scale, two fatigue scales most often used to measure fatigue in individuals living with MS.
- This study provides evidence that both scales effectively measure fatigue in MS, however they measure different aspects of fatigue.
- This study provides guidance for MS researchers and clinicians in selecting the fatigue scale most likely to meet the study's needs with respect to expected levels of fatigue in the sample and precision required.

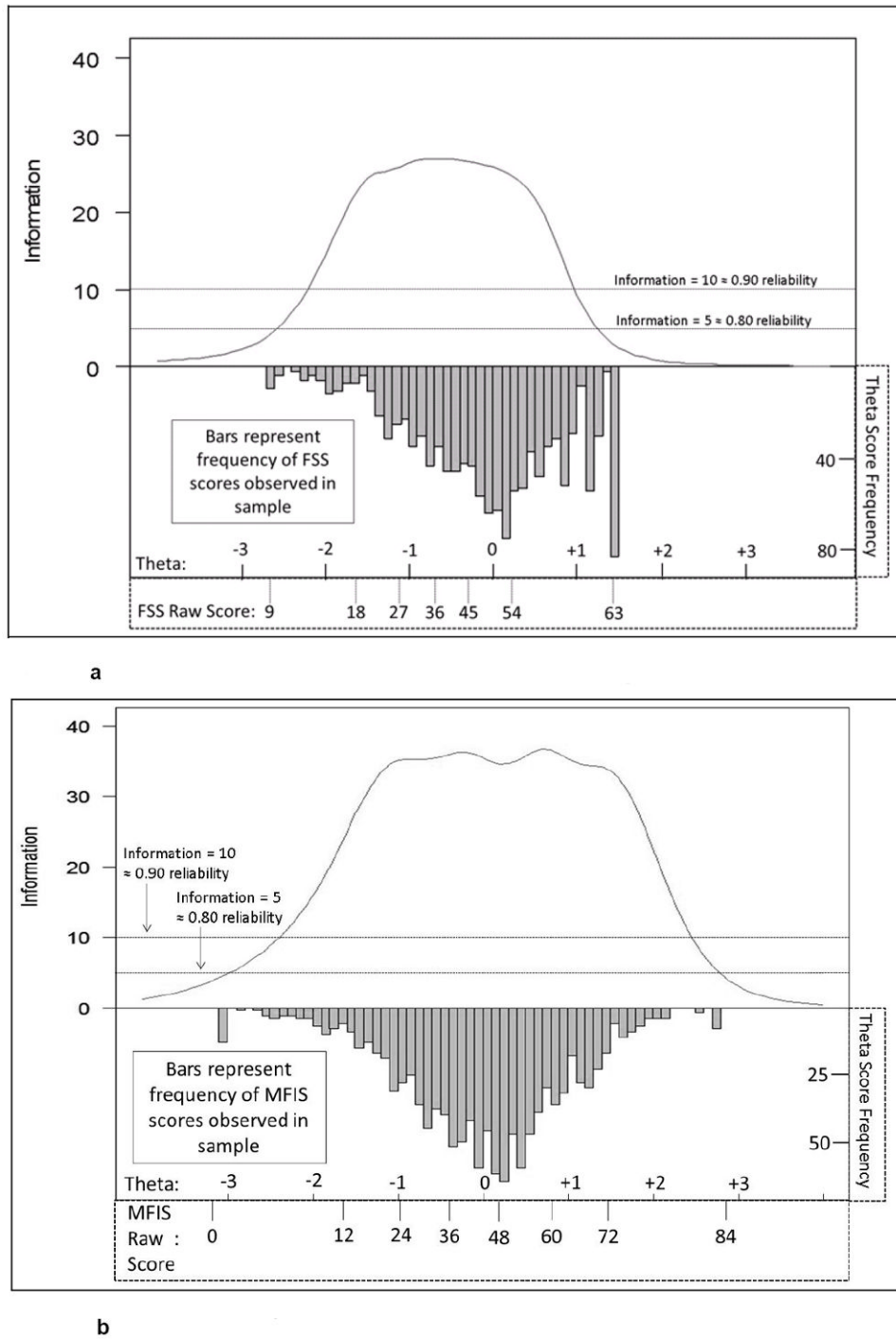


Figure 1.
a Item response theory calculated information in the Fatigue Severity Scale compared to a levels of fatigue in a community dwelling sample of individuals with MS
b Item response theory calculated information for the Modified Fatigue Impact Scale compared to a levels of fatigue in a community dwelling sample of individuals with MS

Table 1

Demographic and disease characteristics of a community sample of individuals with multiple sclerosis (N=1,271).

Variable	n (%) mean ± SD
Age (n=1237)	50.7 ± 11.6
Duration of Disease (n=1261)	13.2 ± 10.1
Sex (n=1237)	
Women	992 (80.2)
Men	245 (19.8)
Race (n=1235) ^a	
Caucasian	1,206 (29)
Native American or Alaska Native	35 (2.8)
Asian	12 (1.0)
African-American	21 (1.7)
Education Completed (n=1237)	
< High School	19 (1.5)
High School / GED	155 (12.5)
Vocational/Some College	465 (37.6)
Bachelors Degree	374 (30.2)
Professional/Graduate	224 (18.1)
Employment Status ^a	
Employed 20+ hrs/wk	447 (36.2)
Employed <20 hrs/wk	63 (5.1)
Unemployed	420 (33.0)
Retired	383 (30.1)
Homemaker	154 (12.5)
Married (n=1234)	
Married/Live with Significant Other	867 (70.3)
Separated/Divorced	215 (17.4)
Never Married	110 (8.9)
Widowed	42 (3.4)
Course of Disease (n=1197)	
Relapsing Remitting	700 (58.5)
Secondary Progressive	240 (20.1)
Primary Progressive	157 (13.1)
Progressive Relapsing	100 (8.4)
Level of Disability (EDSS) (n=1236)	
0 - 4.0	401 (32.4)
4.5 - 6.5	592 (47.9)
7.0 - 10.0	243 (19.7)

^aNumbers may sum to more than 100%, as individuals were allowed to choose multiple answers.

Table 2

Scores, floor effects, and ceiling effects for measures in study population.

Measure	Mean (SD)	Median	Theoretical Range	Floor Responses N (%)	Ceiling Responses N (%)
FSS	5.1 (1.5)	5.55	1-7	11 (0.9)	83 (6.8)
MFIS-Phy	22.0 (8.7)	23.0	0-36	20 (1.6)	20 (1.6)
MFIS-Cog	18.0 (9.5)	18.0	0-40	33 (2.7)	11 (0.9)
MFIS-Pso	4.2 (2.3)	4.0	0-8	92 (7.4)	111 (9.0)
MFIS-Total	44.2 (18.2)	45.0	0-84	13 (1.1)	8 (0.7)
BPI-10 ^a	3.7 (2.5)	3.5	0-10		
PIQ-6 ^a	59.1 (7.6)	59.0	40-78		
HADS-A	5.9 (4.2)	5.0	0-21		
PHQ-9	8.3 (6.1)	7.0	0-27		

^aPIQ and BPI completed in people with pain only (PIQ: n=832; BPI: n=847).

Note: SD: Standard deviation; FSS: Fatigue Severity Scale; MFIS-Phy: Modified Fatigue Inventory Scale-Physical; MFIS-Cog: Modified Fatigue Inventory Scale-Cognitive; MFIS-Pso: Modified Fatigue Inventory Scale-Psychosocial; MFIS-Total: Modified Fatigue Inventory Scale-Total; BPI-10: Brief Pain Inventory-10; PIQ-6: Pain Impact Questionnaire; HADS: 7-item Anxiety Scale from the Hospital Anxiety and Depression Scale; PHQ-9: Patient Health Questionnaire-9.

Table 3

Reliability of the fatigue scale scores

Fatigue Measure	Internal Consistency	Item-to-Scale Correlations
FSS	0.93	0.56 – 0.89
MFIS-Physical	0.94	0.70 – 0.86
MFIS-Cognitive	0.96	0.77 – 0.90
MFIS-Psychosocial	NA	NA
MFIS-Total	0.96	0.66 – 0.80

Note: NA: not applicable.

Table 4

Spearman rank correlations within fatigue scale scores and among fatigue scores and other health measures ^a

Measure	FSS	MFIS-Physical	MFIS-Cognitive	MFIS-Psychosocial	MFIS-Total
FSS	-	0.77	0.55	0.69	0.74
MFIS-Physical		-	0.59	0.77	0.88
MFIS-Cognitive			-	0.60	0.89
MFIS-Psychosocial				-	0.81
BPE-10	0.48	0.57	0.49	0.55	0.60
PIQ-6	0.49	0.56	0.47	0.53	0.57
HADS-Anxiety	0.23	0.26	0.43	0.27	0.38
PHQ-9	0.55	0.60	0.64	0.60	0.70

^aAll correlations are significant p<0.01

Table 5

Summary of ANOVA and t-test results comparing EDSS levels by fatigue scale scores

EDSS Group Comparisons	Omnibus Test	Mild to Moderate	Mild to Severe	Moderate to Severe
FSS				
degrees of freedom	2	981	628	819
number of subjects	1218	983	630	821
F statistic	118.9	NA	NA	NA
t statistic	NA	-14.9	-10.0	0.8
p value	<0.0001	<0.0001	<0.0001	0.41
MFIS Cognitive				
degrees of freedom	2	982	633	825
number of subjects	1224	984	635	827
F statistic	41.2	NA	NA	NA
t statistic	NA	-9.2	-5.4	1.8
p value	<0.0001	<0.0001	<0.0001	0.08
MFIS Physical				
degrees of freedom	2	983	631	820
number of subjects	1221	985	633	822
F statistic	234.2	NA	NA	NA
t statistic	NA	-19.9	-15.6	-1.5
p value	<0.0001	<0.0001	<0.0001	0.13
MFIS Psychosocial				
degrees of freedom	2	987	635	828
number of subjects	1229	989	637	830
F statistic	130.9	NA	NA	NA
t statistic	NA	-14.0	-13.5	-3.0
p value	<0.0001	<0.0001	<0.0001	0.003
MFIS Total				
degrees of freedom	2	979	627	816
number of subjects	1215	981	629	818
F statistic	138.7	NA	NA	NA
t statistic	NA	-15.8	-11.6	0.02

EDSS Group Comparisons	Omnibus Test	Mild to Moderate	Mild to Severe	Moderate to Severe
p value	<0.0001	<0.0001	<0.0001	0.98

Note: NA: Not Applicable; EDSS: Expanded Disability Status Scale; FSS: Fatigue Severity Scale; MFIS: Modified Fatigue Inventory Scale;