Taylor & Francis
Taylor & Francis Group

Fatigue: Biomedicine, Health & Behavior

ISSN: 2164-1846 (Print) 2164-1862 (Online) Journal homepage: www.tandfonline.com/journals/rftg20

Psychometric evaluation of the Multidimensional Fatigue Inventory (MFI) in Portuguese adult samples: a consensus version

Cátia Reis, Madalena L. Santos & João Marôco

To cite this article: Cátia Reis, Madalena L. Santos & João Marôco (2025) Psychometric evaluation of the Multidimensional Fatigue Inventory (MFI) in Portuguese adult samples: a consensus version, Fatigue: Biomedicine, Health & Behavior, 13:3, 224-240, DOI: 10.1080/21641846.2025.2498285

To link to this article: https://doi.org/10.1080/21641846.2025.2498285

FATIGUE: BIOMEDICINE, HEALTH & BEHAVIOR
2025, VOL. 13, NO. 3, 224–240
https://doi.org/10.1080/21641846.2025.2498285

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Psychometric evaluation of the Multidimensional Fatigue Inventory (MFI) in Portuguese adult samples: a consensus version

Cátia Reis [a,b,c], Madalena L. Santos[b] and João Marôco[d]

[a]CRC-W, Católica Research Centre for Psychological, Family and Social Wellbeing, Faculdade de Ciências Humanas, Universidade Católica Portuguesa, Lisboa, Portugal; [b]Gulbenkian Institute for Molecular Medicine, Lisboa, Portugal; [c]Faculdade de Medicina, Instituto de Saúde Ambiental – ISAMB, Universidade de Lisboa, Lisboa, Portugal; [d]Intrepid – Universidade Lusófona, Lisboa, Portugal

ABSTRACT

Background: Fatigue is a multidimensional symptom common across various conditions, often underdiagnosed despite its significant impact on quality of life. The Multidimensional Fatigue Inventory (MFI) is a self-assessment scale designed to measure multiple fatigue dimensions. However, its factor structure has proven difficult to replicate.

Objective: This study evaluates the psychometric properties of the MFI in a Portuguese sample, proposing a refined version and contributing to the debate on its structural validity.

Methods: A total of 1,858 participants (aged 18–85, $M = 30.72 \pm 13.9$; 69.5% female) were recruited through non-probabilistic sampling and split into two subsamples. Six factor structures were tested using confirmatory factor analysis (CFA); the three best-fitting models underwent measurement invariance testing across subsamples and sex. The best model was further validated in the full sample, with a second-order factor (Global Fatigue – GlobF) tested.

Results: Three models were discarded due to poor fit. The retained models showed strong measurement invariance and good psychometric properties. The final model preserved the original Physical Fatigue and Mental Fatigue dimensions. CFA fit indices were strong (RMSEA = 0.076, SRMR = 0.028, CFI = 0.971, TLI = 0.957). Reliability was high ($\alpha = 0.798$–0.881; $\omega = 0.807$–0.880), with adequate construct validity (AVE = 0.519–0.648) and discriminant validity (HTMT = 0.589).

Conclusions: This study proposes a refined, reliable and valid consensus version of the MFI for the Portuguese population. The revised instrument improves fatigue assessment and may support better fatigue identification and management in clinical and non-clinical settings. Further research should examine its responsiveness and cross-cultural applicability.

ARTICLE HISTORY
Received 15 December 2024
Accepted 20 January 2025

KEYWORDS
MFI; fatigue; factor structure; validity; psychometric properties

CONTACT Cátia Reis ccreis@ucp.pt

Supplemental data for this article can be accessed online at https://doi.org/10.1080/21641846.2025.2498285.

Introduction

There is no universally accepted definition of fatigue. While some authors avoid defining it altogether [1], other propose definitions using everyday language synonyms like tiredness or exhaustion [2]. Despite these variations, there is consensus that fatigue is fundamentally a subjective experience. It is important to distinguish tiredness, a general lack of energy or motivation, from sleepiness, which is a physiological drive for sleep often linked to circadian rhythms and homeostatic sleep pressure [3]. This distinction clarifies that fatigue, while overlapping with tiredness and sleepiness, encompasses broader dimensions, including physical, mental, or emotional effort.

Raizen et al. [3] defined fatigue as self-reported sensation of difficulty in initiating or sustaining physical, mental, or emotional activities, characterized by reduced performance capacity despite maintained motivation and a perception of disproportionately high effort. Fatigue is a complex and multidimensional construct. This multidimensional nature of fatigue has been explored in prior research. In this context, fatigue is commonly differentiated into peripheral and central components [4]. Peripheral fatigue is defined as the inability to sustain a specified force output or work rate during physical exertion and, in contrast, central fatigue pertains to the failure to initiate or maintain attentional tasks and self-motivated physical activities [5]. Central fatigue is further comprised of multiple dimensions, including physical fatigue, mental fatigue, and motivational changes. Physical fatigue is marked by challenges in executing physical tasks, whereas mental fatigue manifests as difficulties with concentration and the performance of cognitive activities [1]. As so, fatigue, involve a variety of dimensions with mental, emotional, and/or physical manifestations. These can be experienced by healthy individuals and those with acute or chronic conditions [6], ranging from acquired immunodeficiency syndrome [7] to cancer [8] and rheumatoid arthritis [9]. Fatigue is among the top three most reported symptoms in prevalence studies and may signal disease progression. Moreover, in the general population, it is also linked to higher mortality risk (Odds ratio = 2.14) [10].

In contemporary society, fatigue represents a major challenge primarily due to the high demands of the workplace, extended duty hours, disturbances in circadian rhythms, social and societal pressures, and inadequate sleep [11, 12]. Fatigue is linked to slower reaction times, reduced productivity, impaired ability to handle complex planning, and a higher risk of decision-making errors, particularly in cases of acute fatigue [13]. When fatigue is prolonged in time, can progress to chronic fatigue. In fact, productivity losses due to worker fatigue are estimated to cost employers between $1200 and $3100 per employee annually [14]. Despite its significance, fatigue is often underdiagnosed and neglected partly because, distinguishing it from typical sensations of tiredness or sleepiness has proven challenging, and is many often disregarded [15].

Effective management of fatigue-related disorders requires assessing its nature, frequency, intensity, and duration, yet no universally accepted objective measures for evaluating fatigue currently exist [3, 6]. As a result, subjective methods such as questionnaires play a crucial role in fatigue assessment. There are over 30 self-assessment instruments measuring fatigue (see [3] for lists of non-cancer and cancer self-report fatigue instruments). Some scales, such as the Cancer Fatigue Scale [16], were created for specific medical conditions. Others, like the Multidimensional Fatigue Scale (MFI) [1] are suitable for clinical and non-clinical subjects alike. The MFI was developed to meet the need for a

questionnaire that excludes somatic items (such as headaches) and covers five dimensions of fatigue – General Fatigue, Physical Fatigue, Mental Fatigue, Reduced Motivation and Reduced Activity [1] –, making it possible to distinguish for example general fatigue from mental or physical fatigue. The MFI has been widely used as a research tool in studies of various diseases and the general population [17, 18] making it one of the most utilized scales for assessing multiple dimensions of fatigue [2, 6]. This capability allows for a nuanced understanding of fatigue, facilitating subjects' characterization with respect to which dimension of fatigue is most prominent. For example, in high-stress occupations like air traffic controllers, mental fatigue may be more prominent, while in athletes, physical fatigue may play a more dominant role. By identifying which type of fatigue is most affecting an individual, the MFI can support condition-specific interventions [19].

In Portuguese speaking countries, one study translated and adapted the MFI-20 into Brazilian Portuguese and evaluated its psychometric properties on Brazilian Hodgkin's lymphoma survivors [20]. Furthermore, another study [21] aimed at the cross-cultural adaptation and validation of the MFI-20 specifically for Brazilian patients with Parkinson's disease, providing additional evidence for its applicability in Portuguese-speaking populations.

The MFI has been translated into several languages, including European [22] and Brazilian Portuguese [20, 21] but, to our knowledge, no studies have examined its psychometric properties in Portuguese samples.

Despite the availability of validated versions of the MFI-20 for Brazilian Portuguese, to our knowledge, no studies have yet addressed its validation for the European Portuguese. Given the linguistic and cultural differences between Brazilian and European Portuguese, a formal validation is essential to ensure the tool's accuracy, relevance, and reliability for assessing fatigue in European Portuguese-speaking populations.

Evidence supporting the MFI's psychometric properties has been inconsistent, particularly regarding the replication of its factor structure across different populations, further highlighting the need for a robust and culturally appropriate validation.

Smets et. al. [23] tested three different models in four different samples by confirmatory factorial analysis. The first model was the standard model, the second model combined General Fatigue (GF) and Physical Fatigue (PF) into a single factor, and the third model excluded GF altogether. All models had Adjusted Goodness of Fit Index (AGFI) equal to or higher than 0.94 in all the samples, suggesting a good fit; the lowest Cronbach's Alpha (α), across all samples and all models, was 0.77, a sign of acceptable reliability. The authors acknowledged that the most parsimonious model would be preferable but decided to retain the five-factor solution until more knowledge of the relationships of the scales to other constructs was available. This model was replicated by Smets et al. [23] in Dutch and Scottish samples of cancer patients through a CFA. In both samples AGFI was > 0.9, with $\alpha > 0.75$.

Since then, the original five factor structure has been difficult to replicate. This may be explained, in many projects, by the use of Exploratory Factor Analysis (EFA) as the sole method or as a starting point for further analysis [24–29]: this method will return the factorial structure best suited to the data according to the extraction method used and without inference about the structural validity of the original model. These studies found variations of 3 or 4 factors models, sometimes somewhat similar to the original ones but with loss of items, items cross-loading in different factors leading

to the emergence of factors such as Motivation [30] or Spiritual Fatigue [31], and a strong overlap between GF and PF items, sometimes combined into a single General and Physical Fatigue factor (GPF; e.g. [24, 26]). Other studies used only confirmatory factorial analysis (CFA) to test the validity of the original 5 factors model [32–35], sometimes against alternative models based on construct hypotheses (e.g. assumption of unidimensionality) [32]. All but one failed to replicate the original model, the exception being Chandel et al. [33].

Baussard et al. [29] proposed that one possible source of psychometric trouble was the use of reverse wording in the MFI items. Reverse wording in items express the opposite pole of the construct being measured, i.e. in a fatigue scale reverse worded items are those that express low fatigue, such as *I feel fit,* item #1 in the MFI. Half of the MFI items use reverse wording, a still common strategy to prevent response styles and bias in questionnaires, despite mounting evidence that it is ineffective and has a negative impact on the psychometric properties of the instrument [36–38] uses the MFI for an empirical example. Based on this, Baussard et al. [29] proposed a short version of the MFI using only the straight worded items. They conducted an EFA of the 10 items and found a 3-factor model: Physical Fatigue (PF; four items, almost identical to the original from Smets et al. [1]), Emotional Fatigue (EmF; four items, combining items from RA and RM) and Cognitive Fatigue (CogF; two items from MF). They conducted a CFA on this model that showed good fit (RMSEA = 0.059, CFI = 0.978); they also conducted a CFA on three other models [1, 24, 30] that yielded very poor fit to the data (RMSEA > 0.11, CFI < 0.90).

Despite validation in multiple countries, its construct validity remains unclear. This study aimed to: (a) assess the MFI's psychometric properties in Portuguese-speaking samples from Portugal and Brazil; (b) propose a refined, unified version to support transcultural research; and (c) contribute to the ongoing debate on the MFI's validity by providing additional psychometric evidence, addressing challenges in replicating its factor structure. To achieve this, we developed a consensus version of the MFI – a more concise instrument with fewer items that maintains strong psychometric integrity. This refined version incorporates the most robust items from previous adaptations and demonstrates greater reliability and consistency than its predecessors. We believe this version offers a valuable tool for both research and clinical and non-clinical applications in Portuguese-speaking populations.

Methods and procedures

Participants and procedures

We gathered a global sample of Portuguese speaking adults by combining two subsamples, collected between 6 October 2021, and 20 March 2023. One subsample was composed of 733 participants (ages 18–85, $M = 40.34$, SD = 14.1), of which 69.9% were female. Being under 18 years of age or not having enough grasp of the Portuguese language were used as exclusion criteria. The second subsample included 1125 higher education students (HES) from Portugal (ages 18–66, $M = 22.24$, SD = 5.98; 63.3% female) enrolled in any year of study. Exclusions included minors and insufficient Portuguese proficiency. Participants were recruited via non-probabilistic snowball sampling on

social media, with no incentives. Online surveys are increasingly used in psychometric research and have demonstrated comparable reliability and validity to traditional methods [39]. Research questionnaires on Qualtrics required mandatory informed consent. The use of non-probability sampling in this study was driven by practical and ethical considerations common in psychological research, such as the need for informed consent, participant voluntarism, and adherence to IRB guidelines. While non-probability sampling does introduce limitations regarding the generalizability of the findings, we took steps to enhance the representativeness of our sample by recruiting a large and diverse group ($N > 1000$) to reflect key demographic characteristics of the Portuguese population, including age, gender, and geographic distribution. The global sample included 1,858 participants (ages 18–85, $M = 30.72$, SD = 13.9; 69.5% female, 14.2% male, 16.4% non-respondents), randomly split into a test sample ($n = 928$; 68.9% female, 14.1% male, 17% missing; ages 18–78, $M = 30.59$, SD = 13.77) and a validation sample ($n = 930$; 70.1% female, 14.2% male, 15.7% missing; ages 18–85, $M = 30.85$, SD = 14.12). The test sample was used to refine MFI models, and the validation sample for measurement invariance testing, including gender invariance. This study was approved by ISPA's Ethics Committee (Comissão de Ética do ISPA) under reference D-038-06-2021.

Instrument and scale adaptation

The Multidimensional Fatigue Inventory (MFI) consists of 20 items across five factors: General Fatigue (GF), Physical Fatigue (PF), Reduced Activity (RA), Reduced Motivation (RM), and Mental Fatigue (MF) [1]. Each factor includes four items, two reflecting high fatigue levels, and two reflecting low levels (reverse-worded) [29, 35]. Responses are scored on a 5-point Likert-like scale, from 1 ('yes, it's true') to 5 ('no, it's not true'), with higher scores indicating greater fatigue. Reverse-scoring of low-fatigue items ensures alignment with the scale's direction.

To our knowledge, no psychometric evaluations of the MFI have been published for Portuguese samples, and studies on Brazilian samples remain limited. Additionally, differing versions of the MFI for Brazilian Portuguese and European Portuguese hinder collaborative research between these Portuguese-speaking populations.

Permission to use the MFI was obtained from its original author, and permission for the Brazilian version was granted from Dr. Josiane Lopes. The English version, the European Portuguese [22] and Brazilian Portuguese [21] versions were compared by the authors and two Brazilian sleep and fatigue researchers (see Acknowledgments) to create a version suitable for both countries. This version was tested for response validity [40] in a convenience sample of approximately 20 general population adults from each country. No comprehension issues were identified, and participants consistently based their responses on their subjective fatigue levels.

Data analysis

Missing values were not replaced, since they could not be assumed to be missing at random [41], and removed them listwise. Adequate item sensitivity was determined by ensuring that items had responses across all scale options, did not have a median at the scale's minimum or maximum value, and exhibited absolute Skewness (Sk) ≤ 3 and

absolute Kurtosis (Ku) ≤ 7. Values exceeding these thresholds indicated significant deviations from normality [42, 43]. All descriptive statistics were calculated with IBM SPSS Statistics v29 and Jamovi v2.6. Evidence based on internal structure [40] was obtained through Confirmatory Factor Analysis (CFA) using the robust maximum likelihood method (MLR) with Satorra -Bentler correction. The analysis was conducted using the SEMLj module [44], a Jamovi interface for the lavaan R package [45]. The criteria for goodness of fit were as follows: a robust Root Mean Square of Approximation (RMSEA) ≤ 0.08, a robust Standardized Root Mean Square Residual (SRMS) ≤ 0.08, a robust Comparative Fit Index (CFI) ≥ 0.9 or a robust Tucker Lewis Index (TLI) ≥ 0.9 ([43, 46]). Cronbach's α or Mcdonalds $\acute{\omega}$ [47] ≥ 0.7, were used as criteria for good reliability. An Average Variance Extracted (AVE ([43, 46])) of $\geq$0.5 indicated acceptable convergence validity for a factor. Discriminant validity between factors was established by an heterotrait / monotrait ratio of correlations (HTMT) < 0.85 [43, 46]. Measurement invariance was estimated by imposing successive constraints on the base, free, model (configural invariance): factor loadings (metric invariance), factor loadings plus intercepts (scalar invariance), and factor loadings plus intercepts plus residuals [43, 48] and by calculating the differences between the χ^2 / degrees of freedom (DF) of two consecutive nested models, which should be non-significant, or by having a decrease in consecutive CFIs <0.01 or by having an increase in RMSEA of consecutive models >0.02 [43, 49]. Comparison between alternative models was made by the Akaike Information Criterion (AIC), Bayes Information Criterion (BIC) and by the Expected Cross-Validation Index (ECVI) [43, 46]. Statistical significance was assumed for $p < .05$ for all analysis.

Results

The descriptive statistics for MFI items in the global sample (see supplementary Table 1) show that most items have a median and mode of 3, with responses ranging from 1 to 5. Skewness indicates generally symmetric responses, though items 5 and 20 are left-skewed, reflecting higher scores. Negative kurtosis across items suggests flat distributions with lighter tails, indicating responses are spread evenly rather than clustering around the mean or extremes. All items utilized all Likert options, with no extreme medians, skewness >|3|, or kurtosis >|7|, consistent across the test and validation sub-samples. We decided to follow up on Baussard et al. [29] analysis by comparing the same three models they tested plus two models of our own. Model A was Smets et. al [1] original five factor model (GF, PF, RA, RM, MF); Model B was Gentille et al. [30] four factor model (GF, RA, MF, RM); Model C was Fillion et al. [24] four factor model (GPF, RA, RM, MF); Model D was Baussard et al. [29] three factor model using only the non-reverse worded items (PF, EmF, CogF). To ensure systematic approach, we added two models: Model E was Smets et al. [1] original five factors model, including only the non-reverse-worded items, while Model F represented the same five-factor model but including only the reverse-worded items. The model sample was used to conduct the initial CFAs and refine the models. Table 1 presents the fit and reliability indicators of the initial CFAs.

Model A had encountered identification issues due to the strong correlation between GF and PF (HTMT = 0.948) and between RA and RM (HTMT = 0.908). To address this, we removed GF and refined the model by eliminating items with low βs (items 9 and 15 from RA), and factors with poor reliability or convergent validity. This process continued

Table 1. Fit and reliability indicators for the six models being compared in the test sample.

Initial Models	Factors	N	Free Parameters	χ^2	DF	RMSEA	SRMR	CFI	TLI	Lowest Alpha (factor)	Lowest AVE (factor)	Highest HTMT (factor/ factor)	BIC	AIC	ECVI
Model A Smets et al. (1995) [1]	GF = it1 + it5 + it12 + it16 PF = it2 + it8 + it14 + it20 RA = it3 + it6 + it10 + it17 RM = it4 + it9 + it15 + it18 MF = it7 + it11 + it13*+it19	857	70	1334.00	160	0.100	0.092	0.842	0.813	0.543 (RM)	0.323 (RM)	0.948 (GF/PF)	50385.07	50052.33	58.47
Model B Gentile et al. (2003) [30]	GF = it1 + it12 + it2 + it8 + it20 + it3 + it5 + it16 + it14 RA = it6 + it10 + it17 MF = it7 + it11 + it13 + it19 + it9 + it18 RM = it4 + it15	857	66	1272.00	164	0.096	0.077	0.851	0.828	0.282 (RM)	0.176 (RM)	0.898 (RA/RM)	50282.30	49968.57	58.37
Model C Fillion et. al (2003) [24]	GPF = it1 + it20 + it5 + it16 + it14 + it12 + it8 RA = it9 + it17 + it18 RM = it15 + it4 + it6 MF = it11 + it7	857	51	894.00	84	0.114	0.077	0.834	0.792	0.477 (RM)	0.220 (RM)	0.954 (RA/RM)	38478.26	38235.03	44.67
Model D Baussard et al. (2018) [29]	PF = it2 + it5 + it14 + it16 EmF = it9 + it10 + it17 + it18 CogF = it13 + it19	857	33	237.00	32	0.094	0.046	0.926	0.896	0.694 (EmF)	0.396 (EmF)	0.747 (PF / EmF)	26465.26	26308.40	30.73

Model E Smets et al. with non-reverse worded items only	GF = it5 + it16 PF = it2 + it14 RA = it10 + it17 RM = it9 + it18 MF = it13 + it19	857	40	63.6	25	0.047	0.022	0.986	0.975	0.523 (RM)	0.518 (RM)	0.912 (GF/PF)	26307.15	26117.01	30.51
Model F Smets et al. with reverse worded items only	GF = it1 + it12 PF = it8 + it20 RA = it3 + it6 RM = it4 + it15 MF = it7 + it11	857	40	269	25	0.115	0.064	0.912	0.841	0.282 (RM)	0.198 (RM)	1.350 (RM/RA)	25236.72	25046.58	29.26

Note: χ^2: chi-squared test; DF: degrees of freedom; RMSEA: root mean square error of approximation; SRMR: Standardized root mean square residuals; CFI: comparative fit index; TLI: Tucker Lewis Index; Alpha: level of significance; AVE: average variance extracted; HTMT: Heterotrait-Monotrait ratio of correlations; BIC: Bayesian Information Criteria; AIC: Akaike information criterion; ECVI: expected cross validation index; GF: general fatigue; PF: physical fatigue; RA: reduced activity; RM: reduced motivation; MF: mental fatigue.

until the model stabilized into a final version with good fit. We followed a similar fashion with the other models. Model C faced identification issues in the initial CFA due to overlap between RM and RA (HTMT = 0.954). Model F also had identification issues, with overlaps between GF and PF (HTMT = 1.019) and between RM and RA (HTMT = 1.176). Table 2 presents the fit and reliability indicators for the final refined (*r*) models.

Model AR is composed by the original PF and MF factors. Model DR keeps its original PF and CofF factors, and Model ER retains its original PF, RA, MF factors. These three models showed good fit, good reliability, good convergent validity of each remaining factor and good discriminant validity between factors. Model BR showed poor fit indicators, Model CR had mixed results with unacceptable RMSEA and TLI but a good CFI. Model FR was reduced to a just-identified single factor with only two items, resulting in a perfect fit but limited explanatory interest. They were removed from further analysis.

Before proceeding, and given the significant loss of items and factors, we conducted a quick calculation of each item's success rate (i.e. the ratio of the number of times an item was retained relative to the number of possible models) in the refined models (see Supplementary, Table 2). The item success rate analysis revealed varying degrees of alignment between items and their respective factors. Items associated with MF generally have high success rates, with items 7, 11, and 13 achieving 100% or 80%, indicating strong representation in potential models. Physical fatigue (PF) items also perform well, with item 14 reaching 100% success. General fatigue (GF) items show moderate success, notably item 12 with 75%, though items like 1 and 16 vary around 50-60%. In contrast, reduced motivation (RM) items consistently show lower success rates (0–50%), indicating a weaker fit in the models. This variability suggests that while MF and PF are well-represented, RM may require further model adjustments for better alignment.

If we consider that a successful item must be present in at least two different models, we see that all items from GF, PF and MF are successful, only two items from RA pass the criterion, and that only one item (it.18) from RM remains.

We conducted the AR, DR and ER models through a study of measurement invariance between test and validation samples and between sexes.

The analysis of invariance across samples and sexes for models AR, DR, and ER (see Supplementary Table 3) demonstrated good model fit, with evidence of configural, metric, and scalar invariance across groups. For example, model AR, showed configural invariance across samples (CFI = 0.970, RMSEA = 0.077), and maintained stable fit across metric and scalar levels (ΔCFI = 0.000, ΔRMSEA ≤ -0.006). Across sexes, metric invariance showed a slight decline (ΔCFI = −0.002), while means invariance revealed a more notable drop (ΔCFI = −0.006, $p < .001$). Model DR also showed good configural fit across samples (CFI = 0.979, RMSEA = 0.070), with minimal changes at metric and scalar levels (ΔCFI $\leq$ −0.001, ΔRMSEA ≤ -0.008). Across sexes, invariance testing indicated gradually decreasing fit, particularly for means invariance (ΔCFI = −0.014, ΔRMSEA = 0.011, $p < .001$). Model ER demonstrated very good configural fit across samples (CFI = 0.997, RMSEA = 0.030) and strong stability across all subsequent levels of invariance (ΔCFI $\leq$ 0.001, ΔRMSEA $\leq$ 0.004). Across sexes, fit remained consistent through scalar invariance (ΔCFI = 0.001, ΔRMSEA = −0.006), but means invariance showed a more marked decline (ΔCFI = −0.012, ΔRMSEA = 0.029, $p < .001$). Overall, the models exhibit satisfactory invariance, supporting their reliability across groups, despite some expected minor fit adjustments for specific invariance tests across sexes.

Table 2. Fit and reliability indicators for the final refined models.

Refined Final Models	Factors (*reversed items)	*N*	Free Parameters	χ^2	DF	RMSEA	SRMR	CFI	TLI	Lowest Alpha (factor)	Lowest AVE (factor)	Highest HTMT (factor/ factor)	BIC	AIC	ECVI
Model A Smets et al. (1995) R [1]	PF = it2 + it8 + it14 + it20 MF = it7 + it11 + it13 + it19	857	25	85.70	19	0.076	0.028	0.971	0.957	0.881 (MF)	0.519 (PF)	0.589 (PF/ MF)	19659.29	19540.46	22.83
Model B Gentile et al. (2003) R [30]	GF = it1 + it12 + it2*+it8 + it20 + it3 + it5 + it16* +it14 RA = it6 + it10* +it17 MF = it7 + it11 + it13 + it19 + it18	857	54	1028.00	116	0.104	0.077	0.868	0.845	0.801 (RA)	0.496 (GF)	0.722 (GF/ MF)	41919.19	41662.50	48.67
Model C Fillion et. al (2003) R [24]	GPF = it1 + it20 + it5 + it16 + it14 + it12 MF = it11 + it7	857	25	226.00	19	0.124	0.053	0.927	0.893	0.856 (MF)	0.530 (GPF)	0.634 (GPF/ MF)	19450.34	19331.51	22.58
Model D Baussard et al. (2018) R [29]	PF = it2 + it5 + it14 + it16 CogF = It13 + it19	857	19	40.90	8	0.077	0.027	0.979	0.961	0.799 (CogF)	0.518 (PF)	0.687 (GF/ CogF)	15520.38	15430.06	18.03
Model E Smets et al. with non-reverse worded items only R	PF = it2 + it14 RA = it10 + it17 MF = it13 + it19	857	21	9.83	6	0.030	0.011	0.997	0.993	0.720 (PF)	0.569 (PF)	0.590 (PF/ MF)	16105.82	16006.00	18.70
Model F R Smets et al. with reverse worded items only R	MF = it7 + it11	857	5	0.0	0.0	0.0	0.0	0.0	0.0	0.856	0.76		4761.03	4737.26	5.53

Note: χ^2: chi-squared test; DF: degrees of freedom; RMSEA: root mean square error of approximation; SRMR: Standardized root mean square residuals; CFI: comparative fit index; TLI: Tucker Lewis Index; Alpha: level of significance; AVE: average variance extracted; HTMT: Heterotrait-Monotrait ratio of correlations; BIC: Bayesian Information Criteria; AIC: Akaike information criterion; ECVI: expected cross validation index; GF: general fatigue; PF: physical fatigue; RA: reduced activity; RM: reduced motivation; MF: mental fatigue.

Considering the results and theoretical considerations, we fitted Model AR in the global sample. Psychometric analysis revealed that RMSEA (0.076), and SRMR (0.028) were both within acceptable ranges. The CFI (0.971) and TLI (0.957), indicate a good fit. Reliability was strong for both the performance factors (PF) and measurement factors (MF), with α values of 0.798 and 0.881 and ω values of 0.807 and 0.880, respectively. The Average Variance Extracted (AVE) was 0.519 for PF and 0.648 for MF, and the Heterotrait-Monotrait Ratio (HTMT) was 0.589, further supporting the reliability and validity of the constructs measured in the global sample. We added a 2nd level factor to the model, named Global Fatigue (GlobF) to avoid confusion with the original GF[1]; to prevent identification problems we imposed an equality constraint on 2nd order βs across the model. GlobF trajectories were statistically significant ($p < .001$) with βs = 0.766 but its McDonalds $\acute{\omega}$ was 0.650, a lowest threshold for acceptable reliability. An explanatory diagram illustrating the factorial weights can be found in Figure 1.

Discussion

This study identified a refined factor model for the Multidimensional Fatigue Inventory (MFI) with the best fit for a Portuguese sample, advancing its psychometric validation. Among the six models examined, three (Models B, C, and F) displayed persistent global fit issues, which could not be resolved through model refinement. To preserve the integrity of the existing structure, we opted not to re-specify these models further, focusing instead on evaluating models that more accurately represented the data. Except for

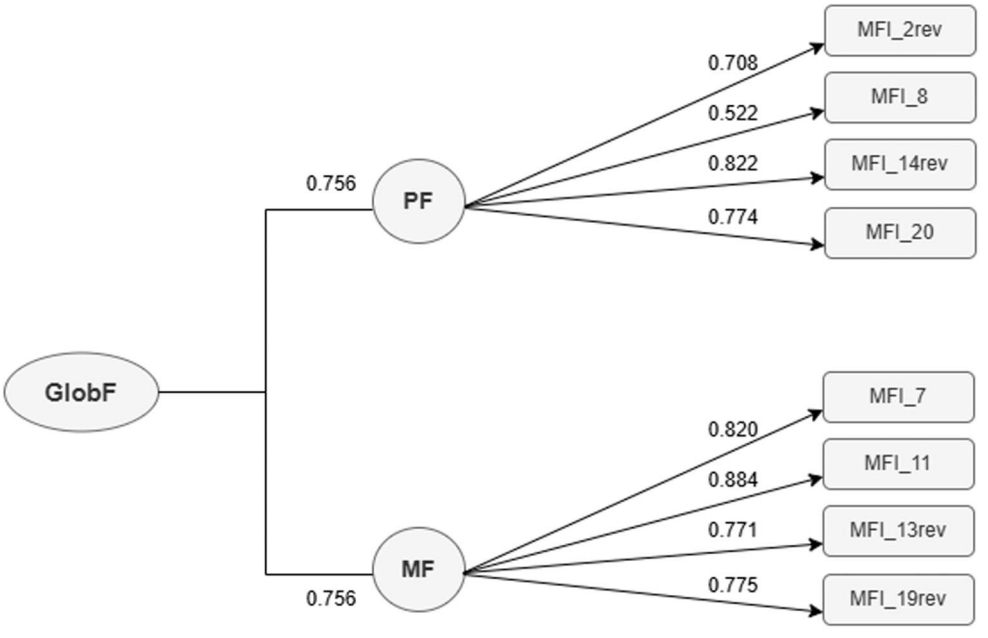

Figure 1. Factorial weights of the Multidimensional Fatigue Inventory **(**MFI) fitted model AR. Second-order model factor reduced version (8 items) structure fit for the sample ($n = 857$). Numeric values are the factor loadings for each first-orders factors and their respective items. GlobF – Global fatigue; PF – Physical fatigue; MF – Mental fatigue; rev – reverse item.

Model E [29] all models had initial problems with strong overlap between factors. This may explain why so many studies have difficulty in finding the original structure, with items loading across multiple factors in EFAs, and with factors being merged. It also supports the idea of abandoning the General Fatigue factor since it duplicates Physical Fatigue a fact identified in the original paper [1], and reproduced in subsequent studies (see [24, 25, 28, 30, 50, 51]). If there is need for a global fatigue score, a second order factor can be used. Refining the models led to a large loss of items, and an analysis of this loss showed that it is far from equal across factors. GF, PF, and MF items are far more likely to be retained than RA items, and three of the RM fatigue items were dropped in all the models they took part in; the fourth, is 18 (*I don't feel like doing anything*, originally from RM in Smets et al. [1]) was only retained in the refined form of Gentile et al. [30], who had moved it to the MF factors. The reason for these overlaps and losses resides at item level (the observed variable) and item 18 is a good example: *not doing anything* is semantically related to reduced motivation, *not feeling like* to mental fatigue. It is difficult to express in an item a construct for which there is no clear definition. Without a clear definition of fatigue, it is difficult to define its dimensions. Raizen et al. [3] specifically exclude reduced motivation from their definition of fatigue. Reduced motivation, and reduced activity, can be viewed as a consequence of fatigue and not a manifestation of fatigue itself, something admitted for reduced activity by Smets et al. [1] in a similar fashion to the loss of personal satisfaction which was once viewed as a constituent factor of burnout [52], and is now viewed as a consequence of burnout and not part of the syndrome itself [53]. Some consideration should be given to removing them from the model for this reason. Our Model AR keeps the same PF and MF factors of the original model, minus RA and RM for poor psychometric qualities, and minus GF due to its redundancy with PF and broader nature. It is for this reason that we considered it to be the best model of our analysis, even though models DR and ER had identical or even better fit and invariance indices. Beyond the fit indices, the consensus version was designed to distinguish between physical and mental fatigue, emphasizing clinical relevance and theoretical clarity. Items related to general fatigue were excluded for their broader nature. Also, items related to reduced motivation, for example, were also excluded as they did not directly contribute to distinguishing physical and mental fatigue. Specifically, items like 'I don't feel like doing anything' (item18) and 'I dread having to do things' (item 9) overlap with emotional disengagement and may better reflect depressive symptoms than fatigue. Additionally, items related to Reduced Motivation (RM) (items 4, 9, 15, 19) had low success rates (see Table S2), ranging from 0% to 50%, further supporting their exclusion. With four items per dimension, the model is balanced, efficient, and low burden, making it suitable for both clinical and non-clinical use, while also demonstrating strong psychometric properties.

In psychometric studies of the MFI, we recommend starting with confirmatory factor analysis and using exploratory factor analysis only if theory-based models fail. Exploratory analysis always yields the best model for a given method, but the key is whether a predefined model fits the data, not what the data suggests. Reliability should be assessed only after validating the factor structure, as it alone does not confirm validity. Factor models should be tested in the general population before use in specific groups (e.g.

patients, athletes, shift workers). If a population differs significantly, a tailored version should be tested across samples to ensure stability.

Our results add evidence to the contention that reverse-worded items have a negative impact on psychometric measurement [38]. One would expect Models E (only straight worded, high fatigue, items) and F (only reverse worded, low fatigue, items) to perform similarly. However, Model E consistently outperformed all other models, despite exhibiting fit issues from the beginning, which ultimately led to its reduction to two items within a single factor. One possible way solution to some of the MFI issues could be to rephrase its reverse worded items, ensuring they do not overlap with the straight worded items, and then assess the stability of the factor models.

Inconsistencies in the MFI's psychometric properties may stem from the lack of a clear, consensual definition of fatigue and its dimensions. A precise definition could clarify the roles of Reduced Motivation and Reduced Activity, facilitating the creation of items that represent fatigue accurately without using reverse-worded sentences.

As some authors have demonstrated [18, 30, 54], the report of fatigue symptoms differs between sexes, with women generally reporting higher fatigue scores and exhibiting greater variance in these scores compared to men. This suggest that, while overall differences between men and women are minimal, a disproportionate number of women experience very high levels of fatigue. In this study, the MFI performed well for both sexes, showing scalar invariance across all gender models, thereby validating its use as a reliable instrument for detecting sex differences in fatigue levels.

Our study presents some limitations. The global sample was collected from two sociodemographic groups using non-probabilistic methods via social media, which may have introduced selection bias and limited the representativeness of the sample. While online surveys offer significant advantages – such as broader geographic reach, cost-effectiveness, and the ability to recruit large and diverse samples efficiently [55, 56] – they also present inherent limitations. These include potential self-selection bias, limited control over the testing environment, and challenges in verifying participant identity and engagement [57]. In our case, this approach resulted in a gender imbalance, particularly in the higher education sample, which included a larger proportion of females compared to males. The use of a non-random sample limits the generalizability of our conclusions. While efforts were made to gather a large and diverse sample size ($n > 1000$) to capture the natural variation within the study population, the lack of random sampling introduces potential biases that may affect the extent to which our findings can be applied to the broader population. Reliability was assessed only through Cronbach's α and McDonald's ω, without temporal stability. Validity evidence, such as convergent, discriminant, and criterion validity, was not evaluated. The sample lacked diversity in factors like age, gender, socioeconomic status, and cultural background, limiting generalizability to broader populations. Future research employing randomized sampling methods is encouraged to further validate these results including other Portuguese groups from different regions of the country as well as Portuguese-speaking countries from diverse cultural contexts to broaden the evidence based. Additionally, a systematic review or even a meta-analysis of the instrument could provide stronger evidence for its relevance and importance. Despite these limitations, we refined the MFI, distinguishing two key dimensions of fatigue (mental and physical), enabling a clearer understanding of their impact on daily performance.

Conclusions

In this study, we propose a consensus version of the MFI that demonstrates good psychometric properties for a normative population. This refined version represents a concise, consensus-driven instrument that combines the most robust items from previous versions and distinguishes between two key dimensions of fatigue: physical and mental. Focusing on these rigorously selected items, the scale offers investigators greater confidence in the reliability and validity of the results, while maintaining efficiency and ease of use in both research and clinical and non-clinical settings. The revised MFI shows strong evidence for its internal structure, including factorial validity, reliability, and measurement invariance across independent samples and sexes. Furthermore, it addresses the variability and lack of consistent validity observed in earlier MFI versions, providing a reliable tool for both researchers and clinicians. Future studies should aim to validate this version in more diverse populations, including both clinical and non-clinical groups, to further establish its broader applicability and effectiveness. It is important that future research uses stratified random sampling to ensure representation across all relevant strata. This includes the general population, special populations (air traffic controllers, pilots, professors or athletes) and clinical groups such as individuals with sleep disorders and other medical disorders.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Cátia Reis, PhD is a Chronobiologist and somnologist.

Madalena L. Santos is an MSc student in neurosciences.

João Marôco is a Full Professor in statistics.

ORCID

Cátia Reis http://orcid.org/0000-0001-6585-3993

References

[1] Smets E, Garssen B, Bonke B, et al. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res. 1995;39(3):315–325. doi:10.1016/0022-3999(94)00125-O

[2] Dittner AJ, Wessely SC, Brown RG. The assessment of fatigue. J Psychosom Res. 2004;56(2):157–170. doi:10.1016/S0022-3999(03)00371-4

[3] Raizen DM, Mullington J, Anaclet C, et al. Beyond the symptom: the biology of fatigue. Sleep. 2023;46(9):1–13. doi:10.1093/sleep/zsad069

[4] Malley D. Deluca, J. (Ed.). (2005). Fatigue as a window to the brain. London: The MIT Press. Neuropsychol Rehabil. 2006;16(5):597–599. doi:10.1080/09602010600685210

[5] Chaudhuri A, Behan PO. Fatigue and basal ganglia. J Neurol Sci. 2000;179(1–2):34–42. doi:10.1016/S0022-510X(00)00411-1

[6] Billones R, Liwang JK, Butler K, et al. Dissecting the fatigue experience: a scoping review of fatigue definitions, dimensions, and measures in non-oncologic medical conditions. Brain Behav Immun Health. 2021;15:100266. doi:10.1016/j.bbih.2021.100266
[7] Peierdun M, Liu W, Renaguli A, et al. Clinical characteristics of abnormal savda syndrome type in human immunodeficiency virus infection and acquired immune deficiency syndrome patients: a cross-sectional investigation in Xinjiang, China. Chin J Integr Med. 2015;21(12):895–901. doi:10.1007/s11655-015-2075-8
[8] Rosas JC, Aguado-Barrera ME, Azria D, et al. (Pre)treatment risk factors for late fatigue and fatigue trajectories following radiotherapy for breast cancer. Int J Cancer. 2023;153(9):1579–1591. doi:10.1002/ijc.34640
[9] Doumen M, Pazmino S, Bertrand D, et al. Longitudinal trajectories of fatigue in early RA: the role of inflammation, perceived disease impact and early treatment response. Ann Rheum Dis. 2022;81(10):1385–1391. doi:10.1136/annrheumdis-2022-222517
[10] Knoop V, Cloots B, Costenoble A, et al. Fatigue and the prediction of negative health outcomes: a systematic review with meta-analysis. Ageing Res Rev. 2021;67:101261. doi:10.1016/j.arr.2021.101261
[11] Luckhaupt SE. Short sleep duration among workers – United States, 2010. Morb Mortal Wkly Rep. 2012;61(16):281.
[12] Sadeghniiat-Haghighi K, Yazdi, Z. Fatigue management in the workplace. Ind Psychiatry J. 2015;24(1):12–17. doi:10.4103/0972-6748.160915
[13] Cunningham TR, Guerin RJ, Ferguson J, et al. Work-related fatigue: a hazard for workers experiencing disproportionate occupational risks. Am J Ind Med. 2022;65(11):913. doi:10.1002/ajim.23325
[14] Fatigue – National Safety Council [Internet]. [cited 2025 Jan 20]. Available from: https://www.nsc.org/work-safety/safety-topics/fatigue
[15] Dittner AJ, Wessely SC, Brown RG. The assessment of fatigue: a practical guide for clinicians and researchers. J Psychosom Res. 2004;56(2):157–170. doi:10.1016/S0022-3999(03)00371-4
[16] Okuyama T, Akechi T, Kugaya A, et al. Development and validation of the cancer fatigue scale. J Pain Symptom Manage. 2000;19(1):5–14. doi:10.1016/S0885-3924(99)00138-4
[17] Chuang L-L, Chuang Y-F, Hsu M-J, et al. Validity and reliability of the traditional Chinese version of the multidimensional fatigue inventory in general population. PLoS One. 2018;13(5):e0189850.
[18] Watt T. Fatigue in the Danish general population. Influence of sociodemographic factors and disease. J Epidemiol Community Health. 2000;54(11):827–833. doi:10.1136/jech.54.11.827
[19] Boter H, Mänty M, Hansen AM, et al. Self-reported fatigue and physical function in late mid-life. J Rehabil Med. 2014;46(7):684–690. doi:10.2340/16501977-1814
[20] Baptista RLR, Biasoli I, Scheliga A, et al. Psychometric properties of the multidimensional fatigue inventory in Brazilian Hodgkin's lymphoma survivors. J Pain Symptom Manage. 2012;44(6):908–915. doi:10.1016/j.jpainsymman.2011.12.275
[21] Lopes J, Araújo HAGO, Smaili SM, et al. Brazilian version of the multidimensional fatigue inventory for Parkinson's disease. Fisioterapia em Movimento. 2020;33:e003362.
[22] Le Gal M, Mainguy Y, Le Lay K, et al. Linguistic validation of six patient-reported outcomes instruments into 12 languages for patients with fibromyalgia. Joint Bone Spine. 2010;77(2):165–170. doi:10.1016/j.jbspin.2010.01.005
[23] Smets EMA, Garssen B, Cull A, et al. Application of the multidimensional fatigue inventory (MFI-20) in cancer patients receiving radiotherapy. Br J Cancer. 1996;73(2):241–245. doi:10.1038/bjc.1996.42
[24] Fillion L, Gélinas C, Simard S, et al. Validation evidence for the French Canadian adaptation of the Multidimensional Fatigue Inventory as a measure of cancer-related fatigue. Cancer Nurs. 2003;26(2):143–154. doi:10.1097/00002820-200304000-00008
[25] Elbers RG, van Wegen EEH, Verhoef J, et al. Reliability and structural validity of the Multidimensional Fatigue Inventory (MFI) in patients with idiopathic Parkinson's disease. Parkinsonism Relat Disord. 2012;18(5):532–536. doi:10.1016/j.parkreldis.2012.01.024

[26] Chuang LL, Chuang YF, Hsu MJ, et al. Validity and reliability of the traditional Chinese version of the multidimensional fatigue inventory in general population. PLoS One. 2018;13(5): e0189850. doi:10.1371/journal.pone.0189850
[27] Gecaite-Stonciene J, Bunevicius A, Burkauskas J, et al. Validation of the multidimensional fatigue inventory with coronary artery disease patients. Int J Environ Res Public Health. 2020;17(21):1–17.
[28] Binz C, Osmanovic A, Thomas NH, et al. Validity and reliability of the German multidimensional fatigue inventory in spinal muscular atrophy. Ann Clin Transl Neurol. 2022;9(3):351–362. doi:10.1002/acn3.51520
[29] Baussard L, Carayol M, Porro B, et al. Fatigue in cancer patients: development and validation of a short form of the Multidimensional Fatigue Inventory (MFI-10). Eur J Oncol Nurs. 2018;36: 62–67. doi:10.1016/j.ejon.2018.07.005
[30] Gentile S, Delarozière JC, Favre F, et al. Validation of the French "Multidimensional Fatigue Inventory" (MFI 20). Eur J Cancer Care. 2003;12(1):58–64. doi:10.1046/j.1365-2354.2003.00295.x
[31] Tian J, Hong J-S. Validation of the Chinese version of Multidimensional Fatigue Inventory-20 in Chinese patients with cancer. Support Care Cancer. 2012;20(10):2379–2383. doi:10.1007/s00520-011-1357-8
[32] Buss T, Kruk A, Wiśniewski P, et al. Psychometric properties of the Polish version of the multidimensional fatigue inventory-20 in cancer patients. J Pain Symptom Manage. 2014;48(4): 730–737. doi:10.1016/j.jpainsymman.2013.11.015
[33] Chandel P, Sultan A, Khan KA, et al. Validation of the Hindi version of the Multidimensional Fatigue Inventory-20 (MFI-20) in Indian cancer patients. Support Care Cancer. 2015;23(10):2957–2964. doi:10.1007/s00520-015-2661-5
[34] Hinz A, Benzing C, Brähler E, et al. Psychometric properties of the Multidimensional Fatigue Inventory (MFI-20), derived from seven samples. J Pain Symptom Manage. 2020;59(3): 717–723. doi:10.1016/j.jpainsymman.2019.12.005
[35] Kieffer JM, Starreveld DE, Boekhout A, et al. A questionable factor structure of the Multidimensional Fatigue Inventory in the general Dutch population. J Clin Epidemiol. 2021;137:266–276. doi:10.1016/j.jclinepi.2021.05.005
[36] Maroco J, Maroco AL, Campos JADB. Student's academic efficacy or inefficacy? An example on how to evaluate the psychometric properties of a measuring instrument and evaluate the effects of item wording. Open J Stat. 2014;4(6):484–493. doi:10.4236/ojs.2014.46046
[37] Assunção H, Lin S-W, Sit P-S, et al. University student engagement inventory (USEI): transcultural validity evidence across four continents. Front Psychol. 2019;10:2796. doi:10.3389/fpsyg.2019.02796
[38] van Sonderen E, Sanderman R, Coyne JC. Ineffectiveness of reverse wording of questionnaire items: let's learn from cows in the rain. PLoS One. 2013;8(7):e68967. doi:10.1371/journal.pone.0068967
[39] Vallejo MA, Jordán CM, Díaz MI, et al. Psychological assessment via the internet: a reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). J Med Internet Res. 2007;9(1):e2. doi:10.2196/jmir.9.1.e2
[40] American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. American Educational Research Association; 2014.
[41] Peng J, Hahn J, Huang K-W. Handling missing values in information systems research: a review of methods and assumptions. Inf Syst Res. 2023;34(1):5–26. doi:10.1287/isre.2022.1104
[42] Finney SJ, DiStefano C. Nonnormal and categorical data in structural equation modeling. In: Hancock GR, Mueller RO, editors. Structural equation modeling: a second course. 2nd ed. Charlotte: IAP Information Age Publishing; 2013. p. 439–492.
[43] Marôco J. Análise de Equações Estruturais: Fundamentos Teóricos, Software & Aplicações. ReportNumber, Lda; 2021.

[44] GitHub – semlj/semlj: structural equation models based on lavaan for Jamovi [Internet]. [cited 2024 Dec 10]. Available from: https://github.com/semlj/semlj
[45] Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw. 2012;48:1–36. doi:10.18637/jss.v048.i02
[46] Henseler J. Bridging design and behavioral research with variance-based structural equation modeling. J Advert. 2017;46(1):178–192. doi:10.1080/00913367.2017.1281780
[47] Hayes AF, Coutts JJ. Use omega rather than Cronbach's alpha for estimating reliability. But … . Commun Methods Meas. 2020;14(1):1–24. doi:10.1080/19312458.2020.1718629
[48] Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev Rev. 2016;41:71–90. doi:10.1016/j.dr.2016.06.004
[49] Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Modeling. 2002;9(2):233–255. doi:10.1207/S15328007SEM0902_5
[50] Lin JMS, Brimmer DJ, Maloney EM, et al. Further validation of the Multidimensional Fatigue Inventory in a US adult population sample. Popul Health Metr. 2009;7:18. doi:10.1186/1478-7954-7-18
[51] Saffari M, Naderi MK, Piper CN, et al. Multidimensional Fatigue Inventory in people with hepatitis B infection. Gastroenterol Nurs. 2017;40(5):380–392. doi:10.1097/SGA.0000000000000250
[52] Maslach C, Jackson SE. The measurement of experienced burnout. J Organ Behav. 1981;2(2):99–113. doi:10.1002/job.4030020205
[53] Lee RT, Ashforth BE. A meta-analytic examination of the correlates of the three dimensions of job burnout. J Appl Psychol. 1996;81(2):123–133. doi:10.1037/0021-9010.81.2.123
[54] Schwarz R, Krauss O, Hinz A. Fatigue in the general population. Onkologie. 2003;26(2):140–144.
[55] Evans JR, Mathur A. The value of online surveys: a look back and a look ahead. Internet Res. 2018;28(4):854–887. doi:10.1108/IntR-03-2018-0089
[56] Wright KB. Researching internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. J Comput Mediat Commun. 2005;10(3):JCMC1034.
[57] Andrade C. The limitations of online surveys. Indian J Psychol Med. 2020;42(6):575–576. doi:10.1177/0253717620957496