Assessment of fatigue among working people: a comparison of six questionnaires

J De Vries, H J Michielsen, G L Van Heck

F

atigue is one of the major complaints in primary care settings. Also in general population studies, fatigue is commonly reported (14–22%). In a recent extensive study, it was found that about 25% of Dutch employees report fatigue at work. Another Dutch study showed that over one third of the recipients of work disability benefit are occupationally disabled on mental grounds. The majority of these individuals suffer from chronic job stress and burnout. The most characteristic component of burnout is emotional exhaustion, a fatigue related concept. Emotional exhaustion refers to feelings of being overextended and depleted of one’s emotional and physical resources.

Fatigue is defined as an experience of tiredness, dislike of present activity, and unwillingness to continue, or as a “disinclination to continue to performing the task at hand and a progressive withdrawal of attention” from environmental demands. As a gradual and cumulative process, fatigue reflects vigilance decrement and decreased capacity to perform, along with subjective states that are associated with this decreased performance. It is a general psychophysiological phenomenon that diminishes the ability of the individual to perform a particular task by altering alertness and vigilance, together with the motivational and subjective states that occur during this transition. As a consequence, there is reduced competence and willingness to develop or maintain goal directed behaviour aimed at adequate performance. This view of fatigue is used in the present study.

There is no standard way to assess fatigue. Fatigue can be measured objectively as well as subjectively. Objective fatigue measures focus on physiological processes or performance such as reaction time or number of errors. Subjective ways to assess fatigue include diary studies, interviews, and questionnaires. Often, questionnaires are used in large scale studies because of their shortness and self report format.

Until about 10 years ago, fatigue questionnaires for particular studies were mainly developed on an ad hoc basis. Two recent reviews showed that most fatigue questionnaires are developed for specific patient groups, such as patients with cancer, or ill persons in general. Little is known, however, about the applicability of these questionnaires in healthy populations, although several fatigue measures are claimed to be useful in patient populations as well as in groups of healthy persons. One of the few questionnaires that has been developed for use in hospital populations as well as community populations is the Fatigue Scale (FS). Furthermore, fatigue is also frequently measured using subscales of broader measures. The Emotional Exhaustion Scale in burnout questionnaires (e.g. MBI) and the Energy and Fatigue Scale of the World Health Organisation Quality of Life assessment instrument (WHOQOL-100) are good examples of this approach.

Before the start of the 1990s, fatigue was predominantly seen as a unidimensional construct. Nowadays, many authors conceive of fatigue as a multidimensional construct. For instance, Smets and coworkers discern five components: general fatigue, physical fatigue, reduction in activity; reduction in motivation, and mental (cognitive) fatigue. Others, for instance, Schwartz and coworkers, have developed a three dimensional scale, distinguishing situation specific fatigue, consequences of fatigue, and response to rest/sleep. The two reviews mentioned state that multidimensional fatigue scales are seen as more comprehensive, and

Abbreviations: CFS, chronic fatigue syndrome; CIS-20, Checklist Individual Strength; CIS-PA, Checklist Individual Strength-reduced Physical Activity; CIS-SEF, Checklist Individual Strength-Subjective Experience of Fatigue; CON, reduced concentration; EE, emotional exhaustion; EF-WHOQOL-100, Energy and Fatigue scale World Health Organisation Quality of Life assessment instrument; FAS, Fatigue Assessment Scale; FS, Fatigue Scale; H, scalability coefficient; MBI-EE, Maslach Burnout Inventory—Emotional Exhaustion; MBI-NL, Maslach Burnout Inventory—Netherlands; MBI, Maslach Burnout Inventory; MOT, reduced motivation; MS, multiple sclerosis; MSF, Mokken Scale analysis for polytomous items; NRS, Need for Recovery Scale; NWQO, Netherlands Organisation for Scientific Work; PA, reduced physical activity; SD, standard deviation; SEF, subjective experience of fatigue; SPSS, Statistical Package for the Social Sciences; VBBA, questionnaire on perception and judgement of work; WHOQOL-100, World Health Organisation Quality of Life assessment instrument; WORC, Work & Organisation Research Centre

Main messages

• The examined fatigue scales are reliable and valid.
• Fatigue is unidimensional in a working population.
• For a working population the FAS is the most promising fatigue measure.

Policy implications

• A distinction in types of fatigue could not be found. Hence, employees experience overall fatigue.
hence as more adequate for providing a complete description of an individual's fatigue experience. However, convincing empirical evidence for the multidimension assumption is still lacking. Moreover, two studies have recently shown that fatigue is best conceived of as a unidimensional construct.

The aim of the present study was to determine the psychometric qualities of several types of fatigue questionnaires in a sample of working persons. The questionnaires were selected on the basis of their use in working and healthy populations, and most were part of the prescribed set of fatigue questionnaires used in a nationwide project. This was done through a strength–weakness analysis that focused on exploring and testing: (1) internal consistency and test–retest reliability; (2) content validity; (3) convergent validity; and (4) the dimensionality of the fatigue instruments. The selected questionnaires are reliable, valid, and frequently employed.

**METHODS**

**Respondents**

Participants (n = 351) were recruited through random telephone calls. A random digit dialing method was used to telephone potential participants. If they agreed to participate, the questionnaire was sent by post. All respondents worked at least 20 hours per week (mean 35.3 hours, SD 9.06; 25% worked 30 hours or less and 25% worked 36–40 hours). In total, 183 men (mean age 45 years, SD 8.4) and 166 women (mean age 43 years, SD 9.5) participated (total response 48%). Gender was unknown for two respondents. Twenty seven per cent of the respondents were single, while 638 persons (73%) were married or living with a partner. Forty six per cent (n = 399) had a college education. Lower educated people were somewhat underrepresented and highly educated persons slightly overrepresented. However, this is not uncommon for this kind of study. With respect to gender, marital status, and age, the data are representative for the Dutch working population. The branch that the participants worked in were: industry/agriculture (n = 32), construction (n = 21), trade/repairs/hotels (n = 33), transport (n = 8), financial services (n = 33), care sector (n = 71), other services (n = 37), public sector (government) (n = 39), education (n = 38), and unknown (n = 13). Concerning being ill, the question was asked: were you ill during the past week? People (n = 25) who were ill, indicated that they were having a cold (n = 13), or had a health problem such as back pain, asthma, or a chronic illness. These individuals were not excluded, because the illness was not severe enough that they had to stop working for more than a week.

**Questionnaires**

The Checklist Individual Strength-20 (CIS-20) consists of 20 statements and provides a total fatigue score and scores for four components of fatigue: subjective experience of fatigue (SEF; eight items), reduced concentration (CON; five items), reduced motivation (MOT; four items), and reduced physical activity level (PA; three items). Respondents use a seven point rating scale (1, yes, that is true, to 7, no, that is not true). The reliability of the CIS is good. Furthermore, the CIS yielded different scores for CFS patients, multiple sclerosis (MS) patients, and patients with abdominal pain. Moreover, the subscales of the CIS correlated significantly with comparable scales. A total score above 76 is considered high. Although the CIS was developed for CFS patients, the questionnaire is claimed to be also appropriate for healthy populations. In a number of recent studies among working persons only the total CIS score has been used, while in other investigations one or more subscales have been employed. In the present study, we evaluated the total score as well as subscale scores in order to provide a complete picture concerning the CIS.

The Emotional Exhaustion subscale (EE scale) of the Dutch version of the Maslach Burnout Inventory (MBI; MBI-NL), comprises five items, each with a seven point rating scale ranging from 1, never, to 7, always. The scale has well established validity and high internal consistency. A score above 2.20 is considered high.

The Energy and Fatigue subscale from the World Health Organisation Quality of Life assessment instrument (EFWHOQOL-100; Dutch version) contains four items. Answers are given on a five point Likert scale (1, never, to 5, always); two positively phrased items using the term “energy” and two negatively phrased featuring the word “fatigue”. This scale has been found to have a good reliability and to have excellent convergent validity.

The 11 item Fatigue Scale (FS; Dutch translation) distinguishes mental fatigue (four items), describing cognitive difficulties, and physical fatigue (seven items). This measure uses a five point rating scale (1, never, to 5, always). It is also possible to calculate a total fatigue score. The scale was found to be both reliable and valid and has shown sensitivity to treatment changes.

The Need for Recovery scale (NRS) from the Questionnaire on Perception and Judgement of Work (VBBA) is designed to measure the short term effects of a day of work. The 11 items are rated on a dichotomous scale, Yes-No. Reliability and validity of the scale are good.

The 10 item Fatigue Assessment Scale (FAS) is a new fatigue scale that was developed in large samples of the Dutch working and general population. The items were selected from an initial item pool consisting of 40 items taken out of existing fatigue questionnaires and represent physical (five items) and mental fatigue (five items). Despite this, based on factor analyses and Mokken Scale analyses, the FAS is considered unidimensional and consequently, only a total score is calculated. The instruction of the FAS is directed at how a person usually feels. The five point rating scale varies from 1, never, to 5, always. Reliability and validity appear to be good. The FAS has a reliability of 0.90 and does not measure emotional stability or depression.

**Statistical procedure**

Means, standard deviations, and Cronbach’s alpha coefficients were calculated for each (sub)scale. Cronbach alpha is a general formula for estimating the reliability of a test by looking at inter-item consistency. It measures how well a set of items measures a single unidimensional latent construct. It is mathematically equivalent to the average of all possible split-half estimates of reliability. Pearson correlations as well as partial correlations, controlling for overlap in items, were used to examine the associations between all questionnaires (convergent validity). In order to scrutinise the dimensionality of fatigue, exploratory principal components analyses were carried out at different levels: (1) the items per questionnaire (this is also used to establish content validity); (2) the pooled items taken from all questionnaires involved; (3) all subscale scores; and (4) all total scores. The scree test was used to identify the number of factors. This graphical method involves inspection of the plotted eigenvalues, representing the variances extracted by the factors, against these factors and the detection of discontinuity in the slope of the plotted points. Subsequently, the dimensionality of fatigue was examined with Mokken Scale Analysis, but only for the analyses involving single item scores because of the fact that for Mokken Scale analysis only single item scores can be used, not summed scores. For our analyses we used the computer program Mokken Scale analysis for Polytomous items (MSA). This program uses cluster analysis techniques for selecting unidimensional subscales from larger sets of items. Each subscale is selected to optimise the scale H for the subset of items selected (the scale H is a weighted mean of the item scores). For reliably ordering persons on a (sub)scale, the scale H has to be at least 0.3 (default in MSA). However, higher values are
desirable because they indicate higher measurement reliability, and a scale $H > 0.5$ is interpreted as indicative of a strong scale. The quality of individual items as contributors to reliable person ordering is guaranteed by only admitting items to a scale if the item scalability coefficient (item $H$; a weighted mean of all item pairs in which the studied item figures) is at least $0.3$.$^{14}$ Based on recommendations, MSP was used with lowerbounds of 0.0, 0.3, 0.4, and 0.5, respectively, for item selection using all 40 items. MSP is one of the few programs for item response theory analysis$^{15}$ that has an automated item selection procedure. All other analyses were performed using SPSS 9.0.$^{14}$ Missing values were not replaced, because missingness did not exhibit any systematic pattern.

Factor analysis uses the correlations or covariances between items, and is vulnerable to the influence of differences in the items' frequency distributions, which may produce artefactual "difficulty factors".$^{14}$ Mokken Scale analysis is based on the scalability coefficient, $H$,$^{16}$ that equals the ratio of the items' covariance and their maximum covariance given the items' univariate frequency distributions. This way, the effect of different frequency distributions is eliminated. Thus, Mokken Scale analysis does not produce artefacts because of differences in frequency distributions.

RESULTS
Table 1 shows mean and standard deviation of the fatigue questionnaires. Inspection of these results reveals that no excessive high or low scores were found in this sample.

Reliability
Table 1 shows that the internal consistency of the fatigue (sub)scales is satisfactory. The alphas ranged from 0.72 (Fatigue Scale-Mental Fatigue) to 0.96 (CIS total).

Content validity
Content validity is concerned with whether the content of a particular fatigue questionnaire supports the assertion that it is a measure of fatigue. Showing that the questionnaire at hand measures one area of content and not a mixture of all sorts of things contributes to its content validity. Six exploratory factor analyses were performed separately for each of the fatigue questionnaires. The scree plot of each factor analysis clearly revealed one factor. The single factor extracted from each questionnaire explained between 43% (Fatigue Scale) and 75% (EF-WHOQOL-100) of the (observed) variance. Because the FS and the CIS are assumed to reflect a multidimensional structure, additional factor analyses were done for each of these scales with a forced number of factors (two for the FS and four for the CIS). The percentage of explained variance increased to 35% and 76%, respectively. However, all CIS factor loadings were higher on the first factor in comparison to the other factors. In general, the other factor loadings differed by 0.3 from the loading on the first factor. The loadings on the first factor ranged from 0.60 to 0.83, on the second from −0.43 to 0.31, on the third from 0.51 to 0.31, and on the fourth from 0.32 to 0.48. This was also true for the forced factor solution of the FS items. Item 11 (“Do you make slips of the tongue when speaking?”) was an exception to this outcome.

Convergent validity
From a theoretical standpoint, the most important kind of validity is construct validity, reflecting the relation of the questionnaire results to the theoretical concept that the questionnaire is trying to measure. For the construct validity of a particular questionnaire convergent validity is important.$^{17}$ Convergent validity refers to the degree to which the operationalisation of a particular construct is similar to other operationalisations that it theoretically should be similar to.

The correlations, corrected for item overlap, between the (sub)scores of the fatigue questionnaires were moderate to strong, ranging from 0.43 (FS-Mental Fatigue with CIS-PA) to $0.81$ (CIS-SEF with EF-WHOQOL-100) (all $p < 0.001$; see table 2). When looking at associations between subscale scores and total scores for the CIS and the FS, it appeared that the subscales correlated at least 0.78 ($p < 0.001$) with their total score.

Dimensionality
The factor structure and the scalability, using coefficient $H$ of the six questionnaires, were explored separately. As already mentioned, when examining the content validity, the scree plots of the exploratory factor analyses revealed that all questionnaires consisted of one factor. In order to check these findings, Mokken Scale analyses were conducted.

Based on the MSP analyses, again all scales were found to be reliable and unidimensional. The scale $H$ ranged from 0.48 (Fatigue Scale) to 0.78 (EF-WHOQOL-100), as depicted in table 3. Values between 0.4 and 0.5 usually are interpreted as “medium” results; values above 0.5 as good. Only one item, from the Fatigue Scale (FS), had to be excluded from the set of FS items, identical to the results of the factor analysis.

Exploratory factor analysis at the item level, using the total set of pooled items, yielded one factor, that explained 44% of the total variance. Moreover, the MSP analysis in which all fatigue items were taken together, showed that 60 items appeared to form one reliable scale (see table 3). In other words, not only do the questionnaires separately measure one construct, they also measure a similar construct, fatigue.

In addition, a factor analysis on all subscales of the six fatigue questionnaires also revealed one factor, explaining 66% of the variance. In the latter analysis, the FAS had the highest loading on this factor.

Separate analyses revealed that the same strong one factor solution was found, when the sample was split according to
When forcing the number of factors to split into bloated specifics.

29 The finding that the fatigue measures used in the present study have a good internal consistency is in accordance with previous studies employing one or more of these measures. Furthermore, in line with a previous study employing these questionnaires, the content validity of the unidimensional measures was good, while the multifaceted structure of the other instruments (CIS and FS) could not be replicated. The convergent validity of all measures was good. The analyses further revealed that such a single cut off score is indicated more than a combination of cut off points for the four dimensions.

All six fatigue measures were moderately to strongly related to each other, indicating good convergent validity. This finding coincides with a previous study that investigated the psychometric properties of these questionnaires. Lewis and Wessely assume that, when fatigue is measured with emotional, behavioural, and cognitive components, it is likely that the concept is multidimensional. This view also reflects the ideas of Smets and colleagues and Gawron and coworkers, who have stated that nowadays there is general agreement to measure fatigue as a multidimensional concept. The present findings, however, suggest that fatigue should be assessed as a unidimensional phenomenon, at least in a working population. Thus far, statements regarding the multidimensionality of fatigue were based predominantly on the outcomes of factor analyses with the employment of the criterion of eigenvalues greater than 1.0 as indicator for the choice of the number of factors. However, this particular criterion greatly overestimates the number of factors and often causes factors to split into bloated specifics. In the analysis of all fatigue items the percentage of the variance explained was 44%. Additional factors do not increase the percentage substantially. Based on the scree plot this is the best exchange dimensionality of fatigue, and claim a good fit for a multidimensional model. Smets and coworkers, however, did not examine whether a one factor solution would have fit their data equally well. In the present study, where scrutinising the agreement to measure fatigue as a multidimensional concept. The present findings, however, suggest that fatigue should be assessed as a unidimensional phenomenon, at least in a working population. Thus far, statements regarding the multidimensionality of fatigue were based predominantly on the outcomes of factor analyses with the employment of the criterion of eigenvalues greater than 1.0 as indicator for the choice of the number of factors. However, this particular criterion greatly overestimates the number of factors and often causes factors to split into bloated specifics. In the analysis of all fatigue items the percentage of the variance explained was 44%. Additional factors do not increase the percentage substantially. Based on the scree plot this is the best exchange.
A possible reason why the results do not support multidimensionality could be that, compared with groups of predominantly healthy persons, patients focus more on symptoms and, therefore, distinguish more aspects of fatigue. Fatigue may be unidimensional for non-patient groups and multidimensional for patients. When looking at scores of chronic fatigue syndrome patients and multiple sclerosis patients on the CIS subscales, these patients did score higher on all subscales. However, the differences in mean scores between those patient populations and the population in the present study were systematically the same for each subscale—that is, the patients did not show a different pattern of scores on the various subscales. Moreover, Studts and colleagues found no difference in dimensionality between chronic pain patients and healthy controls. Hopefully, the outcomes of this study will reopen the discussion about the dimensionality of fatigue.

For a number of questionnaires used in the present study norm scores are available to indicate problematic fatigue. With regard to the CIS, the cut off score for the total score is 76. In the present study 17% had high fatigue scores according to this cut off point. On the MBI-EE scale 19.4% scored high or very high, which indicates that fatigue is a severe health problem. In contrast, 8.9% scored very low and 21.5% low on the MBI-EE scale. The percentages for the other questionnaires are quite similar. Moreover, the response categories were all used. No answer was never mentioned. This indicates that the group of participants was not biased in one direction—that is, the non-responders are mainly non-fatigued individuals or mainly fatigued individuals.

With regard to the demographic variables age, gender, and educational level, hardly any associations were found with fatigue scores. Although this is in contradiction to the observation by Lewis and Wessely, who claimed that women reported two or three times more fatigue than men, it was in line with the results of De Rijk and colleagues. A possible explanation for this phenomenon might be that the studies cited by Lewis and Wessely often measured fatigue with only a single item or scale and/or a dichotomous response format. The finding that different age groups reported similar fatigue experiences might be explained by the healthy worker effect: the phenomenon that people who stay healthy are able to work until their retirement. 57

The present study has some limitations. For example, it was impossible to include all relevant fatigue questionnaires. Therefore, a selection of questionnaires had to be made. The six instruments that were chosen are reliable, valid, and frequently used in Western countries. To our knowledge, this selection of measures forms a good representation of the available fatigue instruments. The use of other measures might have led to different results. Furthermore, the number of questionnaires could influence the scores on the later completed items. However, the comparable percentages of severe fatigue for all measures do not support this idea.

In conclusion, all fatigue questionnaires used in the Fatigue at Work programme measure fatigue unidimensionally in a reliable and valid way. The FAS is the most promising questionnaire to measure fatigue in a working population.

ACKNOWLEDGEMENTS

The present study was supported by a grant from the Netherlands Organisation for Scientific Research (NWO). Research Programme “Fatigue at Work” grant no: S80-202-04, and by WORC, research institute of Tilburg University.

Authors’ affiliations

J De Vries, J H Michielsen, G I Von Heck, Department of Psychology and Health, Tilburg University, and Research Institute for Psychology and Health, Netherlands.

REFERENCES

6 Schaafeli W, Van Dierendonck D, Burnout, a begrip gemeten. De Nederlandse versie van de Maastricht BurnoutInventory (MBI-NL) [Burnout, the measurement of a concept]. The Dutch version of the Maastricht Burnout Inventory (MBI-NL). Gedrag Gezondheid 1994;22:153–72.


