Journal of Occupational and Organizational Psychology (1998), **71**, 331–351 *Printed in Great Britain*

Dimensionality and validity of the Burnout Measure

Dirk Enzmann* and **Wilmar B. Schaufeli**

Utrecht University, Department of Social and Organizational Psychology, PO Box 80.140, 3508 TC Utrecht, The Netherlands

Peter Janssen

Maastricht University, Department of Health Organisation, Policy and Economics, PO Box 616, 6200 MD Maastricht, The Netherlands

Alfred Rozeman

Versloot & Partners, Consultancy, Soesterdijkse weg 15, 3732 HC De Bilt, The Netherlands

The dimensionality and validity of the BM (Burnout Measure) is investigated in Dutch samples of human services professionals and white-collar workers (total $N = 2190$). Results show that, in contrast to the presumed dimensions 'physical exhaustion', 'emotional exhaustion' and 'mental exhaustion', the factors of the BM are 'demoralization', 'exhaustion', and 'loss of motive' The factorial structure is not affected by using different time frames although an unrestricted time frame results in significantly higher mean scores A BM version with a seven-point Likert scale results in more reliable subscales than a five-point scale The factor 'loss of motive' is substantially affected by inconsistent answering patterns of the respondents The three-factor model of the BM fits equally well in samples of professionals with and without patient contact The BM subscales correlate highly with fatigue and with the dimension 'emotional exhaustion' of the Maslach-Burnout-Inventory (MBI) but cannot be distinguished from psychological strain and psychosomatic complaints Problems in the conceptualization of burnout and its operationalization are discussed. It is concluded that the BM captures only a particular aspect of burnout and is rather a measure of general well-being Recommendations for further use of a modified BM are given

Together with the Maslach-Burnout-Inventory (MBI; Maslach, Jackson & Leiter, 1996), the Burnout Measure[1] (BM; Pines & Aronson, 1981) is the most frequently used instrument to assess burnout. Nowadays, a single particular definition of burnout predominates. This definition arose from the inductively developed MBI. Accordingly, burnout is conceived as a three-dimensional construct characterized by emotional exhaustion, depersonalization and feelings of reduced personal

*Requests for reprints
[1]In earlier publications termed 'Tedium Scale'

accomplishment. By the specific phrasing of the MBI items as well as by the theoretical view of its authors, burnout is limited to human services professions in which direct contact with recipients predominates. Behind this lies the theoretical assumption that burnout mainly results from the *emotional* demands of interacting with other people (Maslach, 1993). Thus, burnout is distinguished from classical stress reactions or related constructs such as psychological fatigue. The assumption, however, that demanding interactions with other people are the main cause of burnout could not empirically be verified (Enzmann, 1996; Schulz, Greenley & Brown, 1995). Therefore, alternative attempts of conceptualizing burnout are legitimate and make sense—for instance, a three-dimensional version of the MBI that can be applied to occupational fields other than 'people work' (Schaufeli & Leiter, 1996; Sonnentag, Brodbeck, Heinbokel & Stolte, 1994). A one-dimensional burnout measure, however, would facilitate the diagnosis greatly because a single score is easier to interpret and to communicate than a test profile.

In contrast to the MBI, the BM assesses the degree of burnout with one single score. Furthermore, according the phrasing of the BM items, as well as to the theoretical considerations of its authors, the BM is not restricted to a particular occupational field (i.e. 'people work'). Pines assumes that burnout can occur in any occupational field as well as in other groups such as students, couples, parents, political activists and housewives (Pines, 1988; Pines & Aronson, 1988). However, the broader applicability of the BM is associated with specific problems concerning item phrasing and discriminant validity. Furthermore, there are theoretical and empirical problems associated with the dimensionality of the burnout construct.

The items of the BM are not designed for certain occupational groups. As a consequence, their phrasing is so general that the distinctiveness of the construct which they are meant to measure is threatened. A closer look at the BM items shows that it is very difficult to discriminate them from indicators of similar constructs like fatigue or depression (see Appendix).

Pines & Aronson (1988) define burnout as 'a state of physical, emotional and mental exhaustion caused by long-term involvement in situations that are emotionally demanding' (p. 9). 'Physical exhaustion is characterized by low energy, chronic fatigue, and weakness ... Emotional exhaustion, the second component of burnout, involves primarily feelings of helplessness, hopelessness, and entrapment ... Mental exhaustion, the third component, is characterized by the development of negative attitudes toward one's self, work, and life itself' (p. 12 ff.). Despite this multidimensional definition, however, the three types of exhaustion themselves are not placed into a theoretical framework and a discussion of their differential development over time or their specific causes or consequences is lacking. For example, there is no explanation of *why* and *how* physical exhaustion, which is defined by such diverse symptoms as chronic fatigue, accident proneness, muscle tensions and increased susceptibility to illness, is caused by 'long-term involvement in situations that are emotionally demanding' (Pines & Aronson, 1988, p. 9).

Pines & Aronson compute a single burnout score by summing up the 21 items of the BM (after recoding positively phrased items), although they classify the items according to their definition of burnout into three types of exhaustion (Pines &

Aronson, 1981, p. 203): *physical exhaustion* (items 1, 4, 7, 10, 13, 16, 20); *emotional exhaustion* (items 2, 5, 8, 12, 14, 17, 21); and *mental exhaustion* (items 3, 6, 9, 11, 15, 18, 19) (see Appendix). To date, Pines and her co-workers have not published a study that confirms this factorial structure of the BM. Although the BM is supposed to tap three different aspects of exhaustion, the internal consistency of the whole scale is rather high, ranging from .91 to .93 (Pines & Aronson, 1981, p. 203 ff.). Accordingly, it seems likely that the BM is unidimensional. If, however, the BM should indeed comprise three different aspects of burnout, they can thus be assumed to be highly correlated. Problems in operationalization are also found with regard to single items: for instance, 'entrapment' is claimed to be a symptom of 'emotional exhaustion' (Pines & Aronson, 1981, p. 13), whereas 'feeling trapped' (item 11) is claimed to represent the dimension 'mental exhaustion' (p. 203). Or the item 'feeling burned out' (of the dimension 'emotional exhaustion') has been replaced in the most recent version of the BM by 'I can't take it anymore' without any theoretical or empirical arguments (Pines & Aronson, 1988).

Studies investigating the psychometric properties of the BM are rare. There are hints that the BM is either unidimensional or its factors are highly correlated, but to date there is no convincing empirical evidence of its three-factor structure (i.e. physical, emotional and mental exhaustion). Using a sample of 139 social workers, Corcoran (1986) found four highly correlated factors that showed *no* correspondence to the dimensionality proposed by Pines & Aronson (1988). Because the first factor accounted for 44 per cent of the variance, and because of an internal consistency of $\alpha = .93$ and highly correlated factors, Corcoran considered the BM to be unidimensional. Ray & Miller (1991) found two correlated factors which they termed 'exhaustion' (items 1, 4, 5, 7, 8, 10, 13, 20) and 'affect' (items 9, 11, 12, 14, 15, 18, 21). Somewhat similar results were obtained by Enzmann & Kleiber (1989) and Schaufeli & Van Dierendonck (1993). They found a pattern of three highly correlated factors that were described as *exhaustion* (items 1, 4, 5, 7, 8, 10), a combination of 'physical' and 'emotional exhaustion', *demoralization* (items 9, 11, 12, 13, 14, 16, 17, 18, 21) and *loss of motive* (items 2, 3, 6, 19, 20).

Enzmann & Kleiber (1989) used a rather small heterogeneous sample ($N = 128$) and employed exploratory factor analyses of a German version of the BM. They showed that indeed an oblique rotation accounts better for the data than an orthogonal rotation, with the first unrotated factor accounting for 44 per cent of variance. Using a rather homogeneous sample of 667 nurses and employing confirmatory factor analysis, Schaufeli & Van Dierendonck (1993) found an almost identical result for the Dutch version of the BM. Concerning the validity of the three dimensions it is striking that in the Dutch and in the German samples all four negatively phrased items (3, 6, 19, 20) constitute one single factor. Ray & Miller (1991) seem to have come to a similar result because items 3, 6 and 19 belong to the six items they dropped from the original BM. The distinction between positively and negatively phrased items could be interpreted in terms of negative and positive affect reported in studies on well-being scales (Bradburn, 1969; Diener, Larsen, Levine & Emmons, 1985; Warr, Barter & Brownbridge, 1983). However, it may just as well be explained by errors of respondents who did not recognize the reversed scoring of items 3, 6, 19 and 20. Thus, it is not clear whether the third

factor should be regarded as a genuine burnout dimension of the BM (i.e. loss of motive) or as a methodological artifact.

In recent years, several studies have been conducted in the Netherlands that employed different versions of the BM in different samples. Therefore, it is possible to investigate whether the psychometric properties of the BM fit to its conceptual basis. Specifically, this includes the following issues:

(1) Can the three-factor structure of the BM that has been found in earlier studies be replicated and generalized across different occupational groups with and without direct patient contact?
(2) Is the factorial structure of the BM stable across different response formats and time frames?
(3) If the four positively phrased items of the BM constitute one factor, should this factor be regarded as a genuine dimension of burnout (i.e. loss of motive) or does it reflect a methodological artifact?
(4) Can the discriminant and convergent validity of the BM be established? In other words, can burnout as measured by the BM be distinguished from 'classical' stress reactions and personality traits? And to what extent does it converge with other measures of burnout, most notably the MBI?
(5) Are the mean scores of the BM affected by such factors as the time frame used, the type of client contact, gender or nationality of the sample?

The second and fifth issues need some further explication: it might be expected that different response formats affect only the reliability of the scale, but not its factorial structure. Different time frames, however, are likely to result not only in different scale means but also in different loadings of those items whose meaning depend on the time frame used. For example, by using an unrestricted time frame, item 21 ('feeling anxious') indicates trait anxiety, whereas by using a restricted time frame the same item indicates state anxiety. Because burnout is not regarded to be a personality trait, a restricted time frame is more appropriate than an unrestricted one.

The fifth issue of mean differences can be addressed under the condition that the factorial structure of the BM can indeed be generalized across different versions and samples. Mean differences are not only affected by sample characteristics but also by the time frame used: it can be expected that a more restricted time frame results in lower mean scores than an unrestricted time frame, simply because the likelihood of symptoms increases as the time frame increases. After investigating this, professionals with direct client contact are compared to professionals without direct client contact with respect to their level of exhaustion: if it is true that burnout is a general phenomenon that is not restricted to helping professions, as claimed by Pines & Aronson (1988), no such differences would occur. Because in the samples studied women predominate in the group of helping professionals with direct patient contact, gender will be taken into account simultaneously. Analyses with the MBI have shown that with respect to emotional exhaustion gender differences are less important (Schaufeli & Van Dierendonck, 1994). With respect to the nationality of the samples, there are indications from studies employing the MBI that in European countries the mean scores of burnout are systematically

Table 1. Demographic data of composite samples

Composite sample	Patient contact (%) Yes	No	Unknown	Sex (%) Male	Female	Age x̄	Median	SD	*N*
A	9.2	90 5	0.3	63.5	36.5	37.6	37 0	7.9	902
B	52.3	45 6	2.1	15.5	84.5	41.7	42.0	10.4	608
C	87.2	0 0	12.8	28.1	71.9	33.2	32.0	7.3	680
Total	45.4	49 9	4.7	39.9	60.1	36.6	35.0	8.8	2190

Note Composite sample A white-collar workers from insurance companies (53 per cent), automation business (24 per cent) and bank offices (10 per cent), and professionals and workers from general hospitals (13 per cent) (the response rate (95 per cent) only one sample (*N*=213) is known), *composite sample B* nurses from outdoor services (55 per cent) and nursing homes (2 per cent), white-collar workers from insurance companies (26 per cent), and employees without patient contact (16 per cent) (the mean age is based on 378 cases only, response rates are unknown), *composite sample C* psychiatric nurses (26 per cent), community nurses (32 per cent), hospice nurses (17 per cent), general hospital nurses (15 per cent), and nurses working with mentally retarded patients (10 per cent) (response rates vary between 68 per cent to 89 per cent)

lower than in the United States (Enzmann, Schaufeli & Girault, 1995). Likewise, it should be expected, that the scores of the BM in the Dutch samples are systematically lower than in the numerous samples studied by Pines & Aronson (1981).

Method

Participants

In collapsing the data of 11 Dutch studies into three different composite samples three criteria have been used: (1) the versions of the BM that have been used, (2) the occupational groups, (3) the part of data that has already been used in a study of Schaufeli & Van Dierendonck (1993) The latter serves as a standard of comparison to validate the outcomes of our analyses.

Three types of composite samples are distinguished (see Table 1). the first group (A) includes four samples and consists predominantly of white-collar workers or human services professionals *without* direct patient contact, the second group (B) includes three heterogeneous samples in which a BM with a five-point response format instead of a seven-point format was used; the third group (C) includes four samples in which versions of the BM with different time frames had been used (i.e. with and without a restriction to the last month) This last group consists predominantly of nurses with *direct* patient contact and has previously been investigated by Schaufeli & Van Dierendonck (1993)

Measures

Different versions of a Dutch translation of the BM (Pines & Aronson, 1981) have been employed. In all subsamples of A and B, and two subsamples of C the time frame of the original version had been narrowed down from 'How often do you have any of the following experiences?' to 'How often did you have the following experiences during the last month?' In the two remaining samples of C, the instruction was 'How often did you have the following experiences?' In order to analyse the effects of different versions of the BM, composite sample C will be split up into two subsamples C.1 (*N* = 180, no time frame) and C 2 (*N* = 500, restricted time frame). In composite sample B the response format had been changed from the original seven-point Likert scale ('never', 'once in a great while', 'rarely', 'sometimes', 'often', 'usually', 'always') to a five-point scale ('never', 'rarely',

'sometimes', 'often', 'always') Additionally, in several subsamples some items had been phrased slightly different, e g 'feeling trapped', 'feeling caught', 'not knowing any way out' (item 11) Only with respect to the phrasing of item 15 ('feeling disappointed in others', 'feeling disillusioned and resentful', 'feeling disillusioned') marked differences were found in terms of factor loadings and item total correlations Because dropping this item improved the reliability of the total scale (as well as its subscale) irrespective of its phrasing, item 15 was eliminated from the following analyses. Enzmann & Kleiber (1989) reported a similar improvement of scale homogeneity when this item was dropped

As a convergent validity measure the Dutch version of the MBI (Schaufeli & Van Dierendonck, 1994) was used The MBI consists of three subscales 'emotional exhaustion' (9 items), 'reduced personal accomplishment' (8 items) and 'depersonalization' (5 items) The MBI was employed in one subsample of A and in three subsamples of C ($N = 625$) The scoring of its 22 items ranged from 0 ('never') to 6 ('every day') The internal consistency of the three MBI subscales (Cronbach's α) is sufficient (emotional exhaustion $\alpha = 88$, reduced personal accomplishment: $\alpha = 77$, depersonalization. $\alpha = 71$)

To discriminate burnout from familiar stress reactions, two subscales of the Dutch adaptation of the Work Stress Questionnaire (VOS-D, Bergers, Marcelissen & de Wolff, 1986) developed by Caplan, Cobb, French, Harrison & Pinneau (1975) have been used in three subsamples of C ($N = 593$). The first subscale, 'psychological strain' (14 items), includes anxiety, symptoms of depression and irritation, whereas the second subscale, 'somatic complaints' (11 items), includes psychosomatic symptoms such as sweating palms, upset stomach, sleeping problems, and heart beating faster than usual The scoring of both scales ranges from 1 ('never') to 4 ('very often'); their internal consistencies are satisfactory (psychological strain: $\alpha = .83$, somatic complaints: $\alpha = 82$)

As an additional discriminant validity measure a Dutch scale to assess subjective health (VOEG, Dirken, 1967) was used Seven items of this scale are identical to items of the somatic subscale of the John Hopkins Symptoms Checklist (HSCL, Derogatis, Lipman, Rickels, Uhlenhuth & Covi, 1974) It has been employed in composite sample B and two subsamples of A ($N = 1306$). The subscales of the VOEG are based on an *ad hoc* factor analysis conducted by the authors that resulted in three factors labelled 'fatiguability' (six items), 'cardiovascular problems' (five items), and 'gastrointestinal complaints' (five items). The VOEG is scored dichotomously (yes/no), which is probably one of the reasons for a rather low internal consistency of the second subscale (fatigue $\alpha = 87$, cardiovascular problems· $\alpha = 66$, gastrointestinal complaints $\alpha = .82$)

Because exhaustion is a dominant aspect of burnout, a Dutch scale to assess the quality of sleep (GSKS, Meijman, Vries-Griever, Vries & Kampman, 1988) has been used as a further discriminant validity measure in composite sample B and one subsample of A ($N = 812$) It covers a broad spectrum of sleep complaints (e.g. trouble to fall asleep, to sleep through, recuperation). The internal consistency of the scale with 14 dichotomously scored items is $\alpha = .86$

Finally, a measure to assess the personality trait 'reactivity' has been used to determine the discriminant validity of the BM. Reactivity has been measured by using a 22-item subscale of the Strelau Temperament Inventory (Strelau, 1983) in three subsamples of C ($N = 584$) Reactivity is defined as ' a basic dimension of temperament which determines the intensity of reaction to both internal and external stimuli. Specifically, a high reactive person exhibits a stronger reaction than a low reactive person to an objectively identical stimulus' (Strelau, 1983, p. 199). The reliability of this instrument is rather low ($\alpha = .68$).

Procedure

In a first step, based on a random subsample A.1 (one half of composite sample A) exploratory factor analysis in combination with reliability analysis of the BM is performed in order to arrive at a factor model that optimally accounts for the data In a second step, the resulting factor model (M_4) is one of the four models that is tested in the second half of composite sample A (A 2) by employing confirmatory factor analyses The other models are· a one-factor model (M_1), a model with three correlated factors (emotional, physical, and mental exhaustion) based on the assumptions of Pines & Aronson (1981) (M_2), and the three-factor model (exhaustion, demoralization, loss of motive), identified by Schaufeli & Van Dierendonck (1993) (M_3) The final model (M_5) of the BM is developed in the second step by further optimizing the model fit.

A third step investigates whether the final model fits equally well in the composite sample B, subsample C 1, and C 2 by employing multiple group analyses with the factor loadings constrained to be equal across the groups (cross-validation) The analyses show, whether different *sample characteristics* (in contrast to composite sample A, the majority of the respondents of composite sample C had direct client contact), *time frames* (in subsample C 1 the time frame was not restricted to the last month) or *response formats* (in composite sample B a five-point Likert scale has been employed) have a significant impact on the factorial structure of the BM

If the four positively phrased items of the BM constitute one factor, it is investigated in an additional fourth step whether this factor can be regarded as a genuine dimension of burnout (i e loss of motive), or as a mere methodological artifact. For this purpose, the data of different subsamples are combined. In the final step of the analyses the convergent and discriminant validity of the BM and its subscales are evaluated by means of factor analyses of scale scores of the BM together with the convergent and discriminant validity measures

Results

Step 1: Exploratory factor analyses

Based on a random subsample (A.1, $N = 447$) of composite sample A, an exploratory factor analysis was performed on the 20 items inventory (item 15 had been dropped, see above) using principal axis factoring and oblique rotation. According to several (corresponding) criteria (i.e. eigenvalue >1.0, explained variance >5.0, scree test, parallel analysis (Lautenschlager, 1989)) the optimal number of factors to be extracted proved to be three. The first unrotated factor explained 44.8 per cent of variance and all three factors together explained 58.5 per cent. This finding was consistently shown in other factor analytic studies with different translations of the BM (Corcoran, 1986; Enzmann & Kleiber, 1989, Schaufeli & Van Dierendonck, 1993).

Subsequent reliability analysis indicated that the homogeneity of the three subscales could not be improved further by eliminating particular items. Except for the loading of item 2 on the first factor and the loading of item 13 on the second factor, this factor solution replicates the findings of Enzmann & Kleiber (1989) (see Table 2). With the exception of the loading of item 13 this solution also replicates the findings of Schaufeli & Van Dierendonck (1993) in a sample of nurses who had direct patient contact. The three factors can thus be labelled in the same way as 'demoralization', 'exhaustion' and 'loss of motive', respectively. It is noteworthy that again factor 3 is exclusively defined by the positively phrased items of the BM. Besides, except for item 20, the second factor is identical to the factor 'exhaustion' identified by Ray & Miller (1991). The correlation matrix of the factor scores (Table 2) shows that the factor intercorrelations are rather high and that factor 3 (loss of motive) with exclusively positively phrased items correlates least with both other factors.[2] The reliability (Cronbach's α) of all three factors is high (demoralization: $\alpha = .91$, exhaustion: $\alpha = .91$, loss of motive: $\alpha = .87$).

None of the factors assumed by Pines & Aronson (1981) emerged. Items belonging to 'mental exhaustion' are spread across factors 1 and 3, items belonging to 'emotional exhaustion' are spread across factors 1 and 2, and items belonging to 'physical exhaustion' can be found in every factor. Items 5 ('being emotionally

[2]One of the two differences is highly significant $r_{(\text{factor 1, factor 2})}$ vs $r_{(\text{factor 1, factor 3})}$ $t = 1\,31$, $p = 190$, $r_{(\text{factor 1, factor 2})}$ vs $r_{(\text{factor 2, factor 3})}$ $t = 5\,35$, $p < 001$

Table 2. Factor pattern matrix and factor correlations of the BM (PAF, oblimin rotation; subsample A.1, $N=447$)

Item	Factor 1 (demoralization)	Factor 2 (exhaustion)	Factor 3 (loss of motive)
12 E	.78	− 06	.04
18 M	75	− 08	.01
17 E	.75	04	.01
16 P	.72	13	− .07
11 M	67	.09	.02
21 E	.64	.03	− 08
09 M	63	02	.23
02 E	.60	.14	14
14 E	50	.05	.12
07 P	− .01	.87	− .02
10 P	04	.83	.03
08 E	.10	.77	.05
01 P	− .03	.71	.04
04 P	− .02	.70	− .02
05 E	.22	.59	− .06
13 P	.27	.44	.22
19r M	.06	− .03	79
06r M	.11	− .12	.79
20r P	− .14	.28	.75
03r M	.09	02	.72
Factor correlations	Factor 1	Factor 2	Factor 3
Factor 1	1.00		
Factor 2	.61	1.00	
Factor 3	.56	.42	1.00

Note E=emotional exhaustion, P=physical exhaustion, M=mental exhaustion (lables according to Pines & Aronson, 1981), r=recoded

exhausted') and 8 ('feeling burned out') clearly load on factor 2 which comes most close to the physical aspect of exhaustion as assumed by Pines & Aronson. But the physical aspect can obviously not be distinguished from the emotional aspect. If burnout (for which item 8 is supposed to be the best indicator[3]) is identical to 'physical exhaustion', the concept of burnout as an *emotional* reaction to *emotionally* demanding interactions with other persons is questionable. At this point of the investigation, we neither have sufficient evidence to conclude that burnout is poorly operationalized by the BM nor that burnout has to be conceptualized differently. However, we have to recognize that in the second factor physical *and* emotional aspects of exhaustion converge and that it is dominated by feelings of fatigue. Therefore, we have labelled that factor simply 'exhaustion'.

[3]In a modified version of the BM Pines & Aronson (1988) have replaced this item by 'Can't take it anymore'

Table 3*a*. Confirmatory factor analysis of different factorial models of the BM (calibration subsample A.1, $N=447$)

Model	χ^2	d.f.	GFI	AGFI	CFI	StRMR	Modification
M_1	1698.93	170	637	.551	.726	067	—
	1461.50	151	.654	.565	.746	.066	E20–E19, ¬ V13
M_2	1442.14	167	.694	.616	.771	.067	—
	1148 65	148	745	672	.806	.070	E20–E19, ¬ V13
M_3	624.24	167	.866	.832	918	.036	—
M_4	626 80	167	.874	.842	.918	.040	—
(M_5)	459.99	148	.896	.866	.940	.033	E20–E19, ¬ V13

Note M_1 = one-factor model, M_2 = three-factor model according to Pines, M_3 = three-factor model according to Schaufeli & Van Dierendonck, M_4 = three-factor model based on exploratory factor analysis of subsample A 1, M_5 = modified models M_3 and M_4 with E20–E19 = covariance of errors of items 20 and 19, and ¬ V13 = without item 13, all χ^2 are significant beyond 001

Table 3*b*. Confirmatory factor analysis of different factorial models of the BM (subsample A.2, $N=455$)

Model	χ^2	d f	GFI	AGFI	CFI	StRMR	Modification
M_1	1644.92	170	.653	.571	.757	.058	—
	1417 08	151	674	590	.776	.056	E20–E19, ¬ V13
M_2	1414.81	167	.703	627	.794	.059	—
	1180.27	148	.732	.657	.817	.059	E20–E19, ¬ V13
M_3	627.58	167	871	838	.924	.032	—
M_4	570.06	167	.884	.854	.934	.031	—
(M_5)	445 36	148	900	.872	.947	.029	E20–E19, ¬ V13

Note M_1 = one-factor model, M_2 = three-factor model according to Pines, M_3 = three-factor model according to Schaufeli & Van Dierendonck, M_4 = three-factor model based on exploratory factor analysis of subsample A 1, M_5 = modified models M_3 and M_4 with E20–E19 = covariance of errors of items 20 and 19, and ¬ V13 = without item 13, all χ^2 are significant beyond 001

Step 2: Confirmatory factor analyses

Confirmatory factor analyses employing structural equation modelling[4] have been used to test which of the previously described four models accounts best for the covariance of the BM items. The fit of models M_2, M_3, and M_4 (see notes to Tables 3*a* and 3*b*) can be compared to the fit of model M_1 by employing a χ^2 difference test. This test shows whether a factor model with three correlated factors fits better to the data than a one-factor model. However, the significance of the approximate χ^2 values is strongly affected by sample size. Especially in large samples small deviations will turn out to be significant, whereas in small samples

[4] All structural equation models were tested by using EQS/Windows version 4 02 (Bentler & Wu, 1993)

even mis-specified models might seem to fit to the data. Therefore, and to allow for a comparison of model fit of samples of different size, instead of the χ^2 values the Comparative Fit Index (CFI) (see Gerbing & Anderson, 1993; Tanaka, 1993) is used to evaluate the models The CFI belongs to a set of fit indices that are least affected by sample size. Generally, it is assumed that the CFI of a good-fitting model should be greater than .90. In order to allow for comparisons of other model tests published, the often-used Goodness of Fit Index (GFI) and Adjusted Goodness of Fit Index (AGFI) (Jöreskog & Sörbom, 1989), as well as the standardized root mean square residual (StRMR), are shown.

The model tests have been applied to the calibration subsample A.1 as well as to the seond half of composite sample A (A.2). The latter serves as a cross-validation of the results of the first. As shown in Tables 3*a* and 3*b*, a model with three correlated factors (i.e. 'demoralization', 'exhaustion', and 'loss of motive') clearly fits better to the data than either the one-factor model or the model assumed by Pines & Aronson (1981). Table 3*b* shows that the results of the exploratory factor analysis are replicated in the second half of composite sample A. In all models the fit could be significantly improved by allowing correlated errors of items 20 and 19. The correlated errors are due to common causes influencing the responses to these items that are not accounted for by the latent variable(s) of the model. Whereas models M_3 and M_4 fit equally well in subsample A.1 (Table 3*a*), the fit indices of subsample A.2 (Table 3*b*) show that assuming item 13 ('being weary') to load on the factor 'exhaustion' results in a better model fit. However, in both samples the fit could be improved significantly by allowing item 13 to load simultaneously on the factors 'demoralization' and 'exhaustion'.

Because a simple structure is preferred, item 13 was finally dropped in model M_5 that accounts for correlated errors of items 20 and 19. This further improved the model fit and eliminated the differences between the models M_3 and M_4. It thus seems to be correct to assume that a factor model of the BM with three correlated factors 'demoralization', 'exhaustion', and 'loss of motive' is equally appropriate in a sample of professionals with direct patient contact (composite sample C; see Schaufeli & Van Dierendonck, 1993) as in a sample without direct patient contact (composite sample A).

Step 3: Multiple group analyses

The first research question can be tested directly by employing multiple group analyses by which the loadings and the factor correlations of model M_5 are constrained to be equal across groups. Additionally, by pairwise comparing composite samples A (predominantly professionals without patient contact, restricted time frame, seven-point scale), B (heterogeneous sample, restricted time frame, five-point scale), subsamples C.1 (predominantly nurses with patient contact, no time frame, seven-point scale), and C.2 (predominantly nurses with patient contact, restricted time frame, seven-point scale), it can additionally be shown whether the time frame or the response format used changes the factorial structure of the BM.

Table 4. Pairwise multiple group analyses of model M_5 with factor loadings and factor correlations constrained to be equal

Composite sample		A	B	C.1
B	CFI	928	—	—
	$\Delta\chi^2$	**65.79**	—	—
	p	<.001	—	—
	Items	4, *6*, 17, *19*	—	—
C.1	CFI	.941	903	—
	$\Delta\chi^2$	**31.45**	**32.67**	—
	p	.087	.067	—
	Items	21	2, *19*, *20*, 21	—
C.2	CFI	.941	.917	929
	$\Delta\chi^2$	**28.56**	**29.57**	**22.54**
	p	.159	.129	428
	Items	4	17, *19*	21

Note Items whose loadings differ significantly across groups are indicated in the rows 'Items', d f of $\Delta\chi^2$=22, CFI of composite sample A= 946, composite sample B= 902, subsample C.1 = 915, subsample C 2= 933

Table 4 shows the CFIs of the six multiple group analyses together with the change in χ^2 by releasing all equality constraints.[5] Virtually *no* influence of the sample characteristic 'direct patient contact' was found (the multiple group analysis of composite sample A and subsample C.2 results in a high CFI and a rather low $\Delta\chi^2$; only the loadings of item 4 differ significantly). This means that the factor structure of the BM is generalizable across the different samples studied and that the factors of the BM have the same meaning for persons with and without patient contact.

The largest effect can be found with respect to different response formats. Multiple group analyses with composite sample B (in which a five-point Likert scale has been used) result in the lowest CFIs and highest $\Delta\chi^2$s. Separate analyses of the four samples show the poorest fit with respect to composite sample B (CFI = .902 opposed to 946, .915, and .933 in composite sample A, subsamples C.1, and C.2, respectively). Because in composite sample B the pattern of item loadings does not change, even though the explained variance is less, a shorter response format obviously results in less reliable subscale scores. Consequently, results of studies using BM versions with different response formats cannot be compared. Generally, a seven-point Likert scale results in more reliable subscale scores of the BM than a five-point scale.

Concerning composite sample B, another result is noteworthy. A remarkable proportion of items that differ significantly when composite sample B is compared with the other samples is positively phrased and belongs to the factor 'loss of motive' (items 6, 19 and 20) If this factor (partially) reflects errors of those respondents, who

[5]Because the variances of the items in sample B are systematically smaller due to a more restricted response format, the analyses were based on correlation matrices instead of covariance matrices

did not recognize the reversed direction of items 3, 6, 19 and 20, this error should be strongest with respect to item 19 that follows a series of 12 negatively phrased items. At the same time, it should be expected that such an error has a greater effect if a broader response format is used with a greater number of anchors This reasoning is supported by the fact that the loadings of item 19 always differ significantly when composite sample B is compared with the other samples. Additionally, by releasing the equality constraint of item 19 its loading on the factor 'loss of motive' is *always* smaller in composite sample B. This is a first indication that the factor 'loss of motive' seems to reflect errors of the respondents caused by its reversed scoring.

With respect to the time frame used, *no* systematic differences concerning the factorial structure of the BM can be found when subsample C.1 is compared with the other samples, except for a consistently lower loading of item 21 in subsample C.1. A cautious *post hoc* explanation for the finding with respect to item 21 ('feeling anxious') could be that, by dropping the restriction of the time frame, a trait rather than a state influences the answers. From the literature on the measurement of anxiety it is known that the difference of trait and state is crucial (Spielberger, 1972)

Step 4: Response inconsistencies and the factor 'loss of motive'

To investigate whether the factor 'loss of motive', which exclusively consists of the positively phrased items of the BM, could be explained by careless responses in the sense that some respondents did not recognize the reversed scoring of items 3, 6, 19 and 20, first the answers to three pairs of opposite items were analysed. By comparing item 6 ('being happy') to item 9 ('being unhappy'), item 17 ('feeling hopeless') to item 19 ('feeling optimistic'), and item 16 ('feeling weak') to item 20 ('feeling energetic'), those respondents could be identified who answered inconsistently. For this purpose the composite samples A, subsample C.1 and C.2 were combined ($N = 1582$). Two answers were regarded to be inconsistent, if a person answered with a value less than the mid-point 4 on one item of the pair and a value greater than the mid-point 4 on the other (after recoding all positively phrased items). Concerning the item pairs 6 and 9, 3.2 per cent of the respondents gave inconsistent answers (1.7 per cent having a difference of more than two points); these figures were 5.0 per cent (3.4 per cent) for the item pairs 17 and 19, and 5.0 per cent (2.8 per cent) for the item pairs 16 and 20, respectively. The most inconsistent answers were found for the item pair including item 19, followed by the pair including item 20. Both items appear after a long sequence of 12 negatively phrased items, thereby making an error due to habituation most likely. This can also be shown by comparing the means of the absolute difference of the three item pairs for those persons who answered *any* of them inconsistently ($N = 149$). The means of the first and second pair were 1.00 and 1.72 ($t(148) = -3.52, p < .001$), the mean of the third pair was 1.58 (first vs. third pair: $t(148) = -2.83, p = .005$; second vs. third pair: $t(148) = .68, p = .497$). Comparing the mean of the total BM score (after eliminating items 13 and 15 according to model M_5) of the 149 persons answering inconsistently to the mean of the remaining sample ($N = 1433$), it turns out that

Table 5. Effects of inconsistently answering persons on the factorial structure of the BM

Composite sample	Model	χ^2	d.f.	CFI	StRMR
A+C (N=1582)	M_6	2434.93	150	.881	060
	M_5	1208.34	148	.945	029
		$\Delta\chi^2_{(2)}=1226.59$, $p<.001$			
A+C without inconsistently answering persons (N=1433)	M_6	1838.54	150	905	.058
	M_5	1182.87	148	.942	.034
		$\Delta\chi^2_{(2)}=655.66$, $p<.001$			

Note M_6=model with two correlated factors (items 3, 6, 19, 20 load on 'demoralization'), M_5=model with three correlated factors (items 3, 6, 19, 20 represent the factor 'loss of motive')

inconsistently answering persons seem to have a significant *higher* mean score (3.28 vs. 2.72, $t(1580) = 7.91$, $p<.001$). It is questionable whether this difference should be interpreted as a true difference in the level of burnout, because the inconsistently answering and apparently more burned-out persons at the same time seem to be rather happy, optimistic and energetic. The validity of the positively phrased items with respect to burnout in the group of inconsistently answering persons is also dubious because an analysis of the internal consistency of the total scale reveals that in this group the corrected item total correlations of items 3, 6, 19 and 20 are .26, .15, − .00 and .08, respectively ($N = 149$), as opposed to the corrected item total correlations in the remaining sample (.61, .59, .60, .64; $N = 1433$).

The issue can be investigated further by comparing a three-factor model of the BM (M_5) to a two-factor model in which items 3, 6, 19 and 20 load on the factor 'demoralization' (M_6)—first in the total sample (A + C, $N = 1582$) and then after eliminating the 149 inconsistently answering persons. The degree to which the difference of both models disappears after the elimination of these persons is an indicator of the impact of careless, or at least inconsistent, responding and thus may help to interpret the factor 'loss of motive'. Table 5 shows that inconsistent or careless responding is at least partly responsible for the existence of the third factor. If the most inconsistently answering persons (less than 10 per cent of the total sample) are excluded from the analysis, the difference of model fit between the three- and two-factor models drops from $\chi^2 = 1226.59$ to nearly the half ($\chi^2 = 655.66$). At the same time, the correlations of the factor 'loss of motive' with 'demoralization' and 'exhaustion' in the three-factor model increase from .61 to .74 and from .52 to .63, respectively. This indicates that the factor 'loss of motive' becomes more similar to the other factors if inconsistently answering persons are removed. Thus, although the third factor, 'loss of motive', cannot completely be explained by inconsistent answers with respect to the positively phrased items because it does not disappear entirely after eliminating inconsistently answering persons, such answer patterns play a substantial role in explaining the existence of the third factor.

Table 6. Correlation matrix of BM subscales and convergent and discriminant validity measures ($N=486$)

	BM–DM	BM–EX	BM–LM	MBI–EE	MBI–DP	MBI–PA	VOS–L	VOS–P
BM–EX	.67	—						
BM–LM	60	.59	—					
MBI–EE	63	.77	.60	—				
MBI–DP	42	39	.40	.58	—			
MBI–PA	− 30	− .32	− 58	− 43	− .44	—		
VOS–L	56	.56	40	.55	.32	− .24	—	
VOS–P	70	.66	65	.67	41	− 41	51	—
React	− .40	− .33	− .36	− 35	− .19	.32	− .33	− 38

Note BM–DM = demoralization, BM–EX = exhaustion, BM–LM = loss of motive, MBI–EE = emotional exhaustion, MBI–DP = depersonalization, MBI–PA = personal accomplishment, VOS–L = psychsomatic symptoms, VOS–P = psychological strain, React = reactivity, $p < 001$, two-tailed.

Table 7. Correlation matrix of BM subscales and discriminant validity measures ($N=769$)

	BM–DM	BM–EX	BM–LM	VOEG–F	VOEG–G	VOEG–C
BM–EX	.66	—				
BM–LM	.54	.49	—			
VOEG–F	.40	.60	39	—		
VOEG–G	.22	.22	.17	.30	—	
VOEG–C	.34	.35	.26	.41	.33	—
GSKS	.33	.36	29	.48	.22	.34

Note BM–DM = demoralization, BM–EX = exhaustion, BM–LM = loss of motive, VOEG–F = fatiguability, VOEG–G = gastrointestinal complaints, VOEG–C = cardiovascular problems, GSKS = quality of sleep, $p < 001$, two-tailed

Step 5: Convergent and discriminant validity

Because in analysing the convergent and discriminant validity of the BM a multivariate approach is preferred, the samples in which the MBI, the VOS and reactivity have been measured were combined into one composite sample ($N = 486$). Similarly, a combined sample in which the VOEG and the GSKS were used was built ($N = 769$). Because in the latter samples five-point and seven-point scales of the BM have been employed, before combining these studies the BM scores were converted to z scores. The correlation matrices of both samples are shown in Tables 6 and 7, respectively.

An exploratory factor analysis (principal components) of the (subscale) scores of the BM, MBI, VOS and the reactivity measure resulted in a one-factor solution; the first unrotated factor explains 54.5 per cent of variance. Although this indicates a good convergent validity of the BM and the MBI, it shows on the other hand that

the discriminant validity of the BM with respect to other stress reactions is rather poor, but this holds equally well for the three subscales of the MBI. If one forces a two-factor solution explaining 65.3 per cent of variance, the MBI subscales 'personal accomplishment' and 'depersonalization' dominate the second factor (a result also obtained by Schaufeli & Van Dierendonck, 1993), whereas the BM subscale 'loss of motive' loads equally high on factors 1 ($\lambda = .69$) and 2 ($\lambda = .73$). The loading of the reactivity score on factor 1 is *comparatively* low ($\lambda = -.49$). Table 6 also shows that the correlations of reactivity with the BM subscales are moderate, indicating at least some discriminant validity of the BM with respect to that particular personality trait. It is noteworthy that the MBI subscale 'personal accomplishment' correlates highest with the BM factor 'loss of motive', which could partly be due to the fact that both subscales comprise exclusively the positively phrased items. The speculation that inconsistent answer patterns may also affect the subscale 'personal accomplishment' is supported by the fact that in the subsample of inconsistently responding persons ($N = 20$) 'personal accomplishment' is the only variable that correlates significantly with 'loss of motive' ($r = -.63, p = .003$).

With respect to the discriminant validity of the BM compared to the VOEG subscales 'fatiguability', 'gastrointestinal complaints' and 'cardiovascular problems', and the scale 'quality of sleep' (GSKS), a similar result was found. Although an exploratory principal components analysis shows that there are two eigenvalues greater than 1.0, according to parallel analysis criteria (Lautenschlager, 1989) only one 'significant' component exists which explains 46.7 per cent of variance. This analysis also shows the poor discriminant validity of the BM with respect to fatigue and psychosomatic complaints. If one forces again a two-factor solution (explaining 62.2 per cent of variance), the second factor is defined by 'gastrointestinal complaints' ($\lambda = .77$), 'cardiovascular problems' ($\lambda = .69$), and 'quality of sleep' ($\lambda = .50$). The VOEG subscale 'fatiguability' loads about equally high on factor 1 ($\lambda = .48$) (dominated by the three BM subscales) and factor 2 ($\lambda = .45$). One could regard the latter result as a positive validation of the BM. The rather high correlation of 'fatiguability' with 'exhaustion' ($r = 60$, see Table 7) particularly supports the view that general fatigue is a prominent characteristic of burnout. On the other hand, this makes the assumption questionable that burnout results primarily from emotional demands imposed by demanding interactions with other people (Maslach, 1993; Pines & Aronson, 1988). It could also result from physical and other enduring but non-specific demands that exceed the general adaptability of a person and can thus appear in a wide variety of professions and situations.

Comparison of BM scores between groups

First the question is investigated whether a more restricted time frame results in lower mean scores than an unrestricted time frame. A comparison of the means of subsamples C.1 (unrestricted time frame, $N = 180$) and C.2 (time frame one month, $N = 500$) showed that the mean score is indeed affected by the time frame used. As expected, scores of the unrestricted time frame are significantly *higher* compared to

the one-month time frame. This effect was most pronounced with respect to the subscale 'demoralization' ($\Delta\bar{x} = .73$, $t(355\,4) = 10.58$, $p < .001$), followed by 'exhaustion' ($\Delta\bar{x} = .41$, $t(364.6) = 4.89$, $p < .001$) and 'loss of motive' ($\Delta\bar{x} = .17$, $t(678) = 2.94$, $p = .003$), respectively. One possible interpretation for the different effects regarding the separate subscales of the BM could be the frequency by which the different aspects of burnout occur. Whereas 'demoralization' is less likely to occur within a month, 'loss of motive' (and an inconsistent answer pattern, see above) appears more often. Because of these significant mean differences, in further analyses only data of those samples are investigated in which the restricted time frame of one month has been used.

In comparing different groups (male/female, with/without patient contact) only the scores of the final version of the BM were taken into account. According to the previous analyses, the final version consists of the two subscales 'demoralization' (items 2, 9, 11, 12, 14, 16, 17, 18, 21) and 'exhaustion' (items 1, 4, 5, 7, 8, 10), answered on a seven-point Likert scale with a time frame of one month. Accordingly, only those groups in which this version had been used were analysed (A + C.2, $N = 1405$). The psychometric characteristics of the final BM and its subscales were $\bar{x} = 2.33$, SD = .92, $\alpha = .91$ (demoralization), $\bar{x} = 3.00$, SD = 1.12, $\alpha = .90$ (exhaustion), and $\bar{x} = 2.67$, SD = .93, $\alpha = .93$ (BM total, 15 items); the correlation between 'demoralization' and 'exhaustion' was $r = .67$ ($p < .001$)

An analysis of variance with the factors 'gender' and 'patient contact' including their interaction term showed *no* significant group differences, neither with respect to the *total score* (gender: $F(1,1290) = .11$, $p = .742$; patient contact: $F(1,1290) = .38$, $p = .538$; interaction gender–patient contact: $F(1,1290) = .88$, $p = .349$), nor for the separate scores of *demoralization* (gender: $F(1,1290) = 2.77$, $p = .096$; patient contact: $F(1,1290) = 1.82$, $p = .177$; interaction gender–patient contact: $F(1,1290) = .64$, $p = .425$) or *exhaustion* (gender: $F(1,1290) = 3.63$, $p = .057$; patient contact: $F(1,1290) = .01$, $p = .936$; interaction gender–patient contact: $F(1,1290) = 1.44$, $p = .230$). This result remains unchanged even if the covariate age is taken into account.

Finally, the mean of the total BM score was compared to the mean score obtained in numerous studies of Pines & Aronson (1981). To make a comparison possible, only the mean score of subsample C.1 ($N = 180$) was compared to these data because Pines & Aronson did use an unrestricted time frame. Furthermore, all items of the original BM were used for this comparison. It is important to note that the data published in Pines & Aronson (1981) are based on participants of burnout workshops as well as on other samples. Because a secondary analysis revealed that workshop participants showed significantly higher burnout scores ($\Delta\bar{x} = .29$, $t(3914) = 10.63$, $p < .001$), only the non-participants were used for our comparison. Similar to the Dutch sample the American samples consisted of health-care professionals (nurses, caregivers for the handicapped; $N = 881$, groups 10, 18 and 24 in Pines & Aronson, 1981, pp. 204–206). As a result a slightly but not significantly higher mean score of the Dutch opposed to the American samples was observed ($\Delta\bar{x} = .12$, $t(302.7) = 1.95$, $p = .053$).

Discussion

Problems in theory and operationalization and validity of the BM

In this study the factorial structure of the BM that has previously been found in German (Enzmann & Kleiber, 1989) and Dutch samples (Schaufeli & Van Dierendonck, 1993), was replicated in samples with different characteristics (with and without patient contact). Instead of the hypothesized factors ('emotional exhaustion', 'physical exhaustion' and 'mental exhaustion'), three other factors ('demoralization', 'exhaustion' and 'loss of motive') were consistently found. The result is in line with findings from other factor-analytic studies of the BM. Corcoran (1986) also found a factor structure that did not correspond to the dimensions proposed by Pines *et al.* (1981). Moreover, Ray & Miller (1991) arrived at a two-factor solution in which the factors 'exhaustion' and 'affect' are similar to the factors 'exhaustion' and 'demoralization' of the present investigation. These results do not only question the operationalization of the burnout construct. They also cast doubt on theoretical assumptions underlying the BM. For instance, 'physical exhaustion' includes aspects of psychological exhaustion such as chronic fatigue (Pines & Aronson, 1988, p. 13). It would be theoretically more sound if physical and psychological forms of exhaustion were distinguished and if their relationship were clearly defined. In that case, chronic fatigue as well as emotional exhaustion would probably be part of the category 'psychological exhaustion'. Furthermore, the category 'mental exhaustion' includes negative changes of attitudes towards other people, work, oneself and life in general. It is not clear why this should be called 'exhaustion'. This theoretical indistinctiveness is reflected by the operationalization. First, the items assigned to the aspect of mental exhaustion do not tap negative attitudes towards other people and towards work. For instance, the MBI dimensions 'depersonalization' (negative attitudes towards other people) and 'reduced personal accomplishment' (negative attitudes towards oneself) operationalize these attitudinal aspects of burnout more adequately. Second, all items of the BM cover more or less non-specific negative feelings or thoughts about life in general, and thus tend to aim at general well-being. This is possibly the reason why the three aspects of emotional, physical and mental exhaustion are not retrieved in factor analyses, and why the factors actually found are highly correlated. The high correlations of factors justifies the calculation of one single burnout score—although, theoretically speaking, the items should be assigned to different factors. The advantage of the approximate one-dimensionality of the BM is therefore based on a twofold weakness: theoretical indistinctiveness and imperfect operationalization.

The investigation of the validity of the two main dimensions 'exhaustion' and 'demoralization' showed that the scores of the BM highly correlate with fatigue and with 'emotional exhaustion' of the MBI which also represents a rather general and non-specific aspect of burnout. Therefore, the convergent validity of the BM can be judged positively. Its discriminant validity, however, is unsatisfactory: exhaustion and demoralization can hardly be differentiated from psychosomatic complaints. This result puts the specificity of burnout as measured by the BM into question; it is probably a consequence of the theoretical indistinctiveness of

the conceptualization in which even psychosomatic complaints belong to the symptoms of burnout. Admittedly, the lack of distinction from other constructs is also true for the emotional exhaustion scale of the MBI; however, at least conceptually Maslach & Jackson (Maslach, Jackson & Leiter, 1996) do not conceive psychosomatic complaints as part of their burnout construct.

Consequences for further use of the BM in research and practice

It was pointed out at the beginning of this paper that it would simplify the diagnosis of burnout to have a one-dimensional burnout instrument which were applicable to all kinds of occupational groups and not only to helping professions With regard to one-dimensionality, at first glance the analysis of the BM seems to have produced a positive result. Although there are two dimensions in the modified version of the BM, the high correlation between 'exhaustion' and 'demoralization' justifies to compute a total score from both scales. The internal consistency of the total score based on 15 items ($\alpha = .93$) is excellent. However, the BM does not capture burnout as conceptualized by Pines & Aronson: the aspects 'dehumanization', 'distancing from other people', and 'development of negative attitudes towards others and oneself' are lacking. Only one aspect of burnout (exhaustion) is clearly operationalized by the BM Therefore, the approximate one-dimensionality has been purchased at the expense of a reduction in content. With regard to the applicability of the BM to other than helping professions, at first glance the results can be judged favourably. But again, there is a problem. While there are *no* differences between the factorial structure of the BM in different occupational groups, a comparison to the MBI shows that the BM is rather insensitive to situational differences: although in Dutch samples distinctly lower MBI burnout values of the dimension 'emotional exhaustion' have repeatedly been found than in comparable American samples (Schaufeli & Van Dierendonck, 1994, 1995), the Dutch BM values in the current study do not differ from the scores of similar American samples. This result indicates a serious problem. An instrument may lose its sensitivity for group differences if it is modified in such a way that its applicability becomes broader. Therefore, one may be anxious to know if the MBI–GS (MBI–General Survey; Schaufeli & Leiter, 1996; see below) will succeed in providing an alternative for the MBI for non-helping professions meeting this problem.

Obviously, the psychometric properties of the BM do not fit to its conceptual basis For practical use this implies that the BM is not suitable to measure burnout as a distinct phenomenon. Although the BM is not a diagnostic device that allows to differentiate between physical, emotional and mental exhaustion, it might be useful for the assessment of general deterioration of well-being. In case one wants to use the BM to assess exhaustion, demoralization or general well-being, three recommendations can be drawn from the present study:

(1) The seven-point response format of the BM should not be replaced by a five-point response format, because the shorter response format clearly impairs the reliability of the instrument.

(2) The time frame given in the original instruction should be reduced (e.g. to the last month). Whereas the factorial structure is scarcely impaired by an unrestricted time frame (with the exception of the loading of the item 'feeling anxious'), it leads to significantly higher mean scores. Especially when the BM is to be used in longitudinal studies a restricted time frame should be preferred.
(3) Items 13 and 15 should be eliminated. When calculating burnout scores, one should drop all the positively formulated items (originally 3, 6, 19 and 20) which constitute a separate factor. This factor is partly an artifact because of inconsistent answers of those persons who do not recognize the reversed direction of the items. Rephrasing the items does not make sense because three of them are already included in the item list in a negative formulation. If the items remain in the questionnaire—with the advantage to inhibit the tendency of blind and careless responding—a comparison of the answers of the item pairs 6–9, 17–19 and 16–20[6] can help to recognize inattentive and unreliably answering persons in order to exclude them from further analyses.

Future directions of burnout measurement

To meet the problems related to the multidimensionality and the restricted applicability of the MBI, the BM is not a good alternative. The BM seems to be less sensitive for group differences, and it does not capture the theoretically expected and crucial aspects of burnout. Two other strategies seem to be more promising. On the one hand, the MBI–GS (MBI–General Survey; Schaufeli & Leiter, 1996) can be used, which was developed recently on the basis of the MBI especially for non-helping professions. On the other hand, one should try to operationalize the aspects of burnout regarded as crucial (emotional exhaustion, negative attitudes towards the object of one's work, loss of motive and doubts about personal accomplishment) according to theory by means of deduction. All burnout instruments presently available have essentially been developed inductively, with the exception of the MBI–GS. The MBI–GS has been developed deductively, but the most characteristic (although up to now also most indistinct) burnout dimension '*emotional* exhaustion' has been replaced by a more general dimension of 'fatigue'. Therefore, the task to operationalize the 'classical' burnout construct according to theory remains to be done. However, it seems to be quite impossible to achieve a one-dimensional burnout measure on the basis of a multidimensional burnout construct.

References

Bentler, P M & Wu, E (1993) *EQS/Windows Version 4 02*. Los Angeles, CA: BMDP Statistical Software

Bergers, G P A, Marcelissen, F H G. & de Wolff, C J. (1986) *Handleiding Vragenlijst Organisatiestress* (Manual Organizational Stress Questionnaire). Nijmegen: University of Nijmegen, Department of Psychology

Bradburn, N M (1969) *The Structure of Psychological Well-being* Chicago, IL. Aldine

[6]Here the original item numbers are used, after elimination of items 13 and 15 the pairs are 6–9, 16–18 and 15–19

Caplan, R. D , Cobb, S., French, J R. P , Harrison, R & Pinneau, S R (1975) *Job Demands and Worker Health* Ann Arbor, IL. Institute for Social Research

Corcoran, K J (1986) Measuring burnout A reliability and convergent validity study *Journal of Social Behavior and Personality*, **1**(1), 107–112

Derogatis, L R , Lipman, R S., Rickels, K., Uhlenhuth, E H & Covi, L. (1974) The Hopkins Symptom Checklist (HSCL) A self-report symptom inventory *Behavioural Science*, **19**, 1–15.

Diener, E , Larsen, R J , Levine, S. & Emmons, R A (1985). Intensity and frequency: Dimensions underlying positive and negative affect *Journal of Personality and Social Psychology*, **48**, 1253–1265

Dirken, J. M (1967). *Arbeid en stress Het vaststellen van aanpassingsproblemen in werksituaties* (Work and Stress The Assessment of Adaptational Problems in Work Situations) Groningen Wolters

Enzmann, D (1996) *Gestreßt, erschopft oder ausgebrannt? Einflusse von Arbeitssituation, Empathie und Coping auf den Burnoutprozeß* (Stressed, Burned Out, or Exhausted? The Impact of Working Conditions, Empathy and Coping on the Burnout Process) Munchen Profil

Enzmann, D & Kleiber, D (1989) *Helfer-Leiden: Streß und Burnout in psychosozialen Berufen* (Helpers' Ordeals· Stress and Burnout in Psycho-social Professions). Heidelberg: Asanger

Enzmann, D , Schaufeli, W. B & Girault, N (1995) The validity of the Maslach Burnout Inventory in three national samples. In L. Bennett, D Miller & M W Ross (Eds), *Health Workers and AIDS Research, Intervention and Current Issues in Burnout and Response*, pp. 131–150 Chur, Switzerland: Harwood.

Gerbing, D W. & Anderson, J C (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K A Bollen & J. S. Long (Eds), *Testing Structural Equation Models*, pp 40–65. Newbury Park, CA: Sage.

Joreskog, K. G & Sorbom, D. (1989) *LISREL 7 A Guide to the Program and Applications*, 2nd ed Chicago, IL. SPSS.

Lautenschlager, G. J (1989) A comparison of alternatives to conducting Monte Carlo analysis for determining parallel analysis criteria. *Multivariate Behavioral Research*, **24**, 365–395.

Maslach, C. (1993). Burnout A multidimensional perspective In W. B. Schaufeli, C Maslach & T Marek (Eds), *Professional Burnout: Recent Developments in Theory and Research*, pp. 19–32 Washington, DC Taylor & Francis.

Maslach, C., Jackson, S. E & Leiter, M. P. (1996). *Maslach Burnout Inventory: Manual*, 3rd ed. Palo Alto, CA: Consulting Psychologists Press

Meijman, T. F., Vries-Griever, A. H G de, Vries, G. de & Kampman, R. (1988). *The Evaluation of the Groningen Sleep Quality Scale.* Groningen: University of Groningen, Department for Experimental and Occupational Psychology.

Pines, A M. (1988). *Keeping the Spark Alive: Preventing Burnout in Love and Marriage.* New York St. Martin's Press

Pines, A. M. & Aronson, E. with Kafry, D (1981). *Burnout. From Tedium to Personal Growth.* New York. Free Press.

Pines, A M & Aronson, E. (1988) *Career Burnout: Causes and Cures* New York: Free Press.

Ray, E B. & Miller, K I (1991) The influence of communication structure and social support on job stress and burnout. *Management Communication Quarterly*, **4**(4), 506–527.

Schaufeli, W. B. & Leiter, M P (1996). The MBI—General Survey In C Maslach, S. Jackson & M P Leiter (Eds), *Maslach Burnout Inventory: Manual*, 3rd ed. Palo Alto, CA· Consulting Psychologists Press.

Schaufeli, W B. & Van Dierendonck, D (1993) The construct validity of two burnout measures *Journal of Organizational Behavior*, **14**, 631–647

Schaufeli, W. B & Van Dierendonck, D (1994) Burnout, een begrip gemeten De Nederlandse versie van de Maslach Burnout Inventory (MBI–NL) (Burnout. Measurement of a concept. The Dutch version of the Maslach Burnout Inventory (MBI–NL)) *Gedrag & Gezondheid*, **22**, 153–172

Schaufeli, W B. & Van Dierendonck, D (1995) A cautionary note about the cross-national and clinical validity of cut-off points for the Maslach Burnout Inventory *Psychological Reports*, **76**, 1083–1090

Schulz, R , Greenley, J R & Brown, R (1995) Organization, management, and client effects on staff burnout. *Journal of Health and Social Behavior*, **36**, 333–345

Sonnentag, S, Brodbeck, F. C, Heinbokel, T & Stolte, W (1994) Stressor–burnout relationship in software development teams *Journal of Occupational and Organizational Psychology*, **67**, 327–341.
Spielberger, C D (1972) Anxiety as an emotional state In C. D Spielberger (Ed), *Anxiety Current Trends in Theory and Research*, vol 1, pp 24–49 New York: Academic Press.
Strelau, J (1983) *Temperament, Personality, Activity* New York Academic Press.
Tanaka, J S (1993) Multifaceted conceptions of fit in structural equation models In K. A Bollen & J S Long (Eds), *Testing Structural Equation Models*, pp 10–39 Newbury Park, CA: Sage
Warr, P, Barter, J. & Brownbridge, G (1983) On the independence of positive and negative affect *Journal of Personality and Social Psychology*, **44**, 644–651

Received 3 June 1996; revised version received 6 June 1997

Appendix: Burnout Measure

Item of the BM
How often did you have any of the following experiences during the last month?

1.	PE	E	Being tired.
2	EE	D	Feeling depressed.
3	*ME*	*L*	*Having a good day.*
4.	PE	E	Being physically exhausted.
5	EE	E	Being emotionally exhausted.
6.	*ME*	*L*	*Being happy.*
7	PE	E	Being 'wiped out'.
8	EE	E	Feeling 'burned out'.[a]
9	ME	D	Being unhappy.
10.	PE	E	Feeling rundown.
11	ME	D	Feeling trapped.
12	EE	D	Feeling worthless.
13	PE		Being weary.
14	EE	D	Being troubled.
15	ME		Feeling disillusioned and resentful about people.[b]
16	PE	D	Feeling weak.[c]
17.	EE	D	Feeling hopeless.
18.	ME	D	Feeling rejected
19	*ME*	*L*	*Feeling optimistic.*
20	*PE*	*L*	*Feeling energetic.*
21.	EE	D	Feeling anxious

[a]Modified in Pines & Aronson (1988) to 'Can't take it anymore'
[b]Modified in Pines & Aronson (1988) to 'Feeling disillusioned and resentful'
Modified in Pines & Aronson (1988) to 'Feeling weak and susceptible to illness'
Note PE=physical exhaustion, EE=emotional exhaustion, ME=mental exhaustion, E=exhaustion, D=demoralization, L=loss of motive