

The Validity of the Maslach Burnout Inventory in Three National Samples

DIRK ENZMANN, WILMAR SCHAUFELI AND
NOËLLE GIRAULT

In a recent article on "occupational morbidity and burnout: lessons and warnings for HIV/AIDS carers" Miller (1991) posed ten questions having priority to be answered. The first is: "Does burnout really exist in HIV care?" (Miller, 1991, p. 447). Prerequisites for giving an answer to this question are a clear and consistently used definition of burnout and its operationalization given by a well developed measure. A multitude of publications exist that stress the risk of burning out in AIDS health care. Many of them merely rely on vague descriptions of this "phenomenon" (e.g. Bolle, 1988; Howze, 1988; Roozenburg, 1988; Bor, 1989; LeBourdais, 1989; Voelcker, 1991; Broadhead, 1992). The vast majority of those researchers, however, who study the issue of burnout in AIDS health care by means of standardized questionnaires, employ the Maslach Burnout Inventory (MBI; Maslach and Jackson, 1981, 1986), which is – together with the Burnout Measure (BM; Pines, Aronson and Kafry, 1981, Pines and Aronson, 1988) - the most frequently used instrument in research on burnout (cf. Schaufeli, Enzmann and Girault, 1993). In doing so they adopt the multidimensional definition of burnout according to the subscales of the MBI.

"Burnout is a syndrome of emotional exhaustion, depersonalization, and reduced personal accomplishment that can occur among individuals who do 'people work' of some kind. The Emotional Exhaustion subscale assesses feelings of being emotionally overextended and exhausted by one's work. The Depersonalization subscale measures an unfeeling and impersonal response towards recipients of one's service, care, treatment, or instruction. The Personal Accomplishment subscale assesses feelings of competence and successful achievement in one's work with people." (Maslach and Jackson, 1986, pp. 1-2)

The demand for consistency in the definition of burnout and the prevailing use of a specific instrument puts one on the horns of a dilemma. On the one hand it provides comparability between studies and cross-validation of findings, on the other hand it makes it difficult to improve the psychometric properties of the measure or to adopt it to new theoretical developments. Thus, if the internationally prevailing use of the MBI is to have a favourable effect on the research on burnout, it is especially important to ascertain whether the MBI is psychometrically valid in different translations and populations. In order to answer the first priority question (does burnout really exist in AIDS health care?) and to compare the findings internationally, at least three steps of research are necessary: the reliability and validity of the measure has to be ascertained (1) internationally and (2) in the particular samples of AIDS health carers compared with other human services professionals, (3) the burnout scores in AIDS health care have to be compared with scores in other occupational fields and to other studies in the field of AIDS. The importance of the first issue is not limited to international research on burnout in AIDS health care, but also extends to research on burnout in other fields of human services. Whereas the data required to answer the second question are few and far between, there is sufficient material to answer the first. Pertaining to the third issue, only very few studies on burnout in AIDS health care are published. In this stage of research, it would be too early to give evidence on the true prevalence on burnout in the field of AIDS (cf. Enzmann, Chapter 2).

The purpose of this chapter is to address the first issue and to present findings about the reliability and cross-national validity of the MBI in three European samples of human services providers from the Netherlands, Germany, and France. The findings serve to indicate whether it is possible to translate the MBI into different languages, while preserving its psychometric properties. Three questions will be addressed: (1) Are the three MBI scales reliable indicators of burnout?, (2) Is the MBI a valid measure of burnout?; and (3) Is the factorial structure of the MBI invariant across nations? These questions pertain to inter-item reliability, construct validity, and factorial validity, respectively. Before considering these questions, we will describe briefly the three samples and the measures that were used for validation purposes.

SAMPLES

All three samples consisted of human services professionals. The Dutch sample is relatively homogeneous and consists of nurses who work in general hospitals (20%), mental hospitals (25%), nursing homes (18%), community health (23%), institutions for the mentally handicapped (11%), and other health care institu-

tions (3%). Data were collected independently in four studies: Study A ($n = 68$), Study B ($n = 122$), Study C ($n = 118$), and Study D ($n = 351$). For the current chapter these samples were combined into one composite sample ($n = 659$). The majority of the sample (61%) was female, and 39% was male. The nurses' mean age was 33.2 years ($Sd\ 6.5$).

The *German* sample ($n = 266$) included general hospital nurses (24%), geriatric nurses (9%), social workers (21%), physicians (15%), psychologists (9%), educators (8%), and a remaining group of various professionals, such as occupational therapists and counsellors (11%). Most subjects were female (73%), 27% were male. The mean age was 34.9 years ($Sd\ 8.8$). This sample was combined from two studies.

Study A included 121 professionals from public institutions (e.g. general and mental hospitals, agencies for family care, and youth centres); *Study B* included 145 professionals who work in different public institutions with HIV-infected and AIDS patients (60%), with the elderly (27%), or with cancer patients (13%) (they represent the first respondents of the ABBA-project; see Kleiber and Enzmann in Chapter ?).

The *French* sample ($n = 267$) consisted of health professionals (79%) and personnel officers (21%). The health professionals worked in particularly demanding jobs, either in hospitals (20%, i.e. intensive care units and oncology wards) or in mobile emergency teams (63%, i.e. ambulances or fire brigade). The remaining 17% was employed in other wards where the work is particularly stressful, such as paediatrics. Most professionals in the total sample were physicians (46%), followed by nurses (32%) and personnel officers (21%). Sixty-one percent of the sample was male, and 39% was female. About half (52%) of the sample was between 25 and 35 years of age, 23% was younger than 25, and 25% was older than 35. A subsample of 58 physicians and nurses who worked in ambulance emergency teams were studied in more detail, with several additional measures.

MEASURES

In the three national samples, translated versions of the *MBI* were employed. The term "recipient" was replaced by "patient" in the Dutch and French versions. This was considered more appropriate in the medical context in which the inventory was employed. In the Dutch as well as in the German sample A, the *BM* was completed as an alternative burnout instrument. Internal consistency coefficients for the *BM* were very high in the Dutch and German samples (both Cronbach's α were .93).

Because the national samples were combined from several studies, not all measures were employed in all of the subsamples. In the *Dutch* sample, self-reported *physical stress symptoms* ($\alpha = .83$) and *mental distress* ($\alpha = .82$) were assessed by the corresponding scales of the Dutch adaptation of the Organizational Stress Questionnaire, "Vragenlijst Organisatie Stress" (VOS; Bergers, Marcelissen and de Wolff, 1986). The nurses' level of *reactivity* was assessed by employing the corresponding subscale of the Strelau Temperament Inventory ($\alpha = .68$) (Strelau, 1983). Reactivity refers to the intensity of an individual's reaction to both external and internal stimuli. *Neuroticism* ($\alpha = .82$) and *self-esteem* ($\alpha = .73$) were measured by the N-scale of Eysenck's Personality Questionnaire (Eysenck, 1970) and by a scale developed by Warr and Jackson (1982), respectively.

For both *German* samples, several additional measures were obtained. In *sample A* the level of *depression* ($\alpha = .88$) was assessed by Beck's Depression Inventory (BDI; Beck, Rush, Shaw and Emery, 1972). *Job satisfaction* ($\alpha = .80$) was measured by a self-constructed six-item inventory (sample item: "I like being at my working place"). Finally, *irritability* ($\alpha = .87$) was assessed by an inventory developed by Mohr (1986). Irritability refers to a particular kind of fatigue that is caused by prolonged activity. The individual is forced to remain active, despite being fatigued. In an attempt to meet the demands, the individual activates his or her additional resources. As a consequence, resources are depleted, and the individual feels tense, nervous, and unable to relax, even when extra activities are no longer required. In *sample B* four additional measures were included. Level of *organizational commitment* ($\alpha = .90$) was assessed by the Organizational Commitment Questionnaire (Mowday, Steers and Porter, 1979). *Psychosomatic reactions* ($\alpha = .88$) were measured with a scale developed by Mohr (1986) in a sample of industrial blue collar workers. *Disillusionment* ($\alpha = .60$) was measured with a self-constructed seven-item inventory (sample item: "Only those persons who are willing to make personal sacrifices for the task should choose a job similar to mine"). Finally, *withdrawal* from clients and coworkers ($\alpha = .73$) was measured with a self-constructed thirteen-item inventory (sample item: "I don't bust a gut for clients who are unwilling to cooperate").

In the *French* subsample ($n = 58$), emotional and behavioural reactions that occurred after stressful periods at work were assessed by employing a free-response format. Respondents were classified by the third author into two *a priori* categories: (1) those who showed active reactions such as crying, asking for a time-out for recuperation, bursts of anger, excitement, and looking for social support from colleagues ($n = 26$); (2) those who reacted in a passive way with depressive moods, anxiety, compulsive thinking, inhibition, and withdrawal ($n = 32$). Moreover, withdrawal behaviours towards patients and colleagues were assessed (e.g. cynicism, "escaping" in rest rooms, over involvement in administrative duties). On the basis of their answers, respondents were categorized

into two groups: those who withdrew from their patients and colleagues ($n = 23$), and those who did not ($n = 35$). Finally, the respondents in this subsample were asked whether or not they intended to leave their jobs in the mobile emergency teams within one year. Again two groups were considered: those who intended to quit ($n = 26$), and those who intended to continue in their jobs ($n = 32$).

INTERNAL CONSISTENCY AND INTER-SCALE CORRELATIONS

Table 1 shows the internal consistency coefficients. Because the value of Cronbach's α depends on the number of items included in the scale, Table 1 also presents the corrected α -value which is standardized for a test-length of ten items, according to the Spearman-Brown formula

TABLE 1
Internal consistency (Cronbach's α) of the MBI scales in the Dutch, German, and French samples

scale	Dutch ($n = 659$)		German ($n = 266$)		French ($n = 267$)	
	α	cor. α	α	cor. α	α	cor. α
Emotional exhaustion	.88	.90	.84	.86	.80	.82
Depersonalization	.72	.84	.71	.83	.78	.88
Personal accomplishment	.77	.79	.81	.84	.66	.71

According to the commonly employed criterion of $\alpha > .70$ (Nunnally, 1978, p. 245), the internal consistencies of the MBI scales are satisfactory in every sample, particularly when the test-length is standardized (corrected α -coefficient).

In Table 2 the correlations are presented between the MBI scales in each

TABLE 2
MBI inter-scale correlations in three national samples

	Dutch ($n = 535$)		German ($n = 266$)		French ($n = 267$)	
	EE	DP	EE	DP	EE	DP
DP	0.58***	-	0.52***	-	0.27***	-
PA	-0.40***	-0.43***	-0.25***	-0.15*	-0.10	-0.17**

Notes * $p < .05$, ** $p < .01$, *** $p \leq .001$, EE = emotional exhaustion, DP = depersonalization, PA = personal accomplishment

sample. The inter-scale correlations differ slightly between the samples from the three countries. In the Dutch sample, the scales are substantively and almost evenly correlated, whereas correlations in the German and particularly in the French, samples are much lower. However, the pattern of correlations is quite similar across the three samples: emotional exhaustion and depersonalization are fairly strongly and positively related, whereas personal accomplishment is more weakly and negatively related to both other scales.

CONSTRUCT VALIDITY

In this section we discuss the construct validity of the MBI in each national sample separately, because different additional measures have been included. Two aspects are considered: convergent validity (i.e. to what extent do different burnout inventories measure the same construct?), and discriminant validity (i.e. to what extent can burnout be differentiated from related constructs?). In the next section, the factorial validity of the MBI will be studied cross-nationally (i.e. to what extent does a similar factorial structure emerge in each sample?).

Dutch Sample

In order to study the convergent validity of the MBI, correlations were computed with the total BM score (see Table 3)

TABLE 3
Relationships (Pearson's *r*) between burnout and related concepts (Dutch sample *n* = 659)

	EE	DP	PA	BM
Physical symptoms	0.52***	0.26***	-0.18**	0.59***
Mental distress	0.68***	0.43***	-0.38***	0.78***
BM	0.75***	0.43***	-0.30***	-

Notes ***p* < .01, ****p* ≤ .001, EE = emotional exhaustion, DP = depersonalization, PA = personal accomplishment, BM = burnout measure

It appears that the BM score is highly positively correlated with emotional exhaustion. This was to be expected, since Pines, Aronson, and Kafry (1981) define burnout as emotional, physical, and mental exhaustion. The correlations with depersonalization and personal accomplishment are modestly positive and weakly negative, respectively. The validity of the MBI is not questioned by these findings, however, because burnout is conceptualized multidimensionally by

Maslach and Jackson (1986). Similar relationships between the two burnout instruments have been found in other studies (Corcoran, 1986; Stout and Williams, 1983).

Table 3 also yields information about the discriminant validity of the MBI, (i.e. correlations with self-reported physical stress symptoms and mental distress). Typically, it was expected that correlations of the MBI with physical symptoms and mental distress would be weaker than with the BM. However, correlations of the MBI scales with mental distress are similar to those with the BM. Accordingly, the discriminant validity of the emotional exhaustion scale with self-reported mental distress is relatively poor: 45% of the variance of both variables is shared. The discriminant validity of the BM is even poorer: 60% of its variance is shared with mental distress. In contrast, mental distress is less substantially related to both other MBI scales. Physical stress symptoms are moderately correlated with emotional exhaustion and with the BM, but weakly with depersonalization and personal accomplishment. Thus, without exception the correlations between burnout and mental distress are stronger than with physical stress symptoms. It follows that depersonalization, and particularly personal accomplishment, can be reasonably discriminated from self-reported mental distress and physical stress symptoms. On the other hand, MBI-emotional exhaustion and the BM show a considerable overlap with these general indicators of ill-health.

In order to evaluate the discriminant validity of the MBI more thoroughly, two-factor analyses with scale means were performed in subsamples in which other relevant variables had been assessed. Table 4 shows the results of the principal components analyses after subsequent varimax rotation.

Although the results of both factor analyses are not identical, some important similarities can be observed. Most remarkably, in both samples *no* burnout component emerged on which the BM and all scales of the MBI load exclusively. Instead, personal accomplishment loads on the second component, together with personality characteristics such as reactivity and self-esteem. This finding confirms the results from other similar factor analyses (Brookings *et al*, 1985; Dignam and West, 1988). The first component is dominated by strains such as (emotional) exhaustion, depersonalization, neuroticism, mental distress, and physical stress symptoms.

German Sample

Different sets of measures were used in both German subsamples; therefore, validity is examined separately for each subsample

In order to study the convergent validity of the MBI in *Sample A*, the three scales were correlated with the BM. As can be seen from Table 5, emotional

TABLE 4
Factor analysis of scale means (Dutch subsamples)

Scales	Sample C (n = 188)		Sample D (n = 351)	
	I	II	I	II
BM	0.77	0.46	0.81	-0.43
EE	0.79		0.75	-0.46
DP	0.73		0.35	-0.62
PA	-0.37	-0.52	-0.37	0.69
Physical symptoms	0.34	0.67	0.83	
Mental distress	0.84		0.78	
Reactivity		0.89		-0.66
Self-esteem ^{a)}	-	-		0.79
Neuroticism ^{b)}	0.75		-	-
Explained variance	40.7%	24.5%	35.1%	30.1%

Notes. ^{a)} Not included in sample C

^{b)} Not included in sample D

BM = burnout measure, EE = emotional exhaustion

DP = depersonalization, PA = personal accomplishment

Only factor-loadings $\geq |.30|$ are presented

exhaustion correlated highly positively with the BM score. As in the Dutch sample, the remaining scales of the MBI correlate only modestly or weakly with the BM. Only the relationship between the BM and MBI depersonalization is somewhat weaker in the German sample.

TABLE 5
Relationships (Pearson's r) between burnout and related concepts
(German sample A n = 121)

	EE	DP	PA	BM
Irritability	0.48***	0.29**	-0.36***	0.68***
Depression	0.40***	0.15*	-0.41***	0.66***
Job satisfaction	-0.52***	-0.47***	0.38***	-0.57***
BM	0.69***	0.23**	-0.41***	-

Notes. * $p < .05$, ** $p < .01$, *** $p \leq .001$, EE = emotional exhaustion, DP = depersonalization, PA = personal accomplishment, BM = burnout measure

The discriminant validity of the MBI was investigated by correlating its scales with irritability, depression, and job satisfaction. The three MBI scales were expected to correlate more strongly with the BM than with these three variables. Furthermore, irritability was expected to correlate more strongly with emotional exhaustion than with the other scales of the MBI. The results, presented in Table

5, only partly confirm these expectations. In particular, depersonalization correlates more highly with job satisfaction than with the BM. However, this is not considered a negative result, because the BM correlates highly (and even more strongly than all the MBI scales) with irritability, depression, and job satisfaction. As in the Dutch sample, it is concluded that the BM has a considerably lower discriminant validity than the MBI.

TABLE 6
Factor analysis of scale means (German sample A n = 121)

Scales	I	II	III
BM	0.87		
EE	0.68	0.50	
DP		0.93	
PA			0.93
Depression	0.85		
Irritability	0.77		
Job satisfaction	-0.38	-0.58	0.47
Explained variance	38.9%	21.2%	17.9%

Notes: EE = emotional exhaustion, DP = depersonalization
PA = personal accomplishment
Only factor-loadings ≥ 0.30 are presented

A better understanding of the construct validity of the MBI can be obtained by a factor analysis of the scale means of the MBI, the BM, and the three related concepts (see Table 6). Contrary to the expectation that all burnout scales would load on one component and the related concepts on other components, the first principal component includes all strains (emotional exhaustion, the BM, irritability, and depression), a result much like that found in the Dutch sample. The remaining two components are dominated by depersonalization and personal accomplishment, respectively. Job satisfaction loads substantially on the second as well as the third component. Hence, the results of this analysis indicate the convergent validity of emotional exhaustion and the BM. However, at the same time, these scales lack discriminant validity with such related concepts as irritability and depression.

Additional findings concerning the discriminant validity of the MBI come from Sample B. Broadly speaking, all correlations in Table 7 are moderate. Emotional exhaustion shows the highest correlations with psychosomatic symptoms and with commitment, whereas personal accomplishment and depersonalization are most strongly related to disillusionment and withdrawal, respectively. These moderate relationships underscore the conceptual validity of the MBI scales. Emotional exhaustion reflects the person's level of generic stress and is therefore likely to be related to psychosomatic strains

TABLE 7
Relationships (Pearson's *r*) between burnout and related concepts
(German sample B *n* = 108)

	EE	DP	PA
Psychosomatic reactions	0.54***	0.21*	-0.40***
Disillusionment	0.40***	0.31**	-0.48***
Organizational commitment	-0.40***	-0.20*	0.29**
Withdrawal	0.39***	0.44***	-0.30**

Notes **p* < .05, ***p* < .01, ****p* ≤ .001, EE = emotional exhaustion; DP = depersonalization, PA = personal accomplishment

TABLE 8
Factor analysis of scale means (German sample B *n* = 108)

Scales	I	II	III
EE		0.60	0.64
DP		0.87	
PA	-0.71		-0.31
Disillusionment	0.81	0.35	
Organizational commitment	-0.68		
Withdrawal	0.37	0.73	
Psychosomatic reactions			0.88
Explained variance	26.5%	25.4%	20.1%

Notes EE = emotional exhaustion, DP = depersonalization

PA = personal accomplishment

Only factor-loadings ≥ |.30| are presented

Depersonalization is considered to be a particular coping response that is dominated by distancing and withdrawing from contacts with recipients. Finally, feelings of personal accomplishment result from a positive evaluation of one's work with recipients, which is obviously related negatively to disillusionment.

Again, a factor analysis with the scores of the scales was performed (see Table 8). The results from this analysis confirm the previous interpretations. The first component is dominated by lack of personal accomplishment, disillusionment, and lack of commitment. This indicates that reduced personal accomplishment is accompanied by negative reactions towards the organization (diminished commitment) and by a loss of confidence in the effectiveness of one's professional role. Accordingly, the lack of accomplishment is not restricted to the personal domain but also affects the organizational and professional domains. The second component is dominated by depersonalization and withdrawal. This confirms that withdrawal from coworkers and recipients on a

behavioural level goes along with negative attitudes towards recipients. Finally, the third component consists mainly of both strains that were included in the analysis, i.e. emotional exhaustion and psychosomatic symptoms.

French Sample

Unfortunately, neither an alternative burnout inventory nor other instruments that measure related concepts were included in the French study. Therefore, the validity of the MBI could not be studied analogously to the Dutch and German samples. In the French subsample ($n = 58$), the construct validity of the MBI was examined by relating the MBI scores to emotional and behavioural reactions, to withdrawal, and to the intention to leave the job (cf. Girault, 1989).

The expectation that emotional exhaustion would be associated with passive reactions to stress was confirmed. The mean emotional exhaustion score of the health professionals who reacted in a passive way to stress, was significantly higher than those of their colleagues who reacted in a more active fashion ($t_{(56)} = 3.66, p < .01$). Similar results, which illustrate the inadequacy of passive ways of coping, particularly with respect to the emotional component of burnout, have also been reported in other studies (Gillespie, 1981, Pines and Kafry, 1981). Moreover, as expected, levels of depersonalization were significantly higher for professionals who tended to withdraw from their patients and their colleagues, than for those who did not ($t_{(56)} = 1.88, p < .05$). This result confirms Maslach's (1982) original interpretation of depersonalization as a way of mental distancing. Finally, respondents who indicated that they would like to quit their jobs had significantly lower personal accomplishment scores than those who intended to continue their jobs ($t_{(56)} = 5.25, p < .001$). As expected, feelings of professional inefficacy and demotivation are associated with the desire to leave the job. It is concluded that the three MBI dimensions were successfully validated in the French sample, as a result of assessing different variables for each scale.

THE CROSS-NATIONAL VALIDITY OF THE MBI

The question to be addressed in this section is the extent to which the original three-factorial structure of the MBI (Maslach and Jackson, 1981-1986) can be reproduced in the Dutch, French, and German samples. In order to test the fit of the original theoretical model, structural equations modelling was employed, using the LISREL VII computer program (Jöreskog and Sorbom, 1989). Thus far, only two studies have employed a similar procedure (i.e. confirmatory factor analysis). Unfortunately, in one study seven items were added to the original

MBI (Lee and Ashforth, 1990), so that our results can only be compared with the findings of Gold *et al.*, (1989).

The Procedure of Fitting Models

In structural equations modelling, the traditionally employed goodness-of-fit index (χ^2) strongly depends on sample size. Because in the present study the sizes of the national samples differ considerably, a special strategy was followed. According to this approach, a so-called Type 2 incremental fit index (χ^2 -I2) is computed, which is scarcely affected by sample size (Marsh, Balla and McDonald, 1988) (cf. Equation 1). This index does not provide information about the absolute fit of a particular model; rather, it assesses the fit relative to a (nested) model in that particular sample.

$$\chi^2\text{-I2} = (n - t)/(n - df) \quad (\text{Equation 1})$$

with χ^2 -I2: incremental (Type 2) χ^2
t. χ^2 of target model
n: χ^2 of null model
df: degree of freedom of target model.

At the basis of these model comparisons lies the most restrictive model: the so-called null-model (M_0). In the present case, M_0 corresponds to the hypothesis that there are just as many uncorrelated factors as there are items. The factorial models of the MBI with which M_0 is compared are less restrictive. They can be ordered according to their degree of restriction. For instance, an oblique factor-model that allows the three dimensions of the MBI to correlate is less restrictive than an orthogonal or uncorrelated model.

Each comparison of the null-model with one of these models results in an incremental fit index (χ^2 -I2) which can be compared with the corresponding fit-index of the other samples of different size. Accordingly, the goodness-of-fit of different models within one sample, as well as between different samples, can be assessed.

In order to investigate the cross-national factorial validity of the MBI, three models were tested.

- M_1 : A one-factor model that assumes that all MBI-items load on one single factor.
- M_2 : The three-factor model used by Maslach and Jackson (1981, 1986), which assumes implicitly that the MBI items load on three uncorrelated factors, i.e. emotional exhaustion, depersonalization, and personal

accomplishment. Their factor-interpretation is based upon loadings that resulted from orthogonal rotation (The three scales are correlated, because instead of factor-scores, unweighted items-scores are used for computing the scale-scores. An oblique rotation would probably have resulted in different factor-loadings, and therefore also in a different factor-interpretation)

- M_3 . A three-factor oblique model in which the three factors of M_2 are allowed to be correlated.

Such a confirmatory factor analytic approach is particularly well suited to the study of complex constructs such as burnout because it has the capacity to analyze correlations among factors simultaneously, while considering a hypothesized factor structure.

Absolute and Relative Model Fit

The results of the analyses are presented in Table 9 (model tests) and in Table 10 (model comparisons). Table 9 shows (aside from the χ^2) additional fit-indices that are provided by LISREL, like the GFI (Goodness-of-Fit-Index), the AGFI (Adjusted Goodness-of-Fit-Index), and the RMR (Root Mean Square Residual). Low values of χ^2 and RMR, and high GFI- and AGFI-values are indicative of a good model fit (Jöreskog and Sörbom, 1989).

In each of the national samples, the probability levels of the χ^2 statistic are less than .001, indicating a rather poor absolute fit. This is most probably caused by the large sample sizes.

The χ^2 -I2-values (which are least affected by sample size) are shown in Table 10. High values are indicative of a good model fit. The best relative fit of the three factorial models is found in the German sample, followed by the Dutch, the American, and the French samples, respectively. (The χ^2 -I2-values of the sample of Gold *et al*, (1989) have been computed by the authors.)

In each of the national samples the three-factor oblique model (M_3) fits the data significantly better than the orthogonal model (M_2). It is remarkable that the fit of M_2 in the American sample of Gold *et al*, (1989) (χ^2 -I2 = 771) is poorer compared to the fit obtained in the three European samples, given that the American study used the untranslated English version of the MBI. However, Gold *et al* adapted the MBI slightly in order to make the instrument suitable for their student sample. Perhaps these changes had a negative influence on the factorial structure of the instrument. Moreover, given that the MBI was constructed via an inductive approach using an American population, one would not have expected the even worse fit of M_2 in the Gold *et al* sample.

TABLE 9
Test of MBI factorial model in three national samples

Sample	Model	χ^2	df	GFI	AGFI	RMR
Dutch (n = 535)	M ₀	4473.03	231	.345	.282	.287
	M ₁	1438.59	209	.736	.680	.096
	M ₂	1143.94	209	.832	.797	.189
	M ₃	814.91	206	.870	.840	.097
German (n = 266)	M ₀	1961.28	231	.421	.366	.245
	M ₁	957.83	209	.669	.599	.121
	M ₂	508.99	209	.848	.816	.141
	M ₃	416.78	206	.864	.833	.071
French (n=267)	M ₀	1615.69	231	.542	.498	.192
	M ₁	951.28	209	.713	.653	.117
	M ₂	517.69	209	.856	.826	.098
	M ₃	495.46	206	.860	.829	.078
American ^{a)} (n = 147)	M ₀	1628.62	231	.336	.273 ^{b)}	.293
	M ₁	674.47 ^{c)}	209	.636	.559	.122
	M ₂	534.44 ^{d)}	209	.747	.693	.171
	M ₃	483.26	206	.764	.710	.095

Notes: $p \leq .001$ for all χ^2 -statistics,

M₀ = null model, M₁ = one-factor model, M₂ = three-factor orthogonal model,

M₃ = three-factor oblique model

^{a)} Gold *et al.* (1989), ^{b)} Corrected by the authors, original 0.730, ^{c)} Corrected by the authors, original 764.47, ^{d)} Corrected by the authors, original 543.44

It is important to note that researchers from other countries who want to study the cross-national validity of their MBI versions can compare the obtained fit-indices with the values of Tables 9 and 10.

Improving the Model Fit

LISREL provides information about which items in each national sample are responsible for the rather poor fit of the original three-factorial orthogonal model (M₂). Inspection and subsequent testing of the so-called modification indices (M₀) reveals that in each European sample a significant improvement of fit would occur if M₂ is re-estimated allowing item 12 ("I feel very energetic") to load additionally on emotional exhaustion. (The M₀ for a particular item indicates the increase of the χ^2 -statistic if that item would be "allowed" to load on another particular factor. In other words, the M₀ provides information as to whether or not an item loads on the "right" factor.) In order to compare the M₀s across samples, we computed a corresponding value ($\Delta\chi^2$ -12) which can be interpreted similarly to the χ^2 -12 statistic. The M₀s for item twelve on the

TABLE 10
Comparisons of MBI factorial models in three national samples

Sample	Comparison	χ^2	df	χ^2 -I2
Dutch (n = 535)	M_0 - M_1	3034.44	22	.712
	M_0 - M_2	3329.09	22	.781
	M_0 - M_3	3658.12	25	.857
	M_2 - M_3	329.03	3	.077
German (n = 266)	M_0 - M_1	1003.45	22	.573
	M_0 - M_2	1452.29	22	.829
	M_0 - M_3	1544.50	25	.880
	M_2 - M_3	92.21	3	.051
French (n=267)	M_0 - M_1	664.41	22	.472
	M_0 - M_2	1098.00	22	.781
	M_0 - M_3	1120.23	25	.795
	M_2 - M_3	22.23	3	.014
American ^{a)} (n = 147)	M_0 - M_0	954.15	22	.672
	M_0 - M_2	1094.18	22	.771
	M_0 - M_3	1145.36	25	.805
	M_2 - M_3	51.18	3	.034

Notes $p \leq .001$ for all χ^2 -statistics,

M_0 = null model, M_1 = one-factor model, M_2 = three-factor orthogonal model, M_3 = three-factor oblique model

^{a)} Gold *et al* (1989), all values have been computed by the authors

emotional exhaustion factor are 132.7, 29.0, 35.2 for the Dutch, French, and German samples respectively. The corresponding χ^2 -I2-values are: .031, .020, and .020. A significant factor loading of item twelve on emotional exhaustion clearly makes sense psychologically, because "not feeling energetic" can be considered as just another symptom of exhaustion.

The remaining significant modification indices seem to be rather specific for each of the national samples. In the Dutch sample, items twelve and eighteen ("I feel exhilarated after working closely with my recipients"; personal accomplishment) tend to load on the depersonalization factor ($M_0 = 30.1$, $\Delta\chi^2$ -I2 = .007, and $M_0 = 38.2$, $\Delta\chi^2$ -I2 = .009, respectively). In the French sample, item ten ("I've become more callous toward people since I took this job"; depersonalization) tends to load on emotional exhaustion ($M_0 = 12.0$, $\Delta\chi^2$ -I2 = .008). In the German sample, item sixteen ("Working with people directly puts too much stress on me", emotional exhaustion) tends to load on depersonalization ($M_0 = 22.3$, $\Delta\chi^2$ -I2 = .012). Thus, except for one common adaptation of the three-factor orthogonal model that leads to an improved fit (item twelve), the other (minor) adaptations are country-specific.

Is the MBI theoretically or cross-nationally invalid?

At this point, an interesting question emerges. Can the deviations from the theoretical model (M_2) that we found in the European samples be attributed to cultural differences with the American normative sample in which this theoretical model was developed? In other words, is the MBI cross-nationally invalid, or are the deviations independent of cross-national differences? If the poor fit of the model in the study of Gold *et al.*, (1989) cannot be explained by their somewhat atypical sample (college students) and rephrased items, then their results point in the direction of factorial invalidity of the original model. A test of the assumption that the misspecifications in the model are of a theoretical rather than of a cultural nature is in progress, via a reanalysis of the American normative sample (Schaufeli, Enzmann, Girault and Maslach, forthcoming).

From the current analyses we conclude that the factorial validity of the MBI is not as convincingly demonstrated as one would have been expected from the literature. However, there are some limitations pertaining to our analyses. First, we only compared three national samples. Second, the samples are rather small (particularly the German and French samples). Finally, the samples consist of different occupational groups. In spite of these obvious limitations, certain results are remarkably stable across the samples: (1) significant misfits occur for the orthogonal as well as for the oblique model (model tests); (2) both models capture a substantial aspect of the data (model comparisons); (3) the oblique three-factor model fits the data significantly better than the other models.

In sum, it seems unlikely that sampling effects, cultural influences, or merely differences in translation can be held responsible for these somewhat disappointing results. It is clear that further analyses of the factorial validity of the MBI are needed that apply linear structure modelling techniques to representative samples from different countries.

GENERAL DISCUSSION AND CONCLUSIONS

Particularly in the Dutch and German samples, the MBI positive psychometric results were found. (In the French sample, very few analyses could be performed because almost no additional measures were included.) Moreover, we showed that the three-dimensional structure of the MBI is invariant across samples of three different countries. The fact that the factorial model did not fit perfectly is most probably not caused by cultural bias of some sort, but by misspecifications in the original model. However, a rigorous test of this hypothesis has yet to be done (Schaufeli *et al.*, forthcoming). Golembiewski and his colleagues have also shown the congruence of MBI dimensions across

different nations (Golembiewski, Scherb and Boudreau, 1993). However, they employed a slightly modified version of the MBI and used a less rigorous and more subjective method of testing.

On the other hand, our results also point to the poor discriminant validity of the emotional exhaustion scale of the MBI and of the BM scale. Feelings of exhaustion overlap considerably with similar strains like depression, neuroticism, irritability, psychosomatic reactions, and mental and physical stress symptoms. It has been argued by Watson and Pennebaker (1989) that subjective measures of health and well-being reflect, to a large extent, the perceptions of, and interpretations about, the individual's internal sensations. Thus, a common underlying factor of "negative affectivity" might explain the relatively poor discriminant validity of the MBI emotional exhaustion scale and of the BM scale.

Positive indications were found for the discriminant validity of both of the other MBI scales. Depersonalization is particularly linked to withdrawal from patients and colleagues. This agrees with Maslach's (1982) interpretation of depersonalization as being a way of coping with emotional exhaustion by mental distancing (see also Leiter, 1993). In our study, personal accomplishment is strongly related to favourable personality features, and to positive attitudes towards the organization and one's professional role. This hints at a conceptual relatedness of personal accomplishment with psychological coping resources, such as hardiness (Kobasa, 1982) and self-efficacy (cf. Cherniss, 1993). Generally, these resources play a stress-buffering role because they concern a person's sense of mastery, control and commitment. This view is supported empirically by Koeske and Koeske (1989), who showed that feelings of personal accomplishment moderate the relationship between emotional exhaustion and a particular stress reaction: the intention to quit. In their study, social workers who maintained a sense of accomplishment did not intend to quit despite high levels of exhaustion, whereas those who were exhausted and also had low feelings of accomplishment did intend to quit. This example illustrates a most important point: the discriminant validity of the MBI can only be discussed fruitfully within a conceptual framework that specifies the relationships of the MBI dimensions with each other and with various other variables (cf. Leiter, 1993).

Seven conclusions can be drawn from our cross-national validity study of the MBI.

- First, and most importantly, the psychometric results of the MBI are similar across nations.
- Second, the convergent validity of the emotional exhaustion scale is demonstrated convincingly: about one half of its variance is shared with the BM. The correlations of both other MBI scales with the BM are somewhat less substantive.

- Third, emotional exhaustion is correlated substantially with related concepts, so the discriminant validity of this scale is rather poor. Thus, emotional exhaustion must be considered a generic psychological stress reaction, that is, clearly related to other strains such as depression, neuroticism, irritability, psychosomatic reactions, physical stress symptoms, and mental distress. Moreover, the discriminant validity of the BM is also poor. This one-dimensional indicator of burnout only taps its most generic dimension (i.e. exhaustion) and is also substantially related to similar strains.
- Fourth, the depersonalization and personal accomplishment scales appear to be discriminantly valid. Depersonalization is related to withdrawal, whereas personal accomplishment is related to favourable personality characteristics (e.g. self-esteem), organizational commitment, and feelings of disillusionment. Thus, burnout (as measured by all three of the MBI scales) is a multidimensional construct that cannot be merely reduced to feelings of exhaustion.
- Fifth, as far as the cross-national validation of the factor-structure of the MBI is concerned, results from the three national samples are highly consistent. That is, the factorial structure of the MBI is similar across nations. The three-factor oblique model that allows the three factors to be correlated fits best to the data of each sample. Obviously, the implicit orthogonal model from the test manual that assumed the factors to be uncorrelated shows a relatively poor fit (cf. Gold *et al.*, 1989, Lee and Ashforth, 1990).
- Sixth, our results show that in the original model one particular item loads on the "wrong" factor: additionally on emotional exhaustion instead of only on personal accomplishment. This is not very surprising when the content of this particular item is considered ("I feel energetic").
- Finally, indications were found that the misspecifications that cause the imperfect fit of the original factorial model of the MBI most probably reflect theoretical flaws rather than cultural bias resulting from, for instance, inadequate translation or cross-national differences in perception or expression of symptoms.

DIRECTIONS FOR FUTURE RESEARCH

The MBI can be improved in a number of ways. Additional psychometric development of the depersonalization dimension seems necessary. Depersonalization is the shortest (and therefore least reliable) scale of the MBI, with the most complex factor loadings. Adding a couple of items, preferably about the behavioural element of depersonalization, would increase not only the internal consistency, but would also strengthen the validity of this scale (see

Table 8). Moreover, an equal balance of positively and negatively worded items is strongly recommended. In the present version, the items of the emotional exhaustion and depersonalization scales are phrased negatively, whereas those of the personal accomplishment scale are phrased positively. It cannot be completely ruled out that this fact explains the substantial correlation between the first two scales (see Tables 3, 5, and 7) and the particular behaviour of the latter scale (e.g. its high correlations with favourable personality characteristics). Finally, one item ("I feel energetic") should be deleted from the MBI since it not only loads on the intended personal accomplishment dimension, but also on emotional exhaustion. Inspection of the MBI manual (Maslach and Jackson, 1986, p. 30) reveals that item twelve is the weakest and most complex item in the accomplishment scale, which also loads on the exhaustion factor. Similar results have been obtained in other exploratory factorial validity studies (Koeske and Koeske, 1989; Mor and Laliberte, 1984).

REFERENCES

- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1972). *Cognitive therapy of depression*. New York: Guilford.
- Bergers, G. P. A., Marcelissen, F. H. G., & de Wolff, Ch. J. (1986). *Handleiding Vragenlijst Organisatiestress* [Test-manual Organizational Stress Questionnaire]. University of Nijmegen, The Netherlands, Department of Psychology.
- Bolle, J. L. (1988). Supporting the deliverers of care: Strategies to support nurses and prevent burnout. *Nursing Clinics of North America*, **23**, 843-850.
- Bor, R. (1989). AIDS counselling. *AIDS Care*, **1**, 184-187.
- Broadhead, R. S. (1992). The occupational risks of AIDS outreach workers. In A. S. Trebach, & Zeese, K. (Eds.), *Strategies for Change: New Directions in Drug Policy* (pp. 195-202). Washington, DC: Drug Policy Foundation.
- Brookings, J. B., Bolton, B., Brown, C. E., & McEvoy, A. (1985). Self-reported job burnout among female human service professionals. *Journal of Occupational Behaviour*, **6**, 143-150.
- Cherniss, C. (1993). The role of professional self-efficacy in the etiology and amelioration of burnout. In W. B. Schaufeli, C. Maslach, & T. Marek (Eds.), *Professional Burnout: Recent Developments in Theory and Research* (135-149). Washington, DC: Taylor & Francis.
- Corcoran, K. J. (1986). Measuring burnout: A reliability and convergent validity study. *Journal of Social Behavior and Personality*, **1**, 107-112.
- Dignam, J. T., & West, S. G. (1988). Social support in the workplace: Tests of six theoretical models. *American Journal of Community Psychology*, **16**, 701-724.
- Eysenck, H. J. (1970). *The structure of human personality*. London: Methuen.
- Gillespie, D. F. (1981). Correlates for active and passive types of burnout. *Journal of Social Service Research*, **4**(2), 1-16.
- Girault, N. (1989). Burnout: Emergence et strategies d'adaptation [Burnout: Emergence and strategies of adaptation]. Dissertation. Paris: Université René Descartes.
- Gold, Y., Bachelor, P. A., & Michael, W. B. (1989). The dimensionality of a modified form of the Maslach Burnout Inventory for university students in a teacher-training program. *Educational and Psychological Measurement*, **49**, 549-561.
- Golembiewski, R. T., Scherb, K., & Boudreau, R. A. (1993). Burnout in cross-national settings.

- Generic and model-specific perspectives. In W. B. Schaufeli, C. Maslach & T. Marek (Eds.), *Professional Burnout: Recent Developments in Theory and Research* (pp. 218-236). Washington, DC: Taylor & Francis.
- Howze, E. T. (1988). Dealing with stress and burnout. In T. Edison (Ed.), *The AIDS Caregiver's Handbook* (pp. 138-147). New York: St. Martin's Press.
- Joreskog, K. G., & Sorbom, D. (1989). *LISREL VII: User's reference guide*. Mooresville: Scientific Software.
- Kobasa, S. C. (1982). The hardy personality: Toward a social psychology of stress and health. In J. Suls, & G. Sanders, *Social psychology of health and illness* (pp. 333), Hillsdale, NJ: Erlbaum.
- Koeske, G. F., & Koeske, R. D. (1989). Construct validity of the Maslach Burnout Inventory: A critical review and reconceptualization. *Journal of Applied Behavioral Science*, **25**, 131-144.
- LeBourdais, E. (1989). Hopelessness and helplessness: Treating the doctors who treat AIDS patients. *Canadian Medical Association Journal*, **140**, 440-443.
- Lee, R. T., & Ashforth, B. E. (1990). On the meaning of Maslach's three dimensions of burnout. *Journal of Applied Psychology*, **75**, 743-747.
- Lester, M. P. (1993). Burnout as developmental process: Consideration of models. In W. B. Schaufeli, C. Maslach & T. Marek (Eds.), *Professional Burnout: Recent Developments in Theory and Research* (pp. 237-250). Washington, DC: Taylor & Francis.
- Maslach, C. (1982). Burnout: A social psychological analysis. In J. W. Jones (Ed.), *The Burnout Syndrome: Current Research, Theory, Interventions* (pp. 30-53), Park Ridge, Ill.: London House Press.
- Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behaviour*, **2**, 99-113.
- Maslach, C. & Jackson, S. E. (1986). *Maslach Burnout Inventory Manual* (2nd edn). Palo Alto, CA: Consulting Psychologists Press.
- Miller, D. (1991). Occupational morbidity and burnout: Lessons and warnings for HIV/AIDS carers. *International Review of Psychiatry*, **3**, 439-449.
- Mohr, G. (1986). *Die Erfassung psychischer Befindensbeeinträchtigungen bei Industriearbeitern* [The assessment of psychological distress in blue-collar workers]. Frankfurt/Main: Lang.
- Mor, V., & Laliberte, L. (1984). Burnout among hospice staff. *Health and Social Work*, **9**, 274-283.
- Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior*, **14**, 224-247.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Pines, A. M., & Aronson, E. (1988). *Career Burnout: Causes and Cures*. New York: Free Press.
- Pines, A. M., Aronson, E., & Kafry, D. (1981). *Burnout: From Tedium to Personal Growth*. New York: Free Press.
- Rozenburg, H. (1988). *Burnout in de hulpverlening bij HIV-infectie, voorkomen en voorkomen* [Burnout in the Services for the HIV-Infected, Prevalence and Prevention] (Nationale Commissie AIDS-Bestrijding, Oktober 1988, #6815/HR/88).
- Schaufeli, W. B., Enzmann, D., & Girault, N. (1993). Measurement of burnout: A review. In W. B. Schaufeli, C. Maslach, & T. Marek (Eds.), *Professional Burnout: Recent Developments in Theory and Research* (pp. 199-215), Washington: Taylor & Francis.
- Schaufeli, W. B., Enzmann, D., Girault, N., & Maslach, C. (forthcoming). *The factorial validity of the Maslach Burnout Inventory: A cross-national confirmatory analysis*.
- Stout, J. K., & Williams, J. M. (1983). Comparison of two measures of burnout. *Psychological Reports*, **53**, 283-289.
- Strelau, J. (1983). *Temperament, personality, activity*. New York: Academic Press.
- Voelcker, J. (1991). AIDS groups address caregiver burnout. *The Advocate*, No. 578, 50-53.
- Watson, D., & Pennebaker, J. W. (1989). Health complaints, stress, and distress: Exploring the central role of negative affectivity. *Psychological Review*, **96**, 234-254.