
MEASUREMENT OF BURNOUT: A REVIEW

Wilmar B. Schaufeli
University of Nijmegen

Dirk Enzmann
Technical University of Berlin

Nöelle Girault
René Descartes University

The importance of reliable and valid instruments to measure burnout is evident not only for the purpose of empirical research but ultimately for individual assessment. Because of the vagueness and overinclusiveness of the burnout concept, researchers were faced with a difficult problem. How is it possible to construct a burnout instrument if the boundaries of that phenomenon are unclear? Soon after the "discovery" of the syndrome in the mid-1970s, dozens of symptoms were associated with it, ranging from anxiety to lack of zeal.

In those early years, conceptual confusion prevailed. In the first major review on burnout, Perlman and Hartman (1982) counted more than 48 definitions. They formulated the following "synthetic definition": "Burnout is a

response to chronic emotional stress with three components: (a) emotional and/or physical exhaustion, (b) lowered job productivity, and (c) overdepersonalization" (p. 293). Quite interestingly, their characterization of burnout was rather similar to that of Maslach and Jackson (1981a,b), who a year before had published a burnout inventory that soon became the most widely used instrument. Probably, the recurrently cited summary of Perlman and Hartman with its appealing definition encouraged the use of the Maslach Burnout Inventory (MBI) by providing a conceptual justification. In any case, most burnout researchers were no longer trapped in the vicious circle of an ill-defined construct that could not be measured because it was not properly described.

Use of the MBI implies acceptance of the definition provided by the test authors. Maslach and Jackson (1981a, p. 99) describe burnout as a three-dimensional syndrome characterized by emotional exhaustion, depersonalization ("negative, cynical attitudes and feelings about one's client"), and reduced personal accomplishment ("the tendency to evaluate oneself negatively, particularly with regard to one's work with clients"). In their view the burnout syndrome is restricted to individuals who do "people-work" of some kind. Thus, as a consequence of accepting this definition, the controversy about the nature of burnout was settled by a silent agreement among researchers, at least as far as the human services professions were concerned. This development illustrates how closely the measurement of burnout is related to conceptual issues that have been discussed in Chapter 1 of this book.

The purpose of this chapter is to present a comprehensive overview of measures that have been employed to assess burnout. Several psychometric criteria are used for assessing the various measures, such as inter-item reliability, test-retest consistency, and complexity of the factor structure. Moreover, two types of validity information are considered: (1) consistency with self-report indicators and (2) all too rarely, consistency with indicators not based on self-report. Because most psychometric development has occurred with the MBI and with the Burnout Measure (BM) (Pines, Aronson, & Kafry, 1981), these instruments will be discussed in greater detail.

ALTERNATIVE MEASURES OF BURNOUT

Since the emergence of burnout almost two decades ago, several different kinds of instruments have been proposed to measure burnout. Originally, the syndrome was described rather unsystematically by what can be called clinical *observation* (see also Chapter 1). For example, one of the first people to use the term "burnout" was Freudenberg, a psychoanalytically oriented psychiatrist who observed many mental and physical symptoms among the free clinic staff members with whom he was working. Despite this initial clinical base, a systematic observation method to assess burnout has never been developed. However, a structured *interview* to assess the level of burnout was developed by

Forney, Wallace-Schutzman, and Wiggers (1982). Unfortunately, this approach has not been followed by other researchers.

Another idiosyncratic method to assess burnout involved *projective drawings* (Haack & Jones, 1983). Twenty-six nurses were asked to draw how burned out they felt, and each drawing was independently rated by two psychologists on a 4-point rating scale, ranging from "not burned out" to "very burned out." The nurses were divided into a high-burnout group and a low-burnout group, based on their scores on the Staff Burnout Scale for Health Professionals (SBS-HP; see next section). As expected, the high-burnout group drew pictures of burnout that were rated as expressing significantly more burnout than the low-burnout group. Common themes in the high burnout pictures were exhaustion, isolation, regression, powerlessness, being broken or injured, and feeling overwhelmed. Although empirical testing yielded positive results, this method is not very promising because the criteria for burned-out drawings are unclear. As a result, the ratings may be quite unreliable. Perhaps this is why the authors do not mention inter-rater reliabilities.

An interesting validity study employed an overall *self-assessment* of burnout (Rafferty, Lemkau, Purdy, & Rudisill, 1986). Seventy-six family practice residents were requested to describe themselves over the past several months, keeping in mind the following definition of burnout:

The tendency for committed physicians to lose enthusiasm for their work and to become less effective in managing the stress of emotional contact with patients. Symptoms may include some of the following—fatigue, withdrawal from patients and colleagues, cynicism, irritability, difficulty relaxing off work, physical manifestations of anxiety and depression, and feelings of diminished enthusiasm and effectiveness at work. (p. 489)

Responses were in the form of a 9-point rating scale that ranged from "not at all burned out" to "very burned out." Moreover, each physician completed the MBI. The physician's self-report of overall burnout was correlated moderately with MBI emotional exhaustion ($r = .48$ and $r = .59$ for the frequency and intensity dimensions, respectively) and slightly with the intensity dimension of the depersonalization scale ($r = .34$). These results suggest that physicians assess their global burnout particularly in terms of emotional exhaustion.

In a carefully designed multitrait-multimethod study, Meier (1984) investigated the construct validity of burnout in a sample of 320 male and female faculty members at a university. He included self-ratings of burnout, along with other measures. Subjects were presented with a brief description of burnout and were asked to indicate on a 7-point rating scale the extent to which the construct assessed their current state. This single-item burnout self-rating was moderately correlated with the MBI sum score ($r = .65$) and with two other scales (see

next section): the Meier Burnout Assessment scale ($r = .63$) and the Emener-Luck Burnout Scale ($r = .66$).

Most of the above-mentioned methods for measuring burnout are atypical and have only occasionally been used by individual researchers. The rating scales are an exception, however. Such single-item ratings have been employed more or less successfully as a criterion measure to validate the multi-item inventories that have been the most popular way to assess burnout. Therefore, the remainder of this chapter is devoted to this type of measurement instrument.

BURNOUT MEASURES WITH LIMITED APPLICATION

Self-report burnout measures can be distinguished according to the amount of psychometric research on which they are based. For instance, do-it-yourself inventories have not been investigated empirically at all. Most other self-report measures have been studied occasionally, whereas two instruments—the BM and the MBI—have been examined most extensively. Accordingly, these particular instruments will be discussed separately in the next two sections.

Do-it-yourself inventories of burnout have been published under such appealing headings as: "How burned-out are you?" (Bramhall & Ezell, 1981); "What's your burnout score?" (Steward & Meszaros, 1981); "The burnout-test—Examine your beliefs about work, about leisure, about yourself" (Daily, 1985); "The National Job Burnout Survey" (Veninga & Spradley, 1981). None of these inventories has been studied empirically in any way. Therefore, their use is rather limited. At best, they convey a clearer picture of the author's definition of burnout. At worst, the subject becomes alarmed unduly. The best known do-it-yourself burnout scale was proposed by Freudenberger and Richelson (1980, pp. 17–19). They formulated 15 questions, including: "Do you feel fatigued rather than energetic?", "Are you increasingly cynical and disenchanting?", and "Does sex seem like more trouble than it's worth?" The authors also provide an interpretation of the scale scores: individuals scoring in the lowest category are considered to be "doing fine," whereas others are informed that they are "a candidate for burnout" or that they are in fact "burning out." Individuals scoring in the highest category are warned: "You are in a dangerous place, threatening to your physical and mental well-being." However, the "norms" and the corresponding interpretations are not empirically based and should therefore not be taken as valid indicators of an individual's level of burnout.

A number of other burnout inventories have been employed exclusively in a single study. An example of a purely inductive approach is offered by Blostein, Eldridge, Kilty, and Richardson (1985). They factor-analyzed a set of 49 "burnout indicators" in a sample of 400 child welfare workers and found six underlying dimensions: (1) "classic" burnout (e.g., cynicism, depression); (2) negative feelings about clients (e.g., intolerance, postponement of client con-

tacts); (3) a feeling of being overstimulated (e.g., inability to relax, problems with sleeping); (4) a feeling of being overwhelmed (e.g., need to be alone, feeling pulled in all directions); (5) physical problems (e.g., colds and flu, headaches); and (6) lack of intimacy (e.g., emptiness, feeling unappreciated). Typically, in this type of exploratory study, burnout has been identified as a multidimensional concept. However, the obtained dimensions have not been specified theoretically.

Other studies have yielded more interesting results, particularly when additional burnout instruments have been included or when the burnout instrument has been designed for specific occupational groups. For instance, Farber (1984) developed a 64-item *Teacher Attitude Scale* (TAS), which is a modified version of the MBI. The 25 items of the original version of the MBI were augmented with 40 additional items of exclusive relevance to teachers. A factor analysis in a sample of 365 suburban teachers yielded three factors: (1) general feelings of burnout (mainly emotional exhaustion), (2) commitment to the teaching profession, and (3) gratification in working closely with students.

The *Staff Burnout Scale for Health Professionals* (SBS-HP) is a one-dimensional 30-item measure (Jones, 1980b). Twenty items assess burnout, and the remaining 10 items form a lie scale to detect tendencies to "fake good." According to the test manual, the SBS-HP assesses adverse cognitive, affective, behavioral, and psychophysiological reactions that are considered to constitute the burnout syndrome. The SBS-HP ($\alpha = .93$) has been validated in a series of studies with health professionals (Jones, 1980a). SBS-HP scores were found to be related to external stressors (e.g., high patient-to-staff ratios, high-trauma jobs, working undesirable shifts, and receiving little family support) and to stress reactions (e.g., job turnover, absenteeism, illness, job dissatisfaction, and tardiness rates). However, the samples of the validity studies are very small, ranging from only 27 to 36 health professionals. Moreover, the sizes of the correlation coefficients differ substantially across these studies. More interesting results have come from a larger study of 135 female human service professionals (Brookings, Bolton, Brown, & McEvoy, 1985). Moderate correlations were found between the SBS-HP score ($\alpha = .88$) and MBI emotional exhaustion ($r = .65$) and MBI depersonalization ($r = .54$), whereas the relationship with personal accomplishment was less strong ($r = -.33$). Furthermore, SBS-HP scores were positively correlated with job-related tension and negatively correlated with job satisfaction, internal locus of control, and self-esteem ($.37 < r < .61$).

In a previously mentioned validity study, Meier (1984) employed a self-constructed 23-item true-false burnout test: the *Meier Burnout Assessment* (MBA). Unfortunately, the author does not provide any information about the content of the items. The MBA ($\alpha = .79$) correlated moderately with the MBI sum score in a sample of 320 male and female faculty members ($r = .61$). In

another sample of 120 students, a similar correlation between the two measures ($r = .58$) was found (Meier & Schmeck, 1985).

Ford, Murphy, and Edwards (1983) constructed a 15-item burnout inventory that was meant to be used outside the human service professions. The items of their *Perceptual Job Burnout Inventory* (PJBI) reflect (1) emotional exhaustion and cynicism; (2) demoralized, frustrated feelings and reduced efficiency; and (3) excessive demands on energy, strength, and resources. Separate factor analyses were done for a social services sample ($N = 237$) and a corporate sample ($N = 150$). In the former sample five factors emerged ("winless," "supportless others," "controlless," "supportless organization," and fatigue), whereas in the latter sample only two factors were found ("supportless organization" and emotional exhaustion). Cronbach's α values for the scales ranged from .67 to .86 in both samples. Obviously, the PJBI cannot be used across different occupational groups.

The *Emener-Luck Burnout Scale* (ELBOS) consists of 30 items and has been employed in a sample of 251 professional human service providers (Emener, Luck, & Gohs, 1982). The internal consistency of the scale is satisfactory ($\alpha = .88$). A principal components factor analysis yielded seven factors, of which four were significantly related to the burnout self-rating that was used as a criterion measure. These factors are work-related feelings, work environment provisions, dissonance between the individual's perception of self and others' perception of self, and job alternatives. Neither the reliability coefficients of the scales nor their intercorrelations are presented.

Garden (1987a) employed a four-item *Energy Depletion Index* (EDI) in a sample of 95 MBA students with significant work experience. She argued that energy depletion is the dimension of the burnout experience for which there is most definitorial agreement. The EDI ($\alpha = .82$) is weakly correlated with two aspects of depersonalization: distancing and hostility ($r = .25$ in both cases).

In a similar vein, Shirom and Oliver (1986) define burnout as comprising physical, cognitive, and emotional exhaustion, and the wearing out of resources. In a cross-lagged panel study among 404 Israeli teachers, they employed a six-item *Burnout Index* ($\alpha = .86$; stability across 7 months $r_t = .70$) that correlated .45 with self-reported somatic complaints and .35 with intrinsic job dissatisfaction.

What is common to all the above-mentioned measures is that they assess feelings and emotions that are generated in work-related settings. This applies also for the BM and the MBI. In other words, burnout is considered exclusively on an individual level by the vast majority of the self-report inventories. However, some instruments also assess the situation on the job, i.e., the individual's perception of the organization and its employees. Such instruments that assess both individual and organizational aspects of burnout have been developed for nurses (*Nursing Stress Scale*—NSS; Gray-Toft & Anderson, 1981), teachers

(*Teacher Stress Inventory*—TSI; Fimian, 1984), psychologists (*Psychologist's Burnout Inventory*—PBI; Ackerley, Burnell, Holder, & Kurdel, 1988), and medical personnel (*Medical Personnel Stress Survey*—MPSS; Hammer, Jones, Lyons, Sixsmith, & Afficiando, 1985).

Some promising findings about the convergent validity of the TSI are provided by Fimian and Blanton (1987). There was a moderate positive correlation ($r = .64$) between the MBI and TSI total scores in a sample of 413 teacher trainees and first-year teachers. Hammer et al. (1985) report a similar relationship ($r = .69$) between the MPSS and the SBS-HP in a sample of 116 employees from hospital emergency rooms. It is most likely that the exhaustion subscales that are included in all burnout measures are responsible for this positive relationship. Unfortunately, in both studies no correlations between the subscales of the various burnout measures were presented. However, such information is provided by Ackerley et al. (1988), who found in a sample of over 500 licensed psychologists that the PBI overinvolvement scale was most clearly related to MBI emotional exhaustion ($r = .45$), whereas PBI lack of control was negatively correlated with MBI personal accomplishment ($r = -.55$). The remaining two PBI scales (negative clientele and lack of support) were only weakly correlated with the MBI dimensions.

THE BURNOUT MEASURE (BM)

The second most widely employed burnout questionnaire after the MBI is the Burnout Measure (BM) (Pines & Aronson, 1988). According to the test authors, burnout is defined as "a state of physical, emotional and mental exhaustion caused by long-term involvement in situations that are emotionally demanding" (p. 9). In this current view, burnout is not restricted to certain professional groups. Originally, however, the authors had distinguished between burnout and tedium, which were considered to be similar in symptomatology but different in origin: "*Tedium* can be the result of any prolonged *chronic pressure* (mental, physical, or emotional); *burnout* is the result of repeated *emotional pressure* associated with an intense involvement with people over long periods of time" (Pines, Aronson, & Kafry, 1981, p. 15). In their recent work, Pines and Aronson (1988) abandoned this differentiation by broadening the burnout concept, which now also includes tedium. As a consequence, the former Tedium Measure is currently denoted Burnout Measure.

The BM consists of 21 items that express exhaustion and that are scored on a 7-point rating scale, ranging from "never" to "always." The authors present the BM as an instrument for the self-diagnosis of burnout and offer interpretations for test scores. Occasionally, the proposed norms have been used to classify people who are presumably at risk for developing burnout (Aström, Nilsson, Norberg, Sandman, & Winblad, 1991). However, such use

for individual assessment is not recommended because valid norms of the BM are lacking.

Nevertheless, the BM seems to be a useful research instrument, as is demonstrated by the psychometric findings obtained by the test authors in 30 different samples totaling 3,900 subjects (Pines et al., 1981, pp. 202–222). Seven years later, Pines and Aronson (1988, pp. 220–222) state that meanwhile more than 5,000 subjects with various backgrounds have completed the BM (e.g., human services, business and management, science, art, administration, technical, clerical, teaching, students). However, most of these subjects come from nonrandom samples that included persons who voluntarily participated in burn-out workshops. Accordingly, the possibility cannot be ruled out that selection effects influenced the psychometric results reported by the test authors.

The BM appears to be a reliable instrument with internal consistency coefficients α ranging from .91 to .93. This is not very surprising because many items can be considered synonyms (e.g., “feeling ‘burned out’” and “feeling rundown”) or antonyms (e.g., “being happy” and “being unhappy”). Thus, for reasons of economy, the length of the BM can be shortened by half of its items without negatively affecting its reliability (Schaufeli & Peeters, 1990). Test–retest reliabilities of the BM range from .89 to .66 across a 1- and 4-month interval, respectively (Pines & Aronson, 1988, p. 220).

BM scores have been found to be negatively related with satisfaction from work, from life, and from oneself, in 30 samples (Pines et al., 1981, p. 209). The mean correlations across these samples are $-.35$ (range: $-.31$ to $-.63$), $-.40$ (range: $-.37$ to $-.70$), and $-.50$ (range: $-.34$ to $-.73$), respectively. In a sample of 129 social workers, BM scores were correlated with an intention to leave the job ($r = .58$) (Pines & Aronson, 1988, p. 221). Also, it was shown in a study of 14 residential facilities for the developmentally disabled that mean turnover rates were higher in facilities with high BM scores than those with low BM scores (i.e., 49% against 17%) (Weinberg, Edwards, & Garove, 1983). In studies among 181 telephone operators and 298 police officers, BM scores were found to be positively correlated with poor self-reported physical health ($r = .48$) and with on-duty physical symptoms such as headaches, loss of appetite, nervousness, backaches, and stomach aches (r values ranged from .32 to .38) (Pines & Aronson, 1988, p. 221). Finally, burnout as measured by the BM is positively related to several work features in a number of samples (e.g., lack of social support, lack of autonomy, lack of feedback, and lack of variety). Typically, correlations with work features are modest, seldom exceeding values of .40 (Pines et al., 1981, pp. 213–218).

Despite the multidimensional definition of the burnout syndrome by the test authors, the BM is conceived as a one-dimensional questionnaire. Thus one can argue that the BM is not a proper operationalization of their definition of burnout. Therefore, it is not surprising that two factorial validity studies have failed to distinguish more than one burnout dimension in the BM (Corcoran, 1986;

Justice, Gold, & Klein, 1981). On the other hand, Enzmann and Kleiber (1989) found some weak indications for a three-factor structure of the BM in a sample of 130 German human services professionals. However, their three-factor structure (i.e., "demoralization," "exhaustion," and "loss of motive") did not correspond to the original components distinguished by the test authors. Recently, Schaufeli and Van Dierendonck (1992) replicated this finding using confirmatory factor analysis in a sample of 667 Dutch nurses.

Unfortunately, only a few studies exist in which other burnout measures, such as the MBI, have been employed simultaneously with the BM (Corcoran, 1986; Stout & Williams, 1983). The results of both of these studies (which have been conducted with 139 MA level social workers and 78 mental health workers, respectively) are quite comparable, however. Burnout as measured with the BM is strongly associated with MBI emotional exhaustion and MBI depersonalization ($.50 < r < .70$), and is somewhat less strongly but negatively associated with MBI personal accomplishment ($-.25 < r < -.30$).

It can be concluded that the BM is a reliable and valid research instrument that indicates the individual's level of exhaustion, which is considered to be the core element of the burnout syndrome (Shirom, 1989; see also Chapter 11).

THE MASLACH BURNOUT INVENTORY

As we have seen before, the MBI test authors Maslach and Jackson (1981a, 1986) described burnout as a three-dimensional syndrome characterized by emotional exhaustion, depersonalization, and reduced personal accomplishment. Moreover, in their view the burnout syndrome is restricted to professionals who work with recipients in some capacity. Accordingly, the MBI is meant for use in this particular occupational context.

Problems can arise when the MBI is employed outside the human services professions. In that case some items must be reworded. For example, Iwanicki and Schwab (1981) demonstrated in a sample of 469 teachers that the MBI can be slightly adapted without any problems for use among teachers by substituting "students" for "recipients." In that case, the meaning of the scales remains unchanged. In a special supplement to the test manual, the "MBI Form Ed" for measuring educator burnout is discussed by Richard Schwab (Maslach & Jackson, 1986, pp. 18-22). On the other hand, Garden (1987a) argued that substituting "coworkers" for "recipients" (as was done by Golembiewski, Munzenrider, & Stevenson, 1986) changes the meaning of the items. As a consequence, the depersonalization and personal accomplishment scales, in particular, have to be interpreted differently.

It is important to note that the three dimensions of burnout have not been deduced theoretically before the proper test construction of the MBI commenced. Instead, they were labeled *after* a factor analysis of an initial set of 47

items in a heterogeneous human services sample (see also Chapter 2). Accordingly, an inductive, rather than a deductive, approach was employed. Burisch (1984a) argued that the latter strategy renders less optimal psychometric results than the former strategy. Thus, despite the almost universal acceptance of the MBI, a careful psychometric evaluation is by no means superfluous, particularly when the inventory is applied in other occupational or cultural settings. To date, elaborate psychometric analyses of the MBI performed outside the United States are completely lacking. (See Chapter 13) Originally, the MBI consisted of 25 items, distributed across four scales: emotional exhaustion (9 items), depersonalization (5 items), personal accomplishment (8 items), and involvement (3 items) (Maslach & Jackson, 1981b). However, the latter scale was presented as an optional part of the MBI. In the second version of the test manual the involvement scale is no longer included (Maslach & Jackson, 1986).

Also, in the second version of the test manual the original double rating of each item has been abandoned. Initially, every item was scored twice on a frequency scale, ranging from "a few times a year" (1) to "every day" (6), and on an intensity scale, ranging from "very mild, barely noticeable" (1) to "very strong, major" (7). Because the frequency and intensity ratings appeared to be highly correlated ($r > .80$), only the frequency scoring is recommended in the latest version of the test manual.

The *internal consistency* of the three MBI scales is satisfactory, with Cronbach's α values ranging from .71 to .90 in the normative sample that includes over 11,000 subjects (Maslach & Jackson, 1986). Generally, similar α values have been found in other samples such as psychologists (Ackerley et al., 1988; Huberty & Huebner, 1988); human services professionals (Brookings et al., 1985); teachers (Belcastro, Gold, & Hays, 1983); social workers (Corcoran, 1986); nurses (Constable & Russell, 1986); and prison guards (Dignam & West, 1988; Lindquist & Whitehead, 1986). Occasionally, internal consistency coefficients lower than .70 are found, particularly with the depersonalization scale in non-human services samples, such as gifted students (Fimian, Fastenau, Tashner, & Cross, 1989) and university students (Gold, Bachelor, & Michael, 1989; Powers & Gose, 1986). As mentioned earlier, depersonalization items may have a different meaning in such samples. However, because poor internal consistency coefficients of the depersonalization scale are sometimes also found in human services samples (e.g., Leiter & Maslach, 1988), the scale's shorter length may also be a factor (because the value of the α coefficient depends on the number of scale items). A standardization for test length eliminates these systematic differences in reliabilities between the depersonalization scale and both other MBI dimensions (Schaufeli & Peeters, 1990).

Test-retest coefficients range from .60 to .80 across short periods up to a month (Maslach & Jackson, 1986). Dignam and West (1988) used a structural equation model in a sample of 200 prison guards and found a "true" autocorrelation of .80 of the composite emotional exhaustion and depersonalization score

across a 3-month interval. Two studies investigated the test-retest reliability spanning a period of a year. Stability coefficients for the scales ranged from .33 to .67 in a sample of 700 teachers (Jackson, Schwab, & Schuler, 1986) and from .34 to .62 in a sample of 46 human services professionals (Wade, Cooley, & Savicki, 1986). In all studies, emotional exhaustion appeared to be the most stable burnout dimension, whereas depersonalization was the least stable dimension. These high correlations suggest that burnout is a chronic rather than a transient state of mind.

Because validity is a generic concept, three aspects are briefly discussed: factorial validity, convergent validity, and discriminant validity. The *factorial validity* of the MBI is not completely beyond question. Although the three-dimensional structure of the MBI has been confirmed in several studies (Belcastro et al., 1983; Gold, 1984; Huberty & Huebner, 1988; Fimian & Blanton, 1987; Fimian et al., 1989; Green & Walkey, 1988; Koeske & Koeske, 1989; Lahoz & Mason, 1989; Pierce & Molloy, 1989), other researchers have found two dimensions (Brookings et al., 1985; Dignam, Barrera, & West, 1986; Green, Walkey, & Taylor, 1991), and even four dimensions (Firth, McIntee, McKeown, & Britton, 1985; Iwanicki & Schwab, 1981; Powers & Gose, 1986).

Thus far, virtually no adequate attempts have been made to test the factorial validity of the MBI, by carrying out confirmatory factor analysis. Although Golembiewski and Munzenrider (1988, pp. 19-23) claim that the results of their confirmatory factor analyses support the three-factor structure of the MBI, they provide insufficient information about the rather unusual method they used, i.e., Ahmavaara's technique (see Chapter 14). Moreover, Golembiewski and Munzenrider employed a strongly modified version of the MBI. Four recent studies examined the dimensionality through confirmatory factor analysis with LISREL and found the fit of the original three-factor model to be superior to several alternative models (Gold, et al., 1989; Lee & Ashforth, 1990; Byrne, Schaufeli, & Van Dierendonck, 1992).

Most studies on the *convergent validity* of the MBI have been discussed above. Generally, these studies yield positive results, indicating that to a certain extent the MBI scales measure the same construct as do other burnout instruments such as the BM, PBI, SBS-HP, and MBA. Roughly speaking, these measures share about 25% of their variance. As was concluded before, emotional exhaustion is the best validated dimension of the burnout syndrome. Unfortunately, the relationships with self-ratings, peer ratings, and expert ratings of MBI burnout are less strong (Meier, 1984; Rafferty et al., 1986). For instance, correlations between peer ratings of burnout and MBI scale scores range from .20 to .56 in various samples of social service workers ($N = 91$), physicians ($N = 43$), nurses and social service workers ($N = 180$), and police officers and their spouses ($N = 142$) (Maslach & Jackson, 1986). Generally, only about 10% of the variance of the MBI is shared with information from

external sources, and again emotional exhaustion appears to be the best validated MBI dimension.

There is an evident discrepancy between these results and those found in validation studies that rely exclusively on self-report measures. It is likely that at least some portion of the common variance between self-report measures of burnout is caused by method variance and can be considered an artifact. Thus we are left with the conclusion that, except for the emotional exhaustion scale, the convergent validity of the MBI has not yet been demonstrated convincingly.

Well-designed investigations that assess the *discriminant validity* of the MBI are very rare. Employing a multitrait-multimethod approach, Meier (1984) concluded from a study of 320 faculty members that considerable overlap exists between the MBI and several measures of depression. Unfortunately, he used a total MBI score that summed responses across all items (an unorthodox procedure that was criticized by the test authors). However, Firth et al. (1986b) scored the MBI in the usual way and arrived at an interesting conclusion in their study of 200 nurses. Emotional exhaustion was substantially related to depression ($r = .50$), as measured by Beck's Depression Inventory. Relationships with personal accomplishment ($r = -.17$) and with depersonalization ($r = .32$) were much lower. Similar results were obtained by Landsbergis (1988), who used a less well-known depression scale.

Although in the MBI test manual only weak and insignificant relationships with job satisfaction are reported, several studies show much stronger associations (e.g., Dolan, 1987; Eisenstat & Felner, 1984; Koeske & Koeske, 1989; Landsbergis, 1988; Lindquist & Whitehead, 1986; Penn, Romano, & Foat, 1988; Stout & Williams, 1983). The results from these studies are somewhat consistent, indicating that emotional exhaustion has moderately negative correlations with job satisfaction (coefficients ranging between .35 and .45). Depersonalization is only slightly negatively correlated (coefficients ranging between .25 and .35), whereas personal accomplishment is positively but insignificantly related to job satisfaction.

Finally, some remarks have to be made about the classification of levels of burnout as presented in the test manual. This classification is based on arbitrary statistical norms. The normative sample has been arbitrarily divided into three equally sized groups of 33.3%, assuming that one third of the subjects will experience a high degree of burnout, another third will experience an average level of burnout, and the final third will experience a low level of burnout. However, as Einsiedel and Tully (1982) indicated, no clinically valid reason exists for using the top-third/bottom-third split as the dividing line instead of using, say, the median or the mean. Neither of these cutoff points has been empirically validated, by employing independent expert ratings from clinical psychologists or psychiatrists.

Maslach and Jackson note, correctly, that the categorization of burnout into

three levels "is intended primarily as feedback for individual respondents" (1986, p. 5). Moreover, they explicitly warn that "neither the coding nor the original numerical scores should be used for diagnostic purposes; there is insufficient research on the pattern(s) of scores as indicators of individual dysfunction or the need for intervention." Researchers are directed to use the original numerical scores rather than the categorizations of low, average, and high for statistical analyses. Despite this clear statement, there is a tendency for some researchers to use a particular cutoff point (usually the top third) to differentiate burnout cases from noncases (e.g., Ackerley et al., 1988; Belcastro et al., 1983; Birch, Marchant, & Smith, 1986; Firth et al., 1985; Firth & Britton, 1989; Lindquist & Whitehead, 1986; McGrath, Reid, & Boor, 1989; Lahoz & Mason, 1989; Penn et al., 1988; Turnipseed, 1987; Ursprung, 1986). They then compute the proportion of burned-out persons in their particular sample. With only one exception (Ackerley et al., 1988), less than one third of the respondents are classified as high in emotional exhaustion or in depersonalization. Obviously, the subject populations do not represent the full range of those experiencing burnout, and they have other more skewed distributions than does the normative sample. This is particularly true for personal accomplishment. The percentage of respondents classified as low on this burnout dimension ranges between 1% (Ackerley et al., 1988) and 97% (McGrath et al., 1989)! Such remarkable differences underscore the argument of the test authors not to use the burnout categories but to look at the actual scores instead. Furthermore, no empirical basis exists yet for employing the MBI as a diagnostic tool for individual assessment.

The conclusion concerning the psychometric quality of the MBI is somewhat inconsistent. On the one hand, the factorial and the convergent validity as well as the reliability of the instrument is quite encouraging. On the other hand, burnout as measured with the MBI cannot be validly distinguished from related concepts such as depression and (to a somewhat lesser degree) job satisfaction. It is noteworthy that the scale that is most robust and reliable and that displays the strongest convergent validity (i.e., emotional exhaustion) is the least specific dimension of burnout, showing considerable overlap with related constructs.

CONCLUSIONS AND RECOMMENDATIONS

Our review of the literature shows that during the past decade, much progress has been made in the measurement of burnout. In this closing section, seven tentative conclusions are drawn.

1. The vast majority of instruments that assess burnout are self-report measures. This causes a particular problem in validation studies that use self-

report measures exclusively because at least part of the common variance of the measures has to be attributed to method variance.

2. Most instruments are designed to assess levels of burnout in human services professions. They should not be applied in other occupational contexts because it cannot be assumed that the structure of the burnout syndrome is identical across different occupational groups (see Chapter 1).
3. Two self-report instruments have been intensively studied psychometrically (i.e., BM and MBI). Unfortunately, little information is available about the psychometric properties of the remaining self-report instruments. Mostly, their authors only report satisfactory internal consistencies.
4. Although the psychometric qualities of the BM are promising, its use is rather limited because of the unidimensionality of the measure. The BM reduces a complex psychological phenomenon to mere exhaustion.
5. The psychometric qualities of the MBI (the most widely used self-report instrument) are encouraging but not completely beyond question. In particular, its discriminant validity is rather poor.
6. In most self-report inventories the individual's depletion of emotional resources is included in one way or another. Paradoxically, this crucial dimension of burnout also seems to be its least specific component (see also Chapter 2).
7. Despite the obvious agreement on the core meaning of burnout (i.e., exhaustion of a person's resources), considerable confusion exists about the number, as well as the nature, of other dimensions involved. For instance, some instruments not only assess burnout on the individual level but include organizational aspects as well.

DIRECTIONS FOR FUTURE RESEARCH

In the remainder of this chapter six directions for future research are discussed.

Development of Alternative Burnout Instruments

Currently, burnout is almost exclusively assessed with self-report questionnaires. Occasionally, self-assessment with rating scales or peer ratings is employed. Additionally, standardized interviews should be developed as well as behaviorally anchored rating scales that can be assessed by supervisors and colleagues. Such alternative assessment methods can be cross-validated with the MBI because this currently seems to be the most promising self-report questionnaire to measure burnout. Moreover, specific burnout instruments have to be developed that can be applied outside the human services professions. This means that, in particular, the depersonalization dimension has to be reconsidered. Basically, depersonalization is a specific form of psychological withdrawal from the essence of one's job. In human services professions, recipients

constitute the core element of the job. In management, however, the core element is the organization. It is likely that burned-out managers withdraw psychologically from the organization by developing negative attitudes and behaviors toward the organization and its members. Therefore, a burnout measure for managers should include such particular attitudes and behaviors.

Psychometric Improvement of the MBI

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 regarding the behavioral element of depersonalization, would not only increase the internal consistency (cf. Golembiewski & Munzenrider, 1988, pp. 19-21) but also strengthen the validity of this scale. 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. Finally, factorial validity studies suggest that 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 (Koeske & Koeske, 1989; Mor & Laliberte, 1984; Schaufeli & Van Dierendonck, 1992). Inspection of the MBI manual (Maslach & Jackson, 1986, p. 30) reveals that this item is the weakest and most complex item in the accomplishment scale.

Individual Assessment of Burnout

A great need exists for instruments that can assess burnout on an individual level. The MBI could serve this purpose, but the existing categorization into upper, middle, and lower thirds is arbitrary and lacks any clinical validity. Therefore, this categorization is inappropriate for individual assessment (as is rightfully stressed in the test manual). What is needed are studies that establish clinically valid cutoff points by comparing burnout scores with expert ratings of psychiatrists or psychologists. Such a validation procedure can be difficult, as demonstrated by the discriminant validity study of Rafferty et al. (1986) that was reviewed in the first part of this chapter. The study found rather weak relationships between the MBI scores and expert ratings, possibly because the researchers did not clearly specify the burnout syndrome to the experts. Furthermore, appropriate group norms should be developed based on representative, stratified random samples of different professions. Many researchers employ the "thirds" categorization of MBI scores in order to estimate the number

of burned-out cases in a particular profession. Unfortunately, not only is the use of this categorization invalid, as was previously shown, but the samples are usually inappropriate. Only the few studies that employ appropriate samples, allow for generalizations to the professional group as a whole and thus can be used to develop valid group norms. These notable examples are studies of social workers (Himle, Jayaratne, & Thyness, 1986), prison guards (Lindquist & Whitehead, 1986), librarians (Birch et al., 1986), and psychologists (Ackersley et al., 1988).

Epidemiological Research

In developing adequate group norms along the lines outlined above, one actually studies the epidemiology of burnout in different professions. With the MBI we have a reasonably reliable and valid instrument to assess burnout in human services professions. There is a great need for epidemiological knowledge of burnout in order to identify specific (sub)groups at risk. For instance, virtually no studies exist that provide the sort of information that would be necessary for planning appropriate interventions. To date, most epidemiological studies in the occupational field have used the General Health Questionnaire (Goldberg, 1978) to assess minor psychiatric disorder. However, this instrument has been criticized because it is somewhat unspecific (Fletcher, 1988; Warr, 1990). The advantage of the MBI is its twofold specificity concerning the domain (three aspects of burnout) as well as the occupational group (human services professions).

Cross-National Psychometric Research

The growing popularity of the MBI outside English-speaking countries requires a thorough psychometric evaluation of this instrument in each specific national context. With only two exceptions (Schaufeli & Van Dierendonck, 1992; Schaufeli & Janczur, in press), such studies have only been published in local languages (e.g., Enzmann & Kleiber, 1989; Gil-Monte & Schaufeli, 1992; Girault, 1989; Sirigatti, Stefanile, & Menoni, 1988). Although the results in Chapter 13 suggest that the MBI can be used cross-nationally, valid research that compares levels of burnout between countries is virtually lacking. To date no attempts have been made to calibrate the MBI for different nations, yet such calibration is a prerequisite for using the instrument for diagnostic purposes.

Development of Conceptual Models of Burnout

Finally, and most important, research with the MBI should generate and test models about the etiology and the persistence of burnout in the human services. Strictly speaking, this is not a psychometric endeavor; as was pointed out be-

fore, assessment of the discriminant validity of the MBI is only possible within an elaborate theoretical framework. Because of its multidimensional conceptualization of burnout, the MBI seems to be particularly appropriate to employ in theory-driven research. The fruitfulness of such an approach is demonstrated by several authors who contributed to this volume (Buunk & Schaufeli, Golembiewski et al., Maslach, and Leiter). We strongly believe that, despite its minor weaknesses, the MBI has a promising future in extending our knowledge of professional burnout.