

A CAUTIONARY NOTE ABOUT THE CROSS-NATIONAL AND CLINICAL VALIDITY OF CUT-OFF POINTS FOR THE MASLACH BURNOUT INVENTORY¹

WILMAR B. SCHAUFELI AND DIRK VAN DIERENDONCK

Utrecht University

Summary.—In the present study, burnout scores of three samples, as measured with the Maslach Burnout Inventory, were compared: (1) the normative American sample from the test-manual ($N=10,067$), (2) the normative Dutch sample ($N=3,892$), and (3) a Dutch outpatient sample ($N=142$). Generally, the highest burnout scores were found for the outpatient sample, followed by the American and Dutch normative samples, respectively. Slightly different patterns were noted for each of the three components. Probably sampling bias, i.e., the healthy worker effect, or cultural value patterns, i.e., femininity versus masculinity, might be responsible for the results. It is concluded that extreme caution is required when cut-off points are used to classify individuals by burnout scores; only nation-specific and clinically derived cut-off points should be employed.

The Maslach Burnout Inventory (Maslach & Jackson, 1986) is the most widely used self-report instrument to measure burnout. Generally speaking, its psychometric features are quite encouraging (for a review see Schaufeli, Enzmann, & Girault, 1993). For instance, the three-factor structure, i.e., Emotional Exhaustion—EE, Depersonalization—DP, and reduced Personal Accomplishment—PA, has been confirmed in various samples. Although the discriminant validity is not entirely beyond question, the internal consistencies of the three subscales are satisfactory as well as the convergent validity with other burnout measures. Occasionally, a composite burnout-score is employed by adding the individuals' scores on the three factors. However, Maslach (1993) strongly advised against the use of such a unitary measure since empirical evidence suggests that burnout is a multidimensional construct.

Increasingly, the Maslach Burnout Inventory is employed in non-English-speaking countries (for a review see Golembiewski, Scherb, & Boudreau, 1993). The psychometric qualities of the non-English versions are quite similar to those of the original version. At least, this was shown for the Dutch (Schaufeli & Van Dierendonck, 1993), German (Büssink & Perrar, 1992), and French (Dion & Tessier, 1994) versions. Hence, from a psychometric point of view the inventory seems to be cross-nationally robust. But

¹Address enquiries to either author, Department of Psychology, Utrecht University, P.O. Box 80.140, 3508 TC Utrecht, The Netherlands.

what about levels of burnout? Are there systematic differences in levels of burnout across countries? The first aim of the present study is to compare the scores of the two largest normative samples currently available (USA and The Netherlands).

The classification of levels of burnout remains an unresolved issue in the English version as well as in the non-English versions. As, at least in some countries, the recognition and acceptance of burnout as a work-related disorder seem to grow, valid criteria are needed to classify levels of burnout, for instance in case of insurance claims. In addition, cut-off points might be helpful in deciding which kind of treatment programme is appropriate. However, as long as no valid cut-off points are available which allow differentiation between levels of burnout, the inventory cannot be employed as an individual assessment tool for clinical diagnosis. The test-manual only presents numerical cut-off points based on arbitrary statistical norms. The test-authors divided the normative sample into three equally sized groups of 33.3%, assuming that the top, intermediate, and bottom thirds of the sample would experience "high," "average," and "low" levels of burnout, respectively (Maslach & Jackson, 1986).

Although the test-authors correctly warn that this classification should not be used for diagnostic purposes, there is a strong tendency for many researchers to differentiate burned-out subjects from nonburned-out subjects. This is not only observed in the USA (e.g., Ackerley, Burnell, Holder, & Kurdek, 1988; Poulin & Walter, 1993) but also in Australia (Bennett, Kelaher, & Ross, 1994), Greece (Papadatou, Anagnostopoulos, & Monos, 1994), Spain (Gil-Monte, Valcárcel, & Zornoza, 1993), and Britain (Firth & Britton, 1989). Clearly, this is an erroneous approach, not only since USA cut-off points are expected to be invalid in other countries, but also because these cut-off points have not been clinically derived, for instance, in samples of employees who seek treatment for their work-related mental problems. Therefore, it is quite likely that by employing the classification of burnout levels from the USA test-manual the number of burnout cases is *overestimated*. This is due to the so-called "healthy worker effect" (Karasek & Theorell, 1990): the systematic *underestimation* of levels of burnout because exclusively working, and so more or less healthy, subjects have been included in the normative sample. For instance, those who have left the organization because of burnout have *not* been investigated. Hence, levels of burnout in normative working samples are likely to be relatively low. As far as the second aim of the current study is concerned, we expect to demonstrate this healthy worker effect by comparing burnout levels from the Dutch normative sample with levels of an outpatient sample from that country. It was hypothesized that levels of burnout are higher in the outpatient sample compared to the normative sample.

METHOD

The normative Dutch sample consists of 3,892 human services professionals, e.g., nurses, physicians, correctional officers, hospice staff. Fifty-four percent were women; the mean age of the sample is 34.2 (SD , 8.4) (Schaufeli & Van Dierendonck, 1994). The outpatient sample consists of 142 patients, e.g., health-care professionals, educators, police officers, who have been referred to a psychotherapeutic treatment center because of work-related mental problems. Since to date no valid diagnostic criteria for burnout exist, only exclusion criteria were used; outpatients with major mental illness, i.e., psychosis or personality disorder, were excluded. The majority (66%) are men; the mean age of the sample is 41.9 yr. (SD , 8.9). Most outpatients (64%) are on sick-leave, which has lasted on the average 4.8 mo. (SD , 8.3) prior to referral.

Although in the Dutch version of the inventory two weak items (Nos. 12 and 16) have been eliminated (Schaufeli & Van Dierendonck, 1993, 1994), the 22-item version is used in the present study for reasons of compatibility with the original version. The internal consistencies are satisfactory for the Dutch normative sample (Emotional Exhaustion $\alpha = .88$, Depersonalization $\alpha = .70$, Personal Accomplishment $\alpha = .80$) as well as for the outpatient sample (Emotional Exhaustion $\alpha = .89$, Depersonalization $\alpha = .68$, Personal Accomplishment $\alpha = .74$). The scoring-dimension was anchored by never (0) and every day (6). High Emotional Exhaustion and Depersonalization scores and low Personal Accomplishment scores are indicative for burnout.

RESULTS

Table 1 shows the mean scores on the burnout scale and the cut-off points for the prior American and Dutch normative samples as well as of the present Dutch outpatient sample. A procedure similar to that described in the test-manual (see above) was followed in order to categorize the Dutch samples as having "high," "average," and "low" scores on the burnout inventory.

As can be seen from Table 1, mean scores of Emotional Exhaustion ($t_{14957} = -19.96$, $p < .001$), Depersonalization ($t_{14957} = -14.42$, $p < .001$), and Personal Accomplishment ($t_{14957} = -28.74$, $p < .001$) are significantly lower for the Dutch normative sample than for the American normative sample. Accordingly, relative to the American subjects, the Dutch subjects show *lower* scores on two dimensions (Emotional Exhaustion and Depersonalization) but *higher* scores on the third dimension (reduced Personal Accomplishment). As a consequence, different cut-off points emerge in the Dutch normative sample (see Table 1). Had the American cut-off points been applied to the Dutch sample, the number of subjects with "high" scores on Emotional Exhaustion and Depersonalization would have been severely un-

derestimated (both 13% instead of about 33%), whereas those with "low" scores on Personal Accomplishment would have been likewise *overestimated* (55% instead of about 33%).

TABLE 1
COMPARISON OF SCORES OF DUTCH NORMATIVE ($N=3,892$), DUTCH OUTPATIENT ($N=142$), AND AMERICAN NORMATIVE ($N=11,067$) SAMPLES: CUT-OFF POINTS, MEANS, AND STANDARD DEVIATIONS

Cut-off Points	Dutch Normative Sample			Dutch Outpatient Sample			American Normative Sample		
	EE	DP	PA	EE	DP	PA	EE	DP	PA
High	≥20	≥8	≤28	≥34	≥12	≤25	≥27	≥13	≤31
Average	13-20	5-8	29-32	26-33	6-11	26-28	17-26	7-12	38-32
Low	≤12	≤5	≥33	≤25	≤7	≥29	≤16	≤6	≥39
<i>M</i>	17.86	7.35	30.95	28.64	9.27	27.05	20.30	8.73	34.58
<i>SD</i>	8.50	4.29	5.72	10.13	4.89	5.65	10.75	5.89	7.11

Note.—EE = Emotional Exhaustion; DP = Depersonalization; PA = Personal Accomplishment.

As hypothesized, mean scores on Emotional Exhaustion ($t_{4032}=15.65$, $p<.001$) and Depersonalization ($t_{4032}=5.21$, $p<.001$) are significantly higher in the Dutch outpatient sample than in the normative sample of that country, whereas scores on Personal Accomplishment are significantly lower ($t_{4032}=-15.65$, $p<.001$). A similar pattern is observed when the outpatient sample is compared with the American normative sample, at least as far as Emotional Exhaustion ($t_{11207}=8.43$, $p<.001$) and Personal Accomplishment scores ($t_{11207}=-7.98$, $p<.001$) are concerned. Quite remarkably, *no* significant differences were found in scores on Depersonalization ($t_{11207}=1.09$, *n.s.*). Had the clinically derived cut-off points of the outpatient sample been applied to the Dutch normative sample, the percentage of burnout cases would have been substantially lower than about 33%: 2.6%, 15.5%, and 13.7% for Emotional Exhaustion, Depersonalization, and Personal Accomplishment, respectively. In other words, and in accordance with the healthy worker effect, employing the cut-off points from the normative working sample instead of the clinically derived cut-off points from the outpatient sample, leads to an overestimation of "real" burnout cases of about 30% for Emotional Exhaustion and 18 to 20% for Depersonalization and Personal Accomplishment.

DISCUSSION

The first major finding is that scores on burnout differ significantly between the American and Dutch normative samples. American subjects have higher mean scores on Emotional Exhaustion and Depersonalization but report *less* feelings of reduced Personal Accomplishment than Dutch subjects. This finding underlines the importance of the distinction between the three aspects of burnout as well as the necessity of nation-specific cut-off points.

As was illustrated above, the use of American norms would lead to flagrant misinterpretations of the scores on burnout in Dutch samples.

One can only speculate about the causes of these cross-national differences in scores on burnout. It cannot be ruled out that sampling bias and response tendencies play a significant role. For instance, both national samples are not representative as far as gender, occupation, work experience, region, and so forth are concerned. Further, the fact that the variances of all scales are significantly larger for the American than for the Dutch normative sample (Emotional Exhaustion: $F_{15047} = 1.60$, $p < .001$; Depersonalization: $F_{15047} = 1.87$, $p < .001$; Personal Accomplishment: $F_{15047} = 1.87$, $p < .001$) might point to a different response set; Americans may respond more extremely than the Dutch. Such differences in response styles are occasionally observed in cross-cultural research (Hui & Triandis, 1989). Alternatively, it can be speculated that North American jobs are more stressful than similar jobs in Europe. For instance, as a rule, scores on burnout (particularly Emotional Exhaustion and Depersonalization) among nurses in North American samples (e.g., Brown-Ceslowitz, 1989; Dion & Tessier, 1994; Hare, Pratt, & Andrews, 1988) tend to be higher than those reported in European countries such as France (Girault, 1989), Britain (Firth, McIntee, McKeown, & Britton, 1985), Germany (Enzmann & Kleiber, 1989), Italy (Pedrabissi & Santinello, 1988), and Ireland (Dolan, 1987). Only in Poland have similar burnout scores like those in North America been observed (Schaufeli & Janczur, 1994). It seems unlikely that the translation of the inventory could be held responsible for these differences, since lower burnout scores were found in English-speaking European countries (Britain and Ireland) and higher scores on burnout were found in a French-speaking Canadian sample (Dion & Tessier, 1994). Finally, it can be speculated that American society is more achievement-oriented than most European societies. Given the greater emphasis on individual achievement in the USA, more casualties are expected there in terms of burnout than in a society wherein one is less likely to see career success equated with personal success such as in The Netherlands.

The second major finding is that, as hypothesized, scores on burnout are significantly higher in an outpatient sample than in Dutch or American working samples. It should be noted, however, that our outpatient sample is rather heterogeneous since it consists of different kinds of work-related mental problems. Since no formal diagnostic criteria exist which allow a positive identification of burnout cases, only patients with major mental illness were excluded. Nevertheless, our findings seem to support the existence of a healthy worker effect. When the clinically derived cut-off points from the outpatient sample are applied to the Dutch normative sample, the number of "real" burnout cases drops dramatically, particularly for Emotional Exhaustion. The fact that scores on Emotional Exhaustion are especially high

in the outpatient sample agrees with the observation of Pick and Leiter (1991) who found elevated Emotional Exhaustion scores, but similar Depersonalization and Personal Accomplishment scores in nurses who considered themselves "burned out" compared to those who found they were "coping well." Accordingly, it seems that feelings of emotional exhaustion play a crucial role in the individual's self-perception of "burnout" and might thus, indirectly, encourage the individual to seek professional help.

Quite interestingly, *no* significant differences were found in levels of Depersonalization between the outpatient sample and the American normative sample. Apparently, a Dutch outpatient's score on Depersonalization is quite "normal" in the USA. It can be speculated that this reflects cross-cultural differences. According to Hofstede's (1980) culture dimensions, a feminine value pattern prevails in The Netherlands, whereas in the USA a stronger masculine orientation prevails. Typically, the depersonalized attitude, that is, characterized by objectivity, distance, and instrumentality, is considered to be associated with the male sex-role socialization. Indeed, as a rule, men show higher levels of Depersonalization than women (Greenglass, 1991). Thus the relatively high Depersonalization scores in the American sample might reflect the masculine orientation of that society.

In the present study we had to follow the test-manual by using the top/middle/bottom thirds as a way of distinguishing levels of burnout. However, we prefer percentile ranges as they are being used in the Dutch version of the inventory (Schaufeli & Van Dierendonck, 1994). Percentile ranges have greater flexibility in allowing the delineation of individuals with unusual or extreme values.

In conclusion, our study illustrates that one should be extremely cautious when using cut-off points for the classification of burnout levels. It does not make sense to use cut-off points that are obtained in one country to classify subjects in another country. And, it does not make sense to use cut-off points that are obtained for a "normal" working sample to classify subjects as clinically "burned out." Yet, by using clinically derived, nation-specific cut-off points we can learn more about the incidence of burnout in working populations. Moreover, cross-national research on burnout can clarify this construct and at the same time provide greater insight into the relationship between cultural norms and stress.

REFERENCES

- ACKERLEY, G. D., BURNELL, J., HOLDER, D. C., & KURDEK, L. A. (1988) Burnout among licenced psychologists. *Professional Psychology: Research and Practice*, 19, 624-631.
- BENNETT, L., KELAHER, M., & ROSS, M. (1994) Quality of life in health care professionals: burnout and its associated factors in HIV/AIDS related care. *Psychology and Health*, 9, 273-383.
- BROWN-CESLOWITZ, S. (1989) Burnout and coping strategies among hospital staff nurses. *Journal of Advanced Nursing*, 14, 553-557.

- BÜSSINK, A., & PERRAR, K. M. (1992) Die Messung von Burnout: Untersuchung einer deutschen Fassung des Maslach Burnout Inventory (MBI-D) [The measurement of burnout: the German version of the Maslach Burnout Inventory]. *Diagnostica*, 38, 328-353.
- DION, G., & TESSIER, R. (1994) Validation de la traduction de l'Inventaire d'épuisement professionnel de Maslach et Jackson [Validation of the translation of the burnout questionnaire of Maslach and Jackson]. *Canadian Journal of Behavioural Science*, 26, 210-227.
- DOLAN, A. (1987) The relationship between burnout and job satisfaction in nurses. *Journal of Advanced Nursing*, 12, 3-12.
- ENZMANN, D., & KLEIBER, D. (1989) *Helfer-Leiden: Stress und Burnout in psychosozialen Berufen* [Helper-ordeals: stress and burnout in the human services]. Heidelberg: Asanger.
- FIRTH, H., & BRITTON, P. (1989) 'Burnout,' absence and turnover amongst British nursing staff. *Journal of Occupational Psychology*, 62, 55-60.
- FIRTH, H., MCINTEE, J., MCKEOWN, P., & BRITTON, P. (1985) Maslach Burnout Inventory: factor structure and norms for British nursing staff. *Psychological Reports*, 57, 147-150.
- GIL-MONTE, P. R., VALCÁRCEL, P., & ZORNOZA, A. (1993) Role stress: burnout antecedents in nursing professionals. *European Work and Organizational Psychologist*, 3, 217-227.
- GIRAULT, N. (1989) Burnout: emergence et stratégies d'adaptation [Burnout: emergence and adaptive strategies]. Unpublished doctoral thesis, Université René Descartes, Paris.
- 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*. Washington, DC: Taylor & Francis. Pp. 218-236.
- GREENGLASS, E. (1991) Burnout and gender: theoretical and organizational implications. *Canadian Psychology*, 32, 562-572.
- HARE, J., PRATT, C. C., & ANDREWS, D. (1988) Predictors of burnout in professional and para-professional nurses working in hospitals and nursing homes. *International Journal of Nursing Studies*, 25, 105-115.
- HOFSTEDE, G. (1980) *Cultures' consequences: international differences in work-related values*. Beverly Hills, CA: Sage.
- HUI, C. H., & TRIANDIS, H. C. (1989) Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20, 296-309.
- KARASEK, R., & THEORELL, T. (1990) *Healthy work*. New York: Basic Books.
- MASLACH, C. (1993) Burnout: a multidimensional perspective. In W. B. Schaufeli, C. Maslach, & T. Marek (Eds.), *Professional burnout: recent developments in theory and research*. Washington, DC: Taylor & Francis. Pp. 19-32.
- MASLACH, C., & JACKSON, S. E. (1986) *MBI: Maslach Burnout Inventory: manual research edition*. Palo Alto, CA: Consulting Psychologists Press. (1st ed. 1981)
- PAPADATOU, D., ANAGNOSTOPOULOS, F., & MONOS, D. (1994) Factors contributing to the development of burnout in oncology nursing. *British Journal of Medical Psychology*, 67, 187-199.
- PEDRABISSI, L., & SANTINELLO, M. (1988) Professione infermieristica e sindrome del burnout: Un contributo alla tartura del Maslach Burnout Inventory [The burnout syndrome in the nursing profession: a contribution to the calibration of the Maslach Burnout Inventory]. *Bolletino di Psicologia Applicata*, 187, 41-46.
- PICK, D., & LEITER, M. P. (1991) Nurses' perceptions of burnout: a comparison of self-reports and standardized measures. *The Canadian Journal of Nursing Research*, 23, 33-48.
- POULIN, J., & WALTER, C. (1993) Social worker burnout: a longitudinal study. *Social Work Research & Abstracts*, 29, 5-11.
- 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*. Washington, DC: Taylor & Francis. Pp. 199-215.
- SCHAUFELI, W. B., & JANCZUR, B. (1994) Burnout among nurses: a Polish-Dutch comparison. *Journal of Cross-Cultural Psychology*, 25, 95-113.
- SCHAUFELI, W. B., & VAN DIERENDONCK, D. (1993) The construct validity of two burnout measures. *Journal of Organizational Behavior*, 14, 631-647.
- SCHAUFELI, W. B., & VAN DIERENDONCK, D. (1994) Burnout, een concept gemeten: de Nederlandse versie van de Maslach Burnout Inventory (MBI-NL) [Burnout, the measure-

ment of a construct: the Dutch version of the Maslach Burnout Inventory]. *Gedrag & Gezondheid*, 22, 153-172.

Accepted May 3, 1995.