Stress and Health

Stress and Health **23**: 87–91 (2007)
Published online 1 December 2006 in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/smi.1124
Received 25 February 2006; Accepted 3 August 2006

The factorial validity and measurement invariance of the Maslach Burnout Inventory for human services

Stijn Vanheule*,†, Yves Rosseel and Peter Vlerick

Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium

Summary
This paper first examines the factorial validity of the Maslach Burnout Inventory for human services professions (MBI-HSS). The authors test the original 22-item version and a shortened 20-item version. Model fit is evaluated with confirmatory factor analysis starting with data collected from a sample of hospital nurses (n = 2515) and a sample of nurses and assistants working in residential welfare institutions (n = 1639). Only the fit of the 20-item MBI-HSS was good in both samples. Next the hypothesis of measurement invariance of the 20-item MBI-HSS across the samples is examined, but this hypothesis does not hold. Partial invariance at the level of the subscales also could not be confirmed. The authors conclude that for hospital nurses the MBI-HSS factors mean something other than what they mean for professionals working in welfare institutions, and suggest that future research could focus upon this difference. Copyright © 2006 John Wiley & Sons, Ltd.

Key Words
Burnout; factorial validity; measurement invariance

Introduction

The Maslach Burnout Inventory (MBI) is a commonly used instrument to measure professional burnout worldwide. The scale was constructed by Maslach and Jackson (1981), and contains 22 items that represent three factors: emotional exhaustion (nine items), depersonalization (five items), and reduced personal accomplishment (eight items). In this paper we study the factorial validity of the version for human services professions (MBI-HSS) (Maslach, Jackson, & Leiter, 1996).

During the 1980s and the early 1990s, most studies on the factorial validity of the MBI-HSS or the original MBI—to which the MBI-HSS corresponds—started from exploratory factor analysis (EFA) and principal components analysis (PCA). Some studies observed the three-factorial structure of the MBI that Maslach and colleagues first described, but others favored a factor structure with two or four dimensions (see Schaufeli & Enzmann, 1998). A limitation of EFA and PCA studies is that they are well suited for exploratory purposes, but not for testing hypothesized factor

* Correspondence to: S. Vanheule, Ghent University, Faculty of Psychology and Educational Sciences, H. Dunantlaan 2, B-9000 Ghent, Belgium.
† E-mail: Stijn.Vanheule@UGent.be

structures (see Kalliath, O'Driscoll, Gillespie, & Bluedorn, 2000). For the latter, confirmatory factor analysis (CFA) is more appropriate. CFA tests whether and how well hypothetical models of the mutual relations between items and factors fit empirical data (Hoyle, 1995).

From the 1990s onward CFA has been increasingly used for testing the factor structure of the MBI-HSS. A recurrent finding is that the model fit for the original 22-item MBI-HSS (MBI-HSS-22) is poor (e.g. Byrne, 1994; Kalliath et al., 2000). Some authors preferred to overcome this problem by accepting correlations between the residual variances in the model and by allowing items to load on several factors in the tested model (e.g. Byrne, 1994). A problem of both approaches is that they undermine unidimensional measurement of concepts. Moreover, very few theoretical arguments justify these types of solutions (see Kalliath et al., 2000). Removing items that cause misfit in a hypothesized factor is a more straightforward strategy, but only makes sense if the model thus obtained is further validated.

Several shortened models have been described in the literature. The one that most frequently recurs leaves out items 12 and 16, and several studies have observed acceptable-to-good fit for this 20-item MBI-HSS (MBI-HSS-20) (e.g. Gil-Monte, 2005; Hallberg & Sverke, 2004; Schaufeli & Van Dierendonck, 1993). Item 12 is a personal accomplishment item for which significant loadings on emotional exhaustion were consistently observed, and item 16 is an emotional exhaustion item for which significant loadings on depersonalization were consistently observed. Schaufeli and colleagues (Schaufeli, Bakker, Hoogduin, Schaap, & Kladler, 2001) demonstrated the clinical validity of the MBI-HSS-20. They calculated cut-off scores that discriminate between employees with work-related neurasthenia and those without it.

What is interesting about the MBI-HSS-20 is that the same problematic items were identified by different groups of researchers, and that this version does not differ substantially from the MBI-HSS-22. Both versions are considered equivalent (see Schaufeli & Enzmann, 1998). An important construct validity-related question that remains open is whether this model can be replicated in large samples. Moreover measurement invariance across different human services professions or sectors has not yet been studied for this model.

Other shortened models have been developed as well, mostly by removing still more items from the MBI-HSS-22. So-kum Tang (1998), for example, developed a 19-item model, and Kalliath et al. (2000) a seven-item model. These alternative shortened models were based on a single or limited number of samples and lack cross-validation. Because construct validity of these shortened versions has not been demonstrated, we suggest considering these alternatives when the 20-item version has a poor model fit.

In this article we study the factorial validity of the MBI-HSS-22 and the MBI-HSS-20 by means of CFA and examine metric measurement invariance by comparing data from hospital nurses to data from nurses and assistants in welfare institutions.

Method

Samples

Two samples of human services professionals were studied. The first sample (hospital sample) was composed of hospital nurses (n = 2515). They were recruited in 15 Belgian general hospitals (response rate: 57 per cent). The average age was 32.0 (standard deviation, SD = 7.4); they had been working in the sector for 9.0 years on average (SD = 7.2); and 83 per cent of the group was female. The second sample (welfare sample) was composed of nurses and assistants working in residential welfare institutions for youth with special needs, mentally disabled persons, or geriatric persons (n = 1639). They were recruited in 89 residential institutions in Belgium (response rate: 71 per cent). The average age was 35.0 (SD = 8.5); they had been working in the sector for 11.1 years on average (SD = 8.0); and 81.4 per cent of the group was female. All questionnaires were distributed with a cover letter that asked for anonymous participation and for informed consent, and were collected through letterboxes installed in all participating institutions.

Measures

We used the Flemish version of the MBI, which was obtained through translation and back-translation (Vlerick, 1993). Previous EFA indicated a three-factorial structure underlying the items similar to the one Maslach and Jackson

DOI: 10.1002/smi

(1981) described (Vlerick, 1993), but CFAs did not observe good fit for the 22-item version (Vanheule, Rosseel, & Bogaerts, 2005; Vlerick, 1995). All items ask for frequencies and are scored on a seven-point Likert scale.

Analysis

CFAs were conducted with Lisrel 8.72 using ML-estimation and Satorra–Bentler corrections. We fitted the factor-structure model described by Maslach et al. (1996). Model fit was evaluated starting from: (1) the Comparative Fit Index (CFI) based on the minimum fit function Chi-square, for which values >0.90 indicate good fit; (2) the Tucker–Lewis Index (TLI) based on the minimum fit function Chi-square, for which values of >0.90 indicate good fit; (3) the Standardized Root Mean Square Residual (SRMR), for which values <0.05 indicate good fit; (4) and the Root Mean Square Error of Approximation (RMSEA), for which values <0.05 indicate good fit (Hoyle, 1995). In assessing metric measurement invariance we used the Chi-square difference test, starting from Satorra–Bentler corrected Chi-square (Muthén & Muthén, 2005).

Results

Table I presents descriptive statistics. The fit statistics of the CFAs can be found in Table II. These indicate that in both samples the fit of the MBI-HSS-20 is good at all criteria, whereas the fit of the MBI-HSS-22 is not always acceptable. For the MBI-HSS-22 the CFI and TLI are too low in the welfare sample whereas the SRMR is too high. In the hospital sample the TLI is too low whereas the SRMR and the RMSEA are too high. Figures 1 and 2 depict the well-fitting model with standardized factor-loadings in both samples. We examined the modification indices of the MBI-HSS-22 and observed that in both samples item 12 was the item with the highest cross-loading. When we eliminated this item from the model, items 16 and 18 had the highest cross-loading in the welfare sample, and in the hospital sample item 16 had the second highest cross-loading after item 13. Because item 18 had a low cross-loading in the welfare sample and item 13 a low cross-loading in the hospital sample, item 16 could reasonably be considered the second most problematic item of the set.

In the next stage we tested the metric measurement invariance of the MBI-HSS-20 across both samples and observed that the hypothesis of invariance does not hold: Chi-square difference = 146.27; df = 17; $p < 0.001$. Subsequently we checked whether one of the three factors caused this significant difference (see Table III). All possible models in which one or two factors were

Table I. Descriptive statistics for the 22 MBI-HSS items in the hospital sample and the welfare sample.

MBI-HSS item	Hospital sample		Welfare sample	
	Mean	SD	Mean	SD
1	1.80	1.51	1.74	1.42
2	2.29	1.63	2.13	1.56
3	2.01	1.53	1.70	1.39
4	4.87	1.32	5.47	0.88
5	2.14	1.63	1.78	1.70
6	1.55	1.55	1.12	1.37
7	4.50	1.50	5.28	1.10
8	1.54	1.60	1.27	1.40
9	3.57	1.78	4.28	1.66
10	1.15	1.52	0.56	1.10
11	1.41	1.57	0.99	1.37
12	4.02	1.52	4.63	1.31
13	1.37	1.64	1.20	1.37
14	2.37	1.84	2.26	1.84
15	0.78	1.23	0.32	0.95
16	1.23	1.40	0.62	1.01
17	4.33	1.49	4.89	1.33
18	4.05	1.59	4.74	1.33
19	3.50	1.80	4.37	1.55
20	1.27	1.46	1.03	1.28
21	3.91	1.58	4.59	1.40
22	0.92	1.21	0.92	1.26

Table II. Fit statistics of the 22-item MBI-HSS and the 20-item MBI-HSS.

Model	Sample	Chi-square	df	CFI	TLI	SRMR	RMSEA
MBI-HSS-22 items	Welfare	1092.86	206	0.88	0.86	0.06	0.05
	Hospital	1798.41	206	0.90	0.89	0.06	0.06
MBI-HSS-20 items	Welfare	695.84	167	0.92	0.91	0.05	0.04
(minus items 12 & 16)	Hospital	1171.68	167	0.93	0.92	0.05	0.05

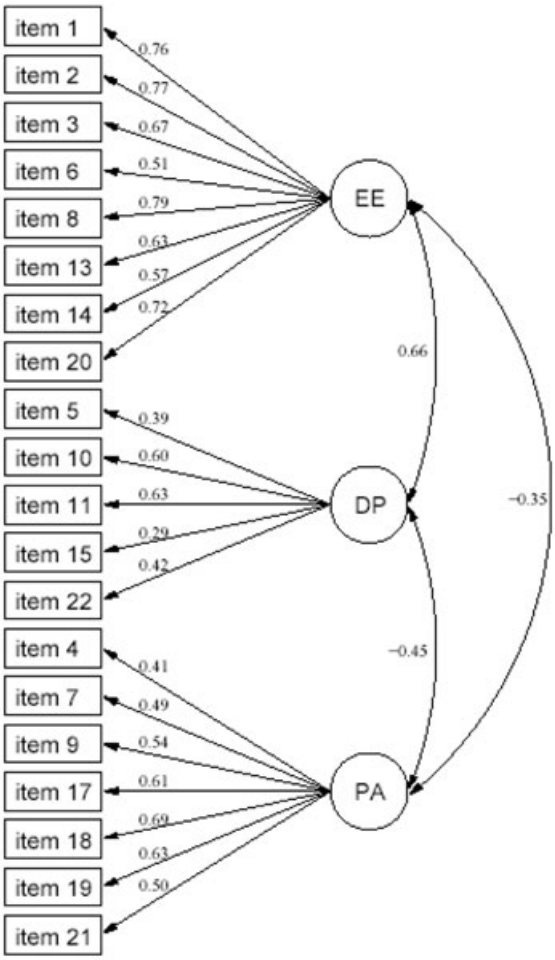

Figure 1. Path diagram with standardized factor-loadings of the 20-item MBI-HSS in the welfare sample (n = 1639): EE = emotional exhaustion, DP = depersonalization, PA = personal accomplishment.

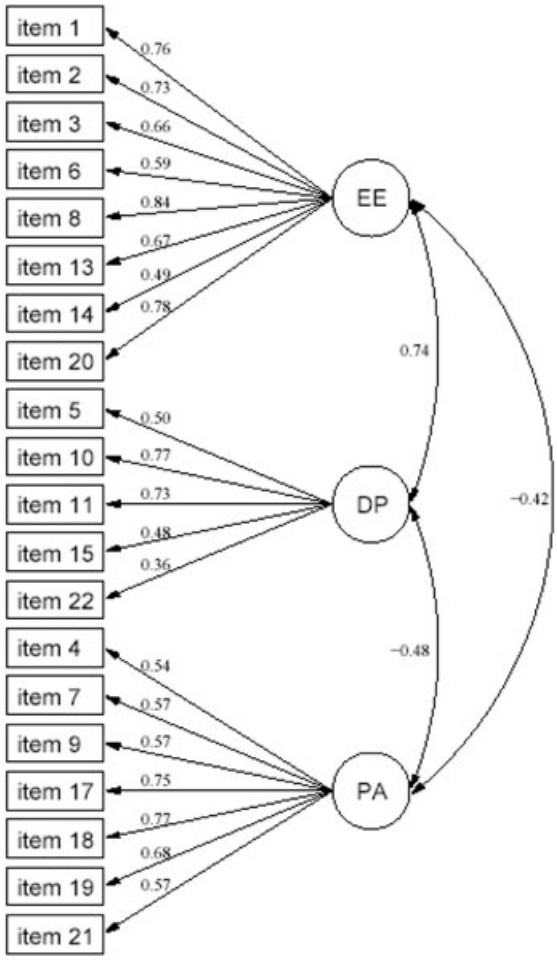

Figure 2. Path diagram with standardized factor-loadings of the 20-item MBI-HSS in the hospital sample (n = 2515): EE = emotional exhaustion, DP = depersonalization, PA = personal accomplishment.

Table III Test of partial measurement invariance of the 20-item MBI-HSS.

Partial invariance models MBI-HSS-20	Chi-square difference	df	p
EE free, DP and PA fixed	100.14	10	<0.001
DP free, EE and PA fixed	94.99	13	<0.001
PA free, EE and DP fixed	102.26	11	<0.001
EE fixed, DP and PA free	47.65	7	<0.001
DP fixed, EE and PA free	52.66	4	<0.001
PA fixed, EE and DP free	45.49	6	<0.001

Note: EE = emotional exhaustion; DP = depersonalization; PA = personal accomplishment.

tested for partial invariance had significant Chi-square difference values.

Discussion

In this paper we evaluated the factorial structure of the MBI-HSS-22 and the MBI-HSS-20 starting with data from a hospital sample and a welfare sample. CFAs indicated good fit for the MBI-HSS-20 and weak fit for the MBI-HSS-22. This result is in line with findings from various researchers (e.g. Gil-Monte, 2005; Hallberg & Sverke, 2004; Schaufeli & Van Dierendonck, 1993) and indicates that the MBI-HSS-20 consists of well-fitting

DOI: 10.1002/smi

unidimensional subscales that measure the burnout dimensions as defined by Maslach and Jackson (see Schaufeli & Enzmann, 1998).

A subsequent test of measurement invariance indicated that the factor-structure model is not invariant across both samples. Neither could partial invariance at the level of the subscales be assumed. This means that the loadings of the items on the factors are substantially different in both samples, and suggests that the interpretation of the meaning of emotional exhaustion, depersonalization, and reduced personal accomplishment is sample specific. For hospital nurses the three MBI-HSS factors seem to mean something other than what they mean for human service professionals employed in welfare institutions. We think this is influenced by differences in job content, work circumstances, and professional goals in the respective sectors. Given the fact that only scant attention in the literature on the MBI-HSS has been devoted to the particular meaning of the three dimensions underlying burnout in different occupational groups, we suggest that future research would do well to focus on it. This type of research could contribute to the further development of occupation-specific knowledge and interventions with respect to burnout.

References

Byrne, B.M. (1994). Testing for the factorial validity, replication, and invariance of a measuring instrument: a paradigmatic application based on the Maslach Burnout Inventory. *Multivariate Behavioral Research*, *29*, 289–311.

Gil-Monte, P.R. (2005). Factorial validity of the Maslach Burnout Inventory (MBI-HSS) among Spanish professionals. *Revista Saude Publica*, *39*, 1–8.

Hallberg, U.E., & Sverke, M. (2004). Construct validity of the Maslach Burnout Inventory: Two Swedish health care samples. *European Journal of Psychological Assessment*, *20*, 320–338.

Hoyle, R.H. (1995). *Structural equation modelling—Concepts, issues and applications*. London: Sage.

Kalliath, T.J., O'Driscoll, M.P., Gillespie, D.F., & Bluedorn, A.C. (2000). A test of the Maslach Burnout Inventory in three samples of healthcare professionals. *Work & Stress*, *14*, 35–50.

Maslach, C., & Jackson, S.E. (1981). *Maslach Burnout Inventory. Research edition.* Palo Alto, CA: Consulting Psychologist Press.

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

Muthén, L., & Muthén, B. (2005). Chi-square difference testing using the S-B scaled chi-square. Note on M-plus website. Retrieved February 1, 2006 from http://www.statmodel.com

Schaufeli, W.B., & Enzmann, D. (1998). *The burnout companion to study and research: A critical analysis*. London: Taylor & Francis.

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

Schaufeli, W.B., Bakker, A.B., Hoogduin, K., Schaap, C., & Kladler, A. (2001). On the clinical validity of the Maslach Burnout Inventory and the Burnout Measure. *Psychology and Health*, *16*, 565–582.

So-kum Tang, C. (1998). Assessment of burnout for Chinese human service professionals: A study of factorial validity and invariance. *Journal of Clinical Psychology*, *54*, 55–58.

Vanheule, S., Rosseel, Y., & Bogaerts, S. (2005). Measuring professional burnout in Dutch-speaking regions: An evaluation of the factorial validity of the Maslach Burnout Inventory. *Psychologica Belgica*, *45*, 147–156.

Vlerick, P. (1993). Exploratieve analyse van de factoriële validiteit van de Maslach Burnout Inventory. *Tijdschrift Klinische Psychologie*, *23*, 235–250.

Vlerick, P. (1995). Confirmatorische analyse van de factorstructuur en de validiteit van de Maslach Burnout Inventory. *Gedrag & Gezondheid*, *32*, 35–40.