Measuring burnout in social work: factorial validity of the Maslach Burnout Inventory-Human Services Survey

Article type: original article

Word count (incl. abstract, figures, tables and references): 5145

Correspondence concerning this article should be addressed to Ann Doherty, School of Psychology, Ulster University, Cromore Road, Coleraine, Co. Londonderry BT52 1SA.

Email: doherty-a80@ulster.ac.uk.

Ann S Doherty and John Mallett

Ulster University Coleraine Campus

Michael P Leiter

Deakin University

Paula McFadden

Ulster University Magee Campus

Affiliations

Ann S Doherty, School of Psychology, Ulster University; John Mallett, School of Psychology, Ulster University; Michael Leiter, School of Psychology, Deakin University; Paula McFadden, School of Applied Social and Policy Sciences, Ulster University.

Abstract

Several studies challenge the three-dimensional structure of the Maslach Burnout Inventory-Human Services Survey (MBI-HSS), citing alternative measurement models including bifactor models. Whilst bifactor models have merit, if data sampling violates assumptions of Stochastic Measurement Theory (SMT) the bifactor model requires modification prior to application. The present study compared five alternative MBI-HSS factor models using both Confirmatory Factor Analysis (CFA) and Exploratory Structural Equation Modeling (ESEM). Data from a cross-sectional survey of United Kingdom (UK) social workers was examined (*N* = 1257), with validation analyses conducted in an independent sample (*N* = 162). Bifactor models, re-specified to account for SMT, provided good fit. However, improved fit was observed for a bifactor-ESEM specification, in both test ($\chi 2$ = 1112.93, *df* = 149, *p* < .001, CFI = .969, RMSEA = .072 [90% CI .068, .076]) and validation ($\chi 2$ = 227.89, *df* = 149, *p* < .001, CFI = .978, RMSEA = .057 [90% CI .042, .072]) samples. The results confirm the MBI-HSS possesses a bifactor structure in UK social workers when SMT is considered, and that bifactor-ESEM may provide a better framework to examine MBI-HSS.

Keywords-burnout; MBI-HSS; social work; bifactor; bifactor-ESEM

Introduction

The Maslach Burnout Inventory-Human Services Survey (MBI-HSS; Maslach, Jackson & Leiter, 1996) is the most cited burnout measure (Worley, Vassar, Wheeler & Barnes, 2008). It comprises three subscales posited to measure the latent dimensions of burnout: emotional exhaustion (EE), depersonalisation (DP) and reduced personal accomplishment (PA). The respondent receives three separate scores, each reflecting a unique aspect of the phenomenon. These dimensions have shown differential relationships with turnover intentions, job satisfaction and organisational commitment (Lee & Ashforth, 1996).

Whilst it has been accepted that these latent dimensions possess some interrelationship, there has been no consensus on the number of dimensions. Two, four and five-factor solutions have been reported across a wide range of occupations (Chao, McCallion & Nickle 2011; Densten, 2001; Walkey & Green, 1992). Furthermore, the literature suggests 16 (Stalker, Harvey, Frensch, Mandell & Adams, 2008), 18 (Yadama & Drake, 1995) and 19 (Kim & Ji, 2009) item scales to be more parsimonious. There is also evidence of additional scale complexity, with some items loading onto more than one factor. Items 12 and 16 are particularly problematic and are frequently omitted in response (Kim & Ji, 2009; Maslach et al., 1996).

Recently, it has been argued the debate surrounding the MBI-HSS factor structure may be settled if one uses a bifactor approach (Mészáros, Ádám, Szabó, Szigeti & Urbán, 2014). Bifactor models allow the simultaneous estimation of a general factor and specific factors, similar to a higher-order model. Unlike higher-order models, the bifactor approach allows all indicators to load onto a general factor and their specific factor. In the higher-order model, the influence of the general factor on the indicators is mediated via the first order factors. Following superior fit with a bifactor model, Mészáros and colleagues (2014) concluded only the PA items explained additional variance in burnout, over and above a general burnout factor. However, the correlated first-order factor model is nested within the bifactor model (Yung, Thissen & McLeod, 1999) and the bifactor model will always provide better fit than a correlated factors model (Reise, 2012).

Despite providing superior fit, Mészáros and colleagues (2014) offer little explanation of unusual factor loading patterns identified within their bifactor solution. Item 4, a PA item, did not load significantly onto their *'global burnout'* general factor. Item 13 exhibited a negative loading on the EE specific factor, which was at odds with the valence of other indicators of this specific factor. Despite this, Mészáros and colleagues (2014) do not speculate as to why they have occurred, how they relate to the literature, or what implications they have for score computation.

Anomalous results using bifactor approaches occur so frequently that Eid, Geiser, Koch and Heene (2017) question the appropriateness of their application. Eid and colleagues note that in many cases indicators are found to have small loadings, non-significant loadings, or even negative loadings on their specific factor, as observed in Mészáros et al., (2014). Such findings are unexpected given loading patterns are often regular when modelled as a simpler correlated first-order factor model (Eid et al., 2017). They further argue that not all indicators load onto the general (*g*) factor, another unexpected anomalous result.

Eid and colleagues (2017) argue traditional bifactor models in many empirical studies violate assumptions underlying Stochastic Measurement Theory (SMT). When a series of observed indicators are used in a random population sample, only one level of sampling has occurred. However, unless the domain specific factors are also selected at random (second level of sampling), *g* cannot be considered a true random variable. Where two-level sampling is impossible, Eid et al., (2017) offer alternative strategies to ensure the general and specific factors remain true random variables. One approach is to use one specific factor as a comparison standard (*S-1* model, Figure 1a), whilst the second approach uses one item as a reference indicator (*S·I-1*, Figure 1b) (Eid et al., 2017).

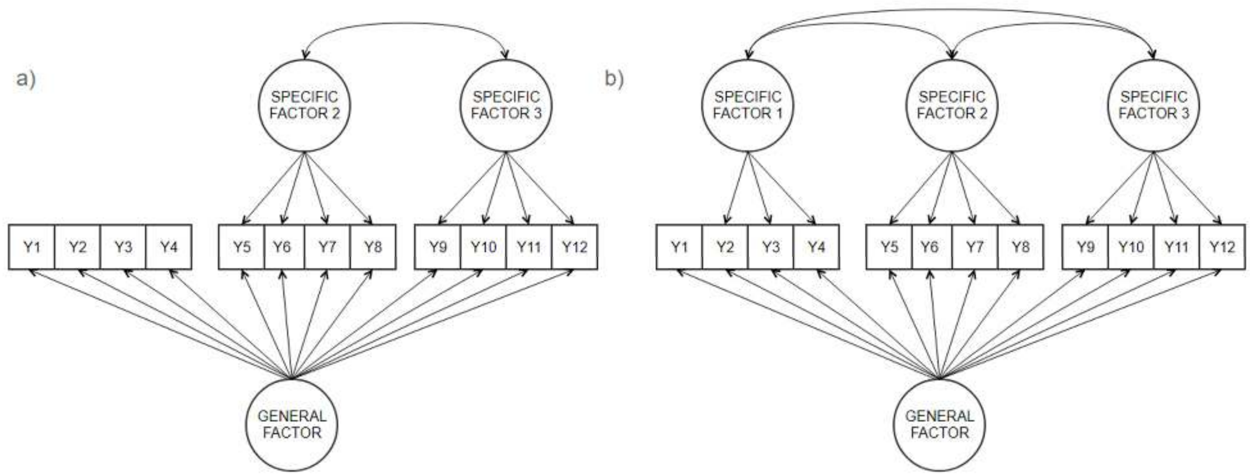


Figure 1: Alternative bifactor model parameterisations where single-level sampling has occurred (Eid et al., 2017), where a) S-1 model and b) S·I-1 model.

The superior performance of bifactor specifications of the MBI-HSS may result from an inherent bias to accommodate unidentified model complexity. Higher-order factor models have fewer parameters than bifactor models, and thus will never fit better from a chi-square perspective (Gignac, 2016). As such, items that cross-load are more easily accommodated by the bifactor approach.

When specified as a correlated factor model, MBI-HSS items 6, 12 and 16 have been found to cross-load onto other latent dimensions (Kanste, Miettunen & Kyngäs, 2006; Kim & Ji, 2009). The Independent Cluster Model (ICM) constraints of CFA assumes that cross-loadings between items and non-target factors are exactly zero (Howard, Gagné, Morin & Forest, 2016). However, many questionnaire items are rarely unique indicators of a single construct and may tap into additional latent factors that have some conceptual relationship

(Morin, Arens & Marsh, 2016). Thus, it is prudent to examine MBI-HSS using Exploratory Structural Equation Modeling (ESEM) to examine these cross-loaded items in greater detail.

Previous research involving social workers has suggested a better fitting latent structure can be achieved using reduced items scales (Kim & Ji, 2009; Stalker et al., 2008; Yadama & Drake, 1995). These authors deleted items due to high inter-item correlations, cross-loadings and small squared multiple correlations, concluding that improved fit indices indicated better fitting models. However, comparing models with different numbers of items is problematic. Furthermore, the failure to cross-validate findings and examine the relationships with external criteria limits the utility of these alternative specifications.

The present study sought to compare a correlated first-order factor model of the MBI-HSS with alternative bifactor specifications compatible with SMT, in a social work sample. Furthermore, we sought to examine whether re-specification within an ESEM framework would provide a better fitting latent structure. Additionally, the performance of the optimal measurement model in a validation sample, and its relationship with an external criterion, resilience, was examined. A clearer understanding of the true measurement model of burnout among social workers is imperative for a more accurate assessment of burnout. Since most previous research has examined child protection social workers, the present study further sought to examine the factorial validity of the MBI-HSS in a more heterogeneous sample of social work specialisms.

Method

Sample

Data from cross-sectional surveys of social workers in the United Kingdom (UK) was examined. In the test sample the survey was emailed to all social workers registered with Community Care©. A total of 1,257 participants completed the MBI-HSS component of the survey. Validation analyses were conducted in independently collected data (N = 162) reported in McFadden, Mallett, Campbell and Taylor (2019). There were no issues of missing data on MBI-HSS items in either sample. Sample characteristics can be observed in Table 1.

Table 1

Frequencies for demographic characteristics of test (N=1257) and validation (N=162) samples

		Test sample	Validation sample
Characteristic		*n (%)*	*n (%)*
Gender	Female	1028 (81.8)	140 (86.4)
	Male	201 (16)	22 (13.6)
Age (years)	18-25	48 (3.8)	14 (8.6)
	26-35	284 (22.6)	76 (46.9)
	36-45	286 (22.8)	43 (26.5)
	46-55	395 (31.4)	23 (14.2)
	56+	229 (18.2)	6 (3.7)
Residence	England	1087 (86.5)	-
	Scotland	61 (4.9)	-
	Wales	67 (5.3)	-
	Northern Ireland	15 (1.2)	162 (100)
Practice area	Child protection	358 (28.5)	162 (100)
	Other children's services	234 (18.6)	-
	Older peoples' services	238 (18.9)	-
	Mental health	131 (10.4)	-
	Adult disability services	121 (9.6)	-

Measures

Burnout was assessed using the 22-item MBI-HSS (Maslach et al., 1996), which asks respondents how frequently they experience feelings in line with item statements. Items were scored on a seven-point scale from 0 (*never*) to 6 (*every day*). The EE subscale (9 items) measured feelings of being emotionally overextended by one's work. The DP subscale (5 items) measured an impersonal response toward recipients of one's care. The PA subscale (8 items) measured feelings of competence in one's work. Higher scores on the EE and DP subscales, accompanied with lower PA scores, indicated a higher degree of burnout. The MBI-HSS has shown good reliability in previous social work studies (Kim & Ji, 2009; Yadama & Drake, 1995).

In the validation sample resilience was measured using the self-report Resilience Scale (RS14: Wagnild & Young, 2009). Items were scored on a seven-point scale from 1 (*strongly disagree*) to 7 (*strongly agree*) and summated to form an overall score whereby higher scores indicate greater resilience. The RS-14 has shown high internal consistency (Aiena, Baczwaski, Schulenberg & Buchanan, 2014; Taku, 2014), as well as significant correlations in expected directions with life satisfaction and psychological distress, in both clinical and non-clinical samples (Aiena et al., 2014). Higher RS14 scores were predictive of lower EE and higher PA among physicians (Taku, 2014).

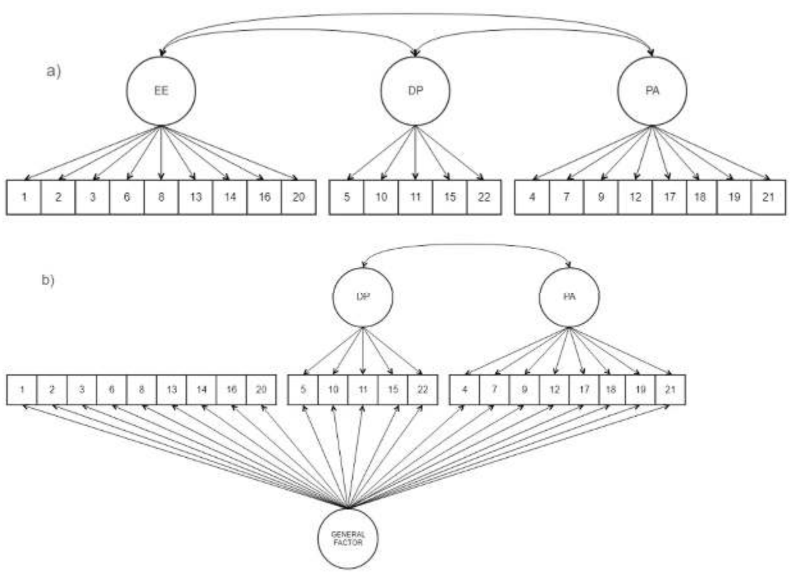
Analyses

Five confirmatory factor models were specified (Figure 2) and estimated in the test sample using Mplus 8.1 (Muthén & Muthén, 2018). Model A comprised the correlated three-factor solution hypothesised by Maslach et al., (1996). Three *S-1* models were specified, aimed at representing item heterogeneity differently, with EE, DP and PA as the reference domains, respectively. In Model B, *g* represents the individuals EE corrected for measurement error. In this model, the DP and PA factors represent whether a participant has higher or lower DP and PA than one would expect given their EE level. In Model C, *g* signifies DP with the EE and PA factors representing whether one would expect higher or lower scores based upon this DP level. Similarly, Model D considers the general factor to be PA, with the EE and DP factors indicating whether one would expect higher or lower scores based upon this PA level. Given the original conceptualisation of the MBI-HSS as a tripartite phenomenon, it was important to examine the performance of these three distinct *S-1* specifications. Eid and colleagues (2017) note model fit can change when the reference domain changes as the selection of a different reference domain represents potential item heterogeneity in a different way. An alternative is to specify the less restrictive *S·I*-1 model to represent item heterogeneity more generally.

For the *S·I*-1 model (Model E), item 8 (*feel burned out*) was selected as the reference indicator. Eid and colleagues (2017) advocate this reference indicator be a 'gold-standard' indicator for the general factor. Mészáros et al., (2014) found item 8 to be the strongest loading item onto their *'global burnout'* general factor. Furthermore, West, Dyrbye, Satele, Sloan and Shanafelt (2012) found that, when used as a single indicator, item 8 revealed strong and consistent relationships with outcomes such as turnover intentions, relative to the use of the full MBI-HSS scale, in a physician sample. An examination of factor loadings for the original MBI-HSS (Model A) in the test sample indicated that item 8 had both the highest loading and squared multiple correlation. Eid et al., (2017) state the *S·I*-1 model should be applied when the inclusion of the additional specific factor makes practical sense. Emotional exhaustion is considered a key aspect of Leiter's (2008) Two Process Model of Burnout, hypothesized to develop from a chronic mismatch in job demands and resources. Therefore, the application of an *S·I*-1 model with an EE specific domain was warranted. Considering this evidence, and the item phrasing, item 8 was selected as the most appropriate reference item.

Analyses were conducted using the WLSMV estimator with THETA parameterisation (Muthén & Muthén, 2018). Goodness of fit was assessed using chi-square (χ^2) test statistic, the comparative fit index (CFI) and the Tucker Lewis index (TLI). A non-significant chi-square, CFI and TFI >.95 indicate good model fit (Hu & Bentler, 1999). Furthermore, a root mean square error of approximation (RMSEA) <.06 and a standardised root mean square residual (SRMR) <.08 indicate acceptable model fit (Hu & Bentler, 1999). As the test sample size exceeded those recommended for confirmatory factor analyses (Comrey & Lee, 1992), the sample was deemed sufficiently powered and formal sample size calculations were not conducted. The data examined are available on request and are not publicly available so as not to compromise participant or organizational identification.

Following the examination of the five alternative CFA models (Figure 2) in both test and validation samples, exploratory structural equation models (ESEM) versions of models A-E were specified using the target rotation function. Thus, indicators could cross-load onto the other reference domains, with these loadings targeted to approximate zero.



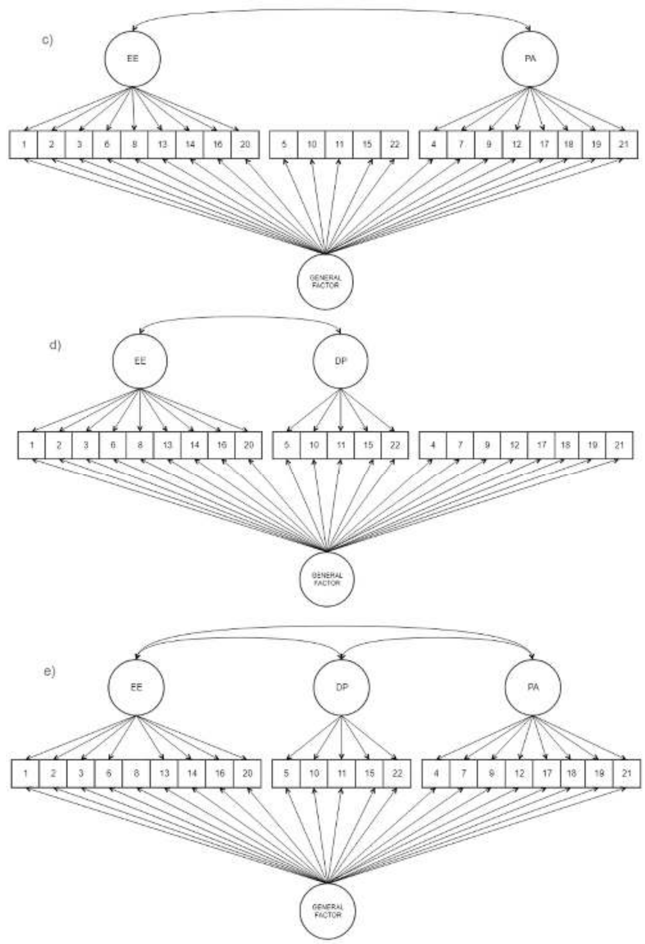


Figure 2: Alternative factor models examined where a) Maslach et al., (1996); b) S-1 with EE as the reference domain; c) S-1 with DP as the reference domain; d) S-1 with PA as the reference domain; e) S·I-1 with item 8 as the reference item.

Results

CFA models

The chi-square values for all CFA models were statistically significant due to the large sample size (Tanaka, 1987). Accordingly, model fit was examined using a variety of fit statistics (Table 2). In order to achieve model identification, factor loadings for four items (items 1, 2, 3 and 6) on the specific factor were fixed to 1 in Model E. Eid et al., (2017) previously reported model identification issues when specifying the *S·I-1* model which were resolved by fixing the item loadings onto the specific factor to 1. As we encountered similar problems these loadings were fixed in a sequential manner until model identification was achieved.

All models showed acceptable fit with respect to CFI and SRMR indices in both samples. The RMSEA values and 90% confidence intervals for the CFA models did not indicate an acceptable fit. However, improved values were obtained within the validation sample. Whilst the utility of threshold values for fit indices has been debated, it has been argued an element of human judgement must be incorporated when assessing fit (Chen, Curran, Bollen, Kirby & Paxton, 2008). Thus, model fit indices were examined in conjunction with the standardised factor loadings and modification indices, as well as a consideration of their theoretical underpinnings.

The bifactor models displayed anomalous loading patterns for PA items 4 and 21 (see Electronic Supplementary Material). In Model B item 4 loaded positively onto the general factor, whereas all other PA items loaded negatively. Personal accomplishment items are positively worded and were thus expected to load negatively onto *g* when specified using the negatively worded EE as the reference domain. Furthermore, item 4 did not load significantly onto *g* in the validation sample for Models B and E, or in either sample when Model C was specified. Item 21 did not load onto *g* when Model B was specified in the test sample, nor did it load significantly onto *g* when Model E was specified in either sample. In both samples moderate intercorrelations between the specific factors were observed for Models D and E.

Of the CFA models the bifactor specifications showed superior fit indices; Model E possessed the highest CFI and lowest RMSEA in both samples. As Model A is nested within Model E, a chi-square difference test was conducted and indicated the correlated first-order factor specification provided significantly worse fit to the test ($\Delta\chi^2(18, N = 1257) = 759.88$, $p < .001$) and validation ($\Delta\chi^2(18, N = 162) = 64.50$, $p < .001$) samples.

Table 2

Comparative fit indices of alternative models of the MBI-HSS in both test and validation samples

	Test sample (N = 1257)							Validation sample (N = 162)						
Model	χ^2	*df*	*p*	CFI	TLI	RMSEA [90% CI]	SRMR	χ^2	*df*	*p*	CFI	TLI	RMSEA [90% CI]	SRMR
CFA-A	3210.98	206	<.001	.903	.891	.108 [.104, .111]	.072	413.03	206	<.001	.942	.935	.079 [.068, .090]	.070
CFA-B	2725.24	195	<.001	.918	.903	.102 [.098, .105]	.054	371.73	195	<.001	.950	.941	.075 [.063, .086]	.058
CFA-C	2421.45	191	<.001	.928	.913	.096 [.093, .100]	.053	388.20	191	<.001	.944	.933	.080 [.068, .091]	.062
CFA-D	2856.46	194	<.001	.914	.897	.104 [.101, .108]	.064	374.33	194	<.001	.949	.939	.076 [.064, .087]	.062
CFA-E	2070.52	188	<.001	.939	.925	.089 [.086, .093]	.046	346.71	188	<.001	.955	.945	.072 [.060, .084]	.054
ESEM-A	1846.52	168	<.001	.946	.925	.089 [.086, .093]	.032	282.18	168	<.001	.968	.956	.065 [.051, .078]	.042
ESEM-B	2709.05	184	<.001	.918	.897	.104 [.101, .108]	.050	338.58	184	<.001	.956	.945	.072 [.060, .084]	.053
ESEM-C	1748.80	176	<.001	.949	.933	.084 [.081, .088]	.038	333.98	176	<.001	.955	.942	.074 [.062, .087]	.053
ESEM-E	1112.93	149	<.001	.969	.952	.072 [.068, .076]	.026	227.89	149	<.001	.978	.966	.057 [.042, .072]	.037

Note: χ^2: chi square test statistic; *df*: degrees of freedom; CFI: comparative fit index; TLI: Tucker Lewis index; RMSEA: root-mean-square error of approximation; CI: confidence interval; SRMR: standardised root-mean-square residual; ESEM: exploratory structural equation model; no convergence was obtained for ESEM model D

Exploratory Structural Equation Modeling (ESEM)

Despite Model E showing acceptable to good fit in the test and validation samples, respectively, moderate factor intercorrelations (see Electronic Supplementary Material) suggested that this may not be the best fitting model for the MBI-HSS. Thus, all five models were re-specified as ESEM models. Model fit indices for ESEM specifications can also be observed in Table 2. The bifactor-ESEM Model E was selected as the final model, as it demonstrated the best CFI, RMSEA and SRMR values in both samples. Furthermore, the decreased factor intercorrelations observed further supported the selection of the bifactor-ESEM Model E as the best fitting model. Standardised factor loadings and factor correlations for Model E-ESEM can be observed in Table 3. Non-significant loadings onto the general factor were observed for items 21 and 4 in the test and validation samples, respectively.

Measurement Invariance

Model identification could not be obtained for bifactor-ESEM Model E when examining configural invariance in the test sample. An examination of the bifactor-ESEM Model E for males and females separately showed good fit for both males and females, however, differences in loading patterns on the *g* factor were observed (see Electronic Supplementary Material). Thus, gender effects were examined using a Multiple Indicators, Multiple Causes Model (MIMIC). No significant association was observed between gender and the latent factors of the bifactor-ESEM Model E; however, the modification indices suggested some misspecification. The inclusion of paths between gender and the items directly revealed significant associations for items 12 (β = -.119, p < .001) and 21 (β = -.098, p = .001), controlling for the effects of the latent factors, χ^2 (165, N = 1229) = 1073.84, p < .001, CFI = .971, RMSEA = .067 [90% CI .063, .071]).

Table 3

Standardised Factor Loadings and Correlations for the S·I-1 bifactor-ESEM model

	Test sample (N=1257)				Validation sample (N=162)			
	GB	EE	DP	PA	GB	EE	DP	PA
MB1	.866**	-.153**	.026	.060*	.825**	.279**	.003	.038
MB2	.853**	-.214**	.037	.010	.835**	.307**	-.054	-.085
MB3	.828**	-.041	-.006	-.042	.765**	.082	-.074	-.220**
MB4	.115*	-.041	-.178**	.339**	.063	.082	-.203*	.301**
MB5	.316**	.025	.552**	-.097**	.434**	-.100	.464**	-.049
MB6	.605**	.528**	.092**	.009	.697**	-.274*	-.061	-.206*
MB7	-.146**	.093*	-.102**	.635**	-.151	.269**	-.237**	.548**
MB8	.896**	N/a	N/a	N/a	.865**	N/a	N/a	N/a
MB9	-.173**	.152**	-.024	.812**	-.210*	-.093	-.145	.659**
MB10	.385**	-.033	.772**	.031	.465**	.018	.805**	-.020
MB11	.428**	-.018	.705**	.075**	.440**	.126*	.619**	-.033
MB12	-.475**	-.012	.112**	.374**	-.334**	-.214**	-.135	.429**
MB13	.718**	-.048	.075*	-.048	.709**	.068	.046	-.123
MB14	.673**	-.048	.097*	.083*	.553**	.085	.166*	.050
MB15	.254**	.146**	.527**	-.116**	.211*	-.159*	.539**	-.155
MB16	.546**	.629**	.041	.029	.650**	-.215*	.010	-.174*
MB17	-.117**	-.159**	-.029	.518**	-.224*	.210*	-.224*	.487**
MB18	-.207**	-.136**	.110**	.585**	-.189*	-.121	.017	.651**
MB19	-.192**	.048*	.037	.777**	-.209*	.180*	.217*	.869**
MB20	.845**	-.004	-.006	-.033	.776**	-.025	-.087	-.090
MB21	.009	-.079*	-.065	.480**	-.209*	.408**	.055	.438**
MB22	.345**	.029	.241**	-.090*	.566**	.149	.278**	-.144*
Correlations								
EE	N/a				N/a			
DP	N/a	.37**			N/a	-.22*		
PA	N/a	-.30**	-.29**		N/a	.13	-.36**	
ω^a	.894	.804	.768	.775	.886	.751	.754	.699
ω_H^b	.608	.029	.319	.569	.670	.038	.287	.648

Note: GB: Global Burnout; EE: Emotional Exhaustion; DP: Depersonalisation; PA: Personal Accomplishment; N/a: non-applicable; ω: McDonald's omega; a: omega subscale in the case of EE, DP and PA specific factors; ω_H: omega hierarchical; b: omega hierarchical subscale in the case of EE, DP and PA specific factors; * $p < .05$; ** $p < .001$

Validity

Correlations between the latent factors of bifactor-ESEM Model E and resilience, modeled as a latent factor, indicated higher resilience was associated with lower global burnout (r = -.32, p < .001, 90% CI -.45, -.19), higher personal accomplishment (r = .53, p < .001, 90% CI .40, .66) and higher emotional exhaustion (r = .68, p = .002, 90% CI .31, 1.04). No significant correlation was observed between resilience and depersonalisation (r = -.11, p = .292, 90% CI -.29, .06).

Discussion

The frequent identification of undesirable loading patterns for the MBI-HSS led researchers to respond by either eliminating cross-loaded items (Kim & Ji, 2009; Stalker et al., 2008; Yadama & Drake, 1995) or to specify a bifactor model (Mészáros et al., 2014). However, Eid et al., (2017) argue that Mészáros and colleagues' (2014) bifactor specification violates Stochastic Measurement Theory (SMT), calling their conclusions into question. The results show that dimensionality within the MBI-HSS is best represented by a bifactor-ESEM specified in accordance with SMT. The *S·I*-1 specification using item 8 (*feel burned out)* as the reference item was the better performing CFA model and when specified within an ESEM framework provided the best fit to both test and validation samples.

Eid et al., (2017) argue that the absence of at least one significant loading onto the general factor was an artefact of violations of SMT by traditional bifactor specifications. Here, we have shown that anomalous loading patterns for item 4 of MBI-HSS persist, despite the use of SMT-compatible bifactor specifications. Stalker and colleagues (2008) argue social worker responses to this item may diverge from their response pattern to other PA items. Differences in life experiences may mean social workers appraise themselves as incapable of truly understanding their clients' feelings, if they have not experienced similar life challenges. Similarly, social workers may display empathy, indicating an ability to understand their clients, and yet still experience a sense of inefficacy.

Few existing studies examining alternative MBI-HSS specifications consider relationships with external criteria. Here, we show the bifactor-ESEM displayed correlations with resilience in the expected direction, supporting the external validity of our model. Higher resilience was associated with higher levels of personal accomplishment and lower global burnout. Taku (2014) found higher resilience predictive of lower emotional exhaustion and higher personal accomplishment among physicians, with no association with depersonalisation identified. However, we observed an unexpected positive association with the emotional exhaustion reference domain, dominated by items 6 (*people work a strain)* and 16 (*people work too stressful)*. We posit that high levels of resilience displayed by social workers does not prevent them from recognising the emotionally draining nature of their role.

The present findings must be interpreted in light of several limitations. A self-selected sample may serve to inflate burnout levels through greater survey engagement. Alternatively, those who experienced burnout may have already exited the profession. Furthermore, the analysis presented here did not offer the opportunity to examine the discriminant and predictive validity of the bifactor-ESEM Model E. Additionally, measurement invariance could not be examined due to model identification issues. Whilst some gender differences in item functioning were observed, further research is required to determine whether a bifactor-ESEM approach is equivalent for males and females. Whilst the gender imbalance observed reflects the natural skew reported within the profession (General Social Care Council, 2010), males were underrepresented in both test and validation samples.

To date, the three-dimensional conceptualisation of burnout has dominated the literature. Reconceptualising burnout as a general phenomenon, with additional specific factors that explain additional variance in the experience, may alter what we already understand about burnout and associated correlates. Future research is required to examine how this novel bifactor structure integrates with correlates identified to date. In practice, examining the relationships between burnout and potential covariates will require the incorporation of structural equation modeling techniques and not merely score computation, which may be beyond the scope of some practitioners.

The present study is unique as it is the first to examine the factorial validity of the MBI-HSS in UK social workers. The study extends the existing literature by confirming that when SMT is considered, the underlying structure of the MBI-HSS remains that of a bifactor specification. Furthermore, the study highlights that bifactor-ESEM provides a better framework within which to examine the latent structure of MBI-HSS. Furthermore, the study highlights the importance of examining the factor structure of the MBI-HSS prior to examining relationships with related constructs. An accurate identification of burnout is critical to develop interventions to assuage the impact on the workforce.

Electronic Supplementary Material

ESM 1. Appendices 1-10 (.doc)

The appendices show additional tables of standardised factor loadings and correlations for tested models.

References

Aiena, B. J., Baczwaski, B. J., Schulenberg, S. E., & Buchanan, E. M. (2015). Measuring resilience with the RS–14: A tale of two samples. *Journal of Personality Assessment*, *97*(3), 291-300. doi:10.1080/00223891.2014.951445

Chao, S. F., McCallion, P., & Nickle, T. (2011). Factorial validity and consistency of the Maslach Burnout Inventory among staff working with persons with intellectual disability and dementia. *Journal of Intellectual Disability Research*, *55*(5), 529-536. doi:10.1111/j.1365-2788.2011.01413.x

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, *36*(4), 462-494. doi:10.1177/0049124108314720

Comrey, A. L. & Lee, H.B. (1992). *A first course in factor analysis (2nd ed.).* Hillsdale, NJ: Lawrence Erlbaum.

Densten, I. L. (2001). Re‐thinking burnout. *Journal of Organizational Behavior*, *22*(8), 833-847. doi: 10.1002/job.115

Dueber, D. M. (2017). Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models. doi.org/10.13023/edp.tool.01

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*(3), 541. doi:10.1037/met0000083

General Social Care Council (2010). *General social care council annual report and accounts 2009-10.* London, UK: The Stationery Office Limited

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57-68. doi:10.1016/j.intell.2016.01.006

Howard, J. L., Gagné, M., Morin, A. J., & Forest, J. (2016). Using bifactor exploratory structural equation modeling to test for a continuum structure of motivation. *Journal of Management*, *44*(7), 2638-2664. doi:10.1177/0149206316645653

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. doi:10.1080/10705519909540118

Kanste, O., Miettunen, J., & Kyngäs, H. (2006). Factor structure of the Maslach Burnout Inventory among Finnish nursing staff. *Nursing & Health Sciences*, *8*(4), 201-207.

Kim, H., & Ji, J. (2009). Factor structure and longitudinal invariance of the Maslach Burnout Inventory. *Research on Social Work Practice*, *19*(3), 325-339. doi:10.1177/1049731508318550

Lee, R. T., & Ashforth, B. E. (1996). A meta-analytic examination of the correlates of the three dimensions of job burnout. *Journal of Applied Psychology*, *81*(2), 123.

Leiter, M. P. (2008). A two-process model of burnout and work engagement: Distinct implications of demands and values. *Giornale Italianol di Medicina del Lavoro ed Ergonomia*, *30*(1 Suppl A), 54-58.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach Burnout Inventory Manual.* Mountain View, CA: CPP. Inc., and Davies-Black.

McFadden, P., Mallett, J., Campbell, A., & Taylor, B. (2019). Explaining self-reported Resilience in child-protection social work: The role of organisational factors, demographic information and job characteristics. *The British Journal of Social Work, 49*(1), 198-216. doi:10.1093/bjsw/bcy015

Mészáros, V., Ádám, S., Szabó, M., Szigeti, R., & Urbán, R. (2014). The bifactor model of the Maslach Burnout Inventory–Human Services Survey (MBI‐HSS)—an alternative measurement model of burnout. *Stress and Health*, *30*(1), 82-88. doi:10.1002/smi.2481

Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 116-139. doi:10.1080/10705511.2014.961800

Muthén, L. K., & Muthén, B. O. (2018). *Mplus statistical modeling software: Release 8.1.* Los Angeles, CA: Muthén & Muthén.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667-696. doi:10.1080/00273171.2012.715555

Stalker, C. A., Harvey, C., Frensch, K., Mandell, D., & Adams, G. R. (2008). Confirmatory factor analysis of the Maslach Burnout Inventory: a replication with Canadian child welfare workers. *Journal of Public Child Welfare*, *1*(3), 77-94.

Taku, K. (2014). Relationships among perceived psychological growth, resilience and burnout in physicians. *Personality and Individual Differences*, *59*, 120-123.

Tanaka, J. S. (1987). How big is big enough? Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, *58,* 134-146.

Wagnild, G. M., & Young, H. M. (2009). *The 14-Item Resilience Scale (RS-14).* Wordem, MT: Resilience Center.

Walkey, F. H., & Green, D. E. (1992). An exhaustive examination of the replicable factor structure of the Maslach Burnout Inventory. *Educational and Psychological Measurement*, *52*(2), 309-323. doi:10.1177/0013164492052002006

West, C. P., Dyrbye, L. N., Satele, D. V., Sloan, J. A., & Shanafelt, T. D. (2012). Concurrent validity of single-item measures of emotional exhaustion and depersonalization in burnout assessment. *Journal of General Internal Medicine*, *27*(11), 1445–1452. doi:10.1007/s11606-012-2015-7

Worley, J. A., Vassar, M., Wheeler, D. L., & Barnes, L. L. (2008). Factor structure of scores from the Maslach Burnout Inventory a review and meta-analysis of 45 exploratory and confirmatory factor-analytic studies. *Educational and Psychological Measurement*, *68*(5), 797-823. doi:10.1177/0013164408315268

Yadama, G. N., & Drake, B. (1995). Confirmatory factor analysis of the Maslach Burnout Inventory. *Social Work Research*, *19*(3), 184-192. doi:10.1093/swr/19.3.184

Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, *64*(2), 113-128. doi:10.1007/BF02294531