Revisiting a global burnout score with the Burnout Assessment Tool across nine country samples

Accepted manuscript version (after peer review):

De Beer, L.T., Schaufeli, W.B., De Witte, H., Hakanen, J.J., Kaltiainen, J., Glaser, J., Seubert, C., Shimazu, A., Bosak, J., Procházka, J., Kajzar, A. & Christensen, M. (2024). Revisiting a global burnout score with the Burnout Assessment Tool across nine country samples. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000839

Revisiting a global burnout score with the Burnout Assessment Tool across nine country samples

Abstract

Studies published on the validity of the Burnout Assessment Tool (BAT), a novel burnout instrument, have gained traction in the literature over recent years. The BAT has been successfully shown to be equivalent across representative samples when modelled as a second-order/higher-order model. However, this specification is not free of criticism and the bifactor approach has been presented as the alternative model specification. Therefore, a study investigating the construct-relevant multidimensionality of the BAT across many representative samples is warranted to reassess a global burnout factor (n = 9041). We implemented bifactor exploratory structural equation modelling to ascertain the relevance of a global burnout factor and specific component factors (bifactor-ESEM). According to the standardised loadings and McDonald's omega coefficients, the results showed that the bifactor-ESEM model had a strong global burnout factor with relevant specific factors beyond the global factor. The model also showed measurement invariance across countries and genders. We also present a figure that compares the global burnout mean scores of the countries. All in all, the results of this study reaffirmed that BAT-assessed burnout can be modelled with an equivalent global burnout score across conditions.

Keywords: burnout, measurement invariance, equivalence, bifactor exploratory structural equation modelling, occupational depression

The measurement of burnout has not been without controversy across the last few decades. However, this has just been one issue of concern in burnout research over recent years, especially as pertaining to the Maslach Burnout Inventory (MBI; Maslach & Jackson, 1981) (De Beer et al., 2024). Other issues include the very conceptualisation of burnout and the associated prevalence rates of burnout in the absence of globally accepted clinical criteria (Bianchi et al., 2021) and the discouragement (or inability) to present a global burnout score even though it is presented as a syndrome (World Health Organization, 2019).

To address these issues, Schaufeli et al. (2020) created the Burnout Assessment Tool (BAT) as a novel measure of burnout based on an *updated* conceptualisation of the phenomenon. Schaufeli and colleagues' approach was unique as the Dutch Occupational Health Authorities recognise burnout as an occupational disease and professionals are well-trained to distinguish between the occupational complaints that employees present (e.g., Verschuren, 2010). This allowed for an inductive approach where some of these independent health professionals could be interviewed to ascertain how burnout presents itself—coupled with a review of 300 potential burnout items that were then used in a deductive phase, made for a robust approach. Burnout received an updated definition of "a work-related state of exhaustion that occurs among employees, which is characterised by extreme tiredness, reduced ability to regulate cognitive and emotional processes, and mental distancing" (Schaufeli et al., 2020, p. 40). A total of 23 items were selected measuring four core components of burnout – including two well-known components, exhaustion and mental distance (Schaufeli & Taris, 2005), and two novel components that closely align with exhaustion, cognitive impairment and emotional impairment. The specific addition of cognitive impairment is essential as research has long held that cognitive problems are a hallmark of burnout (Deligkaris et al., 2014).

Particularly noteworthy about the BAT is its ability and instruction to model burnout as a total score, which users of the gold standard MBI are explicitly discouraged from doing (Maslach et al., 2017). A recent study by De Beer et al. (2020), employed six representative samples from Europe and one from Japan, modelled BAT-assessed burnout as a second-order factor and found scalar (strong) measurement equivalence across countries. However, the use of second-order factor model specifications is not without criticisms, as (for example) items relate only to the second-order factor indirectly through the first-order factors, which could result in a loss of some information (see Gignac, 2016; Morin, 2023; Morin et al., 2016). Bifactor models are considered the alternative hierarchical structure to test as the global and specific factors directly explain item variance and not indirectly as with second-order models.

Subsequently, in this report, we use bifactor exploratory structural equation modelling (bifactor-ESEM) to disaggregate the variance into a global burnout factor with the four specific factors of the BAT. Bifactor-ESEM is also more flexible compared to standard bifactor modelling as it allows for the relaxation of the assumption of precisely zero cross-loadings to approximately zero cross-loadings (Morin, 2023) – a much more realistic assumption when one considers context- and potential wording-effects. Indeed, recent studies have shown that not only is the bifactor structure of the BAT the best fit, but the correlations compared between the normal four-factor CFA model and the four-factor ESEM model also reduce, revealing the presence of an underlying global factor (Morin, 2023).

Therefore, the purpose of this report is to revisit the validity and measurement invariance of a global burnout score with BAT-assessed burnout across nine largely representative samples (Austria, Belgium, Czech Republic, Finland, Germany, Ireland, Japan, the Netherlands, and Norway) using the bifactor-ESEM approach. We also inspect the level of global burnout risk in each country and across gender. The findings of this report can contribute to the refinement of the definition of burnout, its measurement, and how to create a global burnout score in practice.

Methods

Participants and procedures[1]

[1] For a more complete sample description please see the supplementary material (p. S2). Furthermore, the samples from Austria, Belgium, Germany, Ireland, Japan, and the Netherlands were used from De Beer et al. (2020).

The study comprised data from nine countries (n = 9041). Each sample was sampled with the aim to be largely representative of each country at least by gender using a quota sampling procedure. Surveys were conducted in the most spoken language of each specific country (e.g., Austria and Germany in the German language; Belgium and The Netherlands in Dutch). In Austria (n = 1059), the average age was 42.98 years (SD = 13.32), with a gender distribution of 50.10% male and 49.90% female. In Germany (n = 1073), the average age was 41.79 years (SD = 13.14), and the gender split was 51.50% male and 48.50% female. Belgium's Flanders region (n = 1500) had an average age of 41.37 years (SD = 11.46) with a 54.30% male and 45.70% female distribution. Finland (n = 703) had participants averaging 48.13 years in age (SD = 9.99) and were 49.22% male and 50.78% female. In Ireland (n = 431), the average age was 42.10 years (SD = 12.30), with a gender distribution of 53.60% male and 46.40% female. Japan (n = 1032) had an average age of 40.24 years (SD = 11.69) and an even gender split of 50% male and 50% female. The Netherlands (n = 1500) had an average age of 41.26 years (SD = 13.36) and a gender split of 54.10% male and 45.90% female. The Czech Republic (n = 1020) showed an average age of 41.82 years (SD = 11.65) and a gender split of 53.90% male and 46.10% female. Finally, Norway (n = 493) had an average age of 45.55 years (SD = 11.54) with a gender distribution of 49.54% female and 50.46% male.

Measures

Burnout was measured with the BAT-23 (Schaufeli et al., 2020). The BAT-23 is considered the core of burnout and comprises 23 items measuring: exhaustion with eight items, and mental distance, cognitive impairment, and emotional impairment with five items each. All items are measured on a 5-point scale ranging from 1 (Never) to 5 (Always).

Data analysis

We used latent variable modelling techniques in Mplus 8.10 (Muthén & Muthén, 2023). The bifactor-ESEM approach was used (Morin, 2023; Morin et al., 2020), and we specifically used mean- and variance-adjusted weighted least squares estimation (Millsap, 2011). First, we estimated the model in each country separately (see Table S1), and then we tested the measurement invariance between countries and gender (see Table 1). We deferred to the standard fit metrics to consider the fit of the models: CFI and TLI above 0.90; RMSEA and SRMR below 0.08 (Hoyle, 2023). For measurement invariance, we tested full measurement invariance by considering increasingly constrained models in the following order: configural (factor structure), weak/metric (loadings), strong (thresholds), and strict (item uniquenesses) (Putnick & Bornstein, 2016). Measurement invariance was considered to hold if the CFI did not deteriorate by -0.01 or more and if the RMSEA did not deteriorate by more than +0.015 (Chen, 2007) in each subsequent model.

Results

Table S1 and S2 in the supplementary material present the bifactor-ESEM model's fit statistics in each country and gender. All models showed acceptable fit metrics, surpassing the more stringent CFI and TLI > .950 cut-off criteria in every instance.

Given that the model fitted in each country, we moved on to multigroup measurement invariance and found full (strict) invariance. We also tested an additional step of invariance: latent variance-covariance invariance. Indeed, latent variance-covariance invariance was also apparent without any concern and this model was preferred for further reporting. Gender also showed all these levels of invariance up to and including latent variance-covariance invariance.

-INSERT TABLE 1 HERE-

We used the latent variance-covariance model parameters for both countries and gender to estimate McDonald's omega coefficient (ω). Supplementary Tables S3 and S4 provide the details, but the global factor was very well-defined in the countries ($\lambda \geq .664$; $\omega = .981$) and gender ($\lambda \geq .665$; $\omega = .979$) with the specific factors all retaining meaningful specificity beyond the global factor (ω's > .700), but with no loadings exceeding the value of the corresponding item on the global factor – once again indicating the strength of the global factor.

Figure 1 below shows the visual presentation of the observed global burnout mean scores in each country – having achieved strict and latent variance-covariance a comparison of observed scores become possible and are more easily comparable with future studies than latent means. We added dashed grey lines to indicate overlaps in confidence intervals – indicating no meaningful difference between scores. Most European countries scored similarly, and where there was deviation, these were small. However, the Japanese sample scored somewhat higher compared to the Western countries. Regarding gender, no statistically significant mean difference was present for the global burnout score.

-INSERT FIGURE 1 HERE-

Discussion

This report investigated and demonstrated the validity and measurement invariance of a BAT-assessed global burnout score, estimated with bifactor-ESEM across nine largely representative country samples. Specifically, it was found that the BAT presents a strong global score, with the four specific factors remaining as meaningful factors. Previous research has shown with most of these countries, that a second-order total score is also valid (De Beer et al., 2020). Therefore, even given the criticisms against the second-order model, we believe the BAT can function as a second-order factor due to the validity evidence surrounding that specification. Subsequently, comparing the substantive differences from the conclusions from De Beer et al. (2020) to this study shows similarities and no severe deviations, indicating the robustness of the global burnout score across models. The Japanese sample also showed the highest level of burnout risk. All in all, the results from this report support the validity of the BAT. Specifically, that the BAT can be modelled as a total score in line with a syndrome comprising the four components of BAT-defined burnout, something other measures of burnout, such as the MBI, cannot do. The level of invariance also indicates that observed scores can be compared between countries and genders.

This study is not without limitations. First, cross-sectional data was used; therefore, no test-retest reliability or stability investigations over time could be made. Second, these samples were not collected at the same points in time, which could affect the intensity of burnout levels. For example, the Norwegian sample is the most recent and the effects of the pandemic, the energy crisis, and the war in Ukraine may have impacted these employees and could explain the slightly elevated burnout score.

Future studies should focus on the predictive validity of the BAT, that is, how the BAT, as a self-report measure, may predict real-world outcomes such as sickness absence and costs in an organisation, also longitudinally. Moreover, the parameters of the latent variance-covariance model are presented in the supplementary material, which could be used as starting values or priors (Bayesian) in future BAT models with smaller samples. For example, these sets of priors could be used with a small and/or medium variance (uncertainty) to compare for sensitivity to make substantive conclusions against the uninformed prior model or against another set of relevant priors.

Conclusion

The BAT is a valid and reliable instrument to measure burnout complaints and model the phenomenon as a global score within organisations. The supplementary material contains an example Mplus script to model the BAT with bifactor-ESEM.

References

Bianchi, R., Verkuilen, J., Schonfeld, I. S., Hakanen, J. J., Jansson-Fröjmark, M., Manzano-García, G., ... & Meier, L. L. (2021). Is burnout a depressive condition? A 14-sample meta-analytic and bifactor analytic study. *Clinical Psychological Science, 9*(4), 579-597. https://doi.org/10.1177/2167702620979597

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504. https://doi.org/10.1080/10705510701301834

De Beer, L. T., & Morin, A. J. S. (2022). *(B)ESEM invariance syntax generator for Mplus.* https://statstools.app/b_esem/ https://doi.org/10.6084/m9.figshare.19360808

De Beer, L. T., Schaufeli, W. B., De Witte, H., Hakanen, J. J., Shimazu, A., Glaser, J., Seubert, C., Bosak, J., Sinval, J., & Rudnev, M. (2020). Measurement invariance of the Burnout Assessment Tool (BAT) across seven cross-national representative samples. *International Journal of Environmental Research and Public Health, 17*(15), 5604. https://doi.org/10.3390/ijerph17155604

De Beer, L.T., van der Vaart, L., Escaffi-Schwarz, M., De Witte, H., & Schaufeli, W.B. (2024). Maslach Burnout Inventory – General Survey: A systematic review and meta-analysis of measurement properties. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000797

Deligkaris, P., Panagopoulou, E., Montgomery, A. J., & Masoura, E. (2014). Job burnout and cognitive functioning: A systematic review. *Work & Stress, 28*(2), 107-123. https://doi.org/10.1080/02678373.2014.909545

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. https://doi.org/10.1016/j.intell.2016.01.006

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior, 2*(2), 99-113. https://doi.org/10.1002/job.4030020205

Maslach, C., Leiter, M.P., & Jackson, S. E. (2017). *Maslach Burnout Inventory Manual* (4th ed.). Mind Garden.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Taylor & Francis.

Morin, A. J. S. (2023). Exploratory structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed.) (pp. 503-524). Guilford.

Morin, A.J.S., 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, 23*, 116-139. https://doi.org/10.1080/10705511.2014.961800

Morin, A. J. S., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM) and bifactor-ESEM. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of Sport Psychology* (pp. 1044-1073). https://doi.org/10.1002/9781119568124.ch51

Muthén, L., & Muthén, B. (2023). Mplus User's Guide, Version 8.10. Muthén & Muthén.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90. https://doi.org/10.1016/j.dr.2016.06.004

Schaufeli, W. B., Desart, S., & De Witte, H. (2020). Burnout Assessment Tool (BAT)—development, validity, and reliability. *International Journal of Environmental Research and Public Health, 17(24), Article 9495*. https://doi.org/10.3390/ijerph17249495

Schaufeli, W. B., & Taris, T. W. (2005). The conceptualisation and measurement of burnout: Common ground and worlds apart. *Work & Stress, 19*(3), 256-262. https://doi.org/10.1080/02678370500385913

Verschuren, C. (2010). Eén lijn in de eerste lijn bij overspanning en burnout. *Psychopraktijk, 2*(6), 27-31. https://doi.org/10.1007/s13170-010-0091-0

World Health Organization (2019). Burnout QD85. In *International statistical classification of diseases and related health problems* (11th ed.). https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/129180281

Open Science Statements

Open Science: We confirm that there is sufficient information for an independent researcher to reproduce all the reported methodology. We report how we determined our samples, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. For all psychometrically relevant inferential tests and parameters, we report exact p-values, effect sizes, and 95% confidence or credible intervals.

Open Data: The data for this study can be requested from the corresponding author, who will consider all reasonable requests.

Open Materials: The information needed to reproduce of the reported methodology is available online at https://doi.org/10.17605/OSF.IO/5YF3T

Open Analytic Code: I confirm that the scripts, code and output needed to reproduce the results are provided online at https://doi.org/10.17605/OSF.IO/5YF3T

Preregistration of Studies and Analysis Plans: This study was not preregistered.

Table 1

Results of the BESEM Measurement Invariance Testing for Country and Gender

Country	χ^2	**d*f***	**CFI**	**TLI**	**RMSEA**	**CM**	**ΔCFI**	**ΔRMSEA**
M1: Configural	5138.139	1332	.992	.987	.053 [.052, .055]	-		
M2: Weak / Metric (λ)	9124.067	2052	.986	.984	.059 [.057, .060]	M1	-.006	+.006
M3: Strong (λ, ν)	10988.549	2564	.983	.985	.057 [.056, .058]	M2	-.003	-.002
M4: Strict (λ, ν, δ)	12962.84	2748	.980	.983	.061 [.060, .062]	M3	-.003	+.004
M5: Latent variance-covariance (λ, τ, δ, ξ/φ)	14613.445	2868	.977	.982	.064 [.063, .065]	M4	-.003	+.003
Gender	χ^2	**d*f***	**CFI**	**TLI**	**RMSEA**	**CM**	**ΔCFI**	**ΔRMSEA**
N1: Configural	3147.767	296	.993	.988	.046 [.045, .048]	-		
N2: Weak / Metric (λ)	2375.638	386	.995	.993	.034 [.032, .035]	N1	+.002	-.012
N3: Strong (λ, τ)	2137.615	450	.996	.995	.029 [.028, .030]	N2	+.001	-.005
N4: Strict (λ, τ, δ)	1710.32	473	.997	.997	.024 [.023, .025]	N3	+.001	-.005
N5: Latent variance-covariance (λ, τ, δ, ξ/φ)	779.583	493	.999	.999	.011 [.010, .013]	N4	+.002	-.013

Note: χ^2 = robust chi-square; df = degrees of freedom; CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = Root mean square error of approximation; λ = factor loadings; τ = thresholds; δ = uniquenesses; CM = comparison model; ΔCFI = change in CFI; ΔRMSEA = change in RMSEA

Figure 1

The Global Burnout Scores Across Countries with 95% Confidence Intervals