PSYCHOLOGY & HEALTH
https://doi.org/10.1080/08870446.2024.2321358

Routledge
Taylor & Francis Group

OPEN ACCESS

Check for updates

The burnout-depression conundrum: investigating construct-relevant multidimensionality across four countries and four patient samples

Leon T. De Beer[a,b], Jari J. Hakanen[c], Wilmar B. Schaufeli[d,e], Hans De Witte[d,f], Jürgen Glaser[g], Janne Kaltiainen[c], Christian Seubert[g] and Alexandre J. S. Morin[h]

[a]Department of Psychology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway; [b]WorkWell Research Unit, North-West University, Potchefstroom, South Africa; [c]Workability and Work Careers, Finnish Institute of Occupational Health, Helsinki, Finland; [d]Research Unit Occupational & Organizational Psychology and Professional Learning, KU Leuven, Leuven, Belgium; [e]Department of Psychology, Utrecht University, Utrecht, The Netherlands; [f]Optentia Research Unit, Vanderbijlpark Campus, North-West University, Vanderbijlpark, South Africa; [g]Department of Psychology, University of Innsbruck, Innsbruck, Austria; [h]Substantive Methodological Synergy Research Laboratory, Concordia University, Montreal, Canada

ABSTRACT

This research seeks to contribute to the ongoing discussion about the distinctive nature of burnout and depression. In a first study, we relied on employee samples from four European countries (N=5199; 51.27% women; M_{age} = 43.14). In a second study, we relied on a large sample of patients (N=5791; 53.70% women; M_{age} = 39.54) who received a diagnosis of burnout, depressive episode, job strain, or adaptation disorder. Across all samples and subsamples, we relied on the bifactor exploratory structural equation modelling to achieve an optimal disaggregation of the variance shared across our measures of burnout and depression from the variance uniquely associated with each specific subscale included in these measures. Our results supported the value of this representation of participants' responses, as well as their invariance across samples. More precisely, our results revealed a strong underlying global factor representing participants' levels of psychological distress, as well as the presence of equally strong specific factors supporting the distinctive nature of burnout and depression. This means that, although both conditions share common ground (i.e. psychological distress), they are not redundant. Interestingly, our results also unexpectedly suggested that suicidal ideation might represent a distinctive core component of depression.

ARTICLE HISTORY
Received 9 August 2023
Accepted 15 February 2024

KEYWORDS
Burnout; depression; psychological distress; job strain; Burnout Assessment Tool (BAT)

Introduction

Burnout was recently identified as one of the leading occupational diseases in the Netherlands (Weel, 2021), and even though evaluation criteria may differ, some form of 'burnout syndrome may be acknowledged as an occupational disease' in at least

CONTACT Leon T. De Beer leondb@gmail.com Department of Psychology, Norwegian University of Science and Technology (NTNU), Dragvoll Campus, Edvard Bulls veg 1, Trondheim, 7491, Norway; WorkWell Research Unit, North-West University, Potchefstroom, South Africa

Supplemental data for this article can be accessed online at https://doi.org/10.1080/08870446.2024.2321358.

eight other European countries (Lastovkova et al., 2018, p. 160). Effective 1 January 2022, The World Health Organization [WHO] (2019a) recognized burnout in the 11th revision of the International Classification of Diseases (ICD-11), as an occupational phenomenon defined as: '… a syndrome … resulting from chronic workplace stress that has not been successfully managed'. According to this definition, burnout is seen as encompassing: (i) feelings of energy depletion or exhaustion; (ii) increased mental distance from one's job, or feelings of negativism or cynicism related to one's job; and (iii) a reduced sense of professional efficacy (WHO, 2019a). This definition matches the classical definition of burnout proposed by Maslach and colleagues, as operationalized in the Maslach Burnout Inventory (MBI; Maslach & Jackson, 1981; Maslach & Leiter, 2016; Maslach et al., 2017).

However, even though the MBI has long served as the gold standard in burnout measurement, several problems were recently identified in relation to this measure and its operationalization. For instance, (i) the utility of reduced professional efficacy as a core component of burnout has been seriously challenged (Bresó et al., 2007; De Beer & Bianchi, 2019; Sandrin et al., 2022; Schaufeli & Taris, 2005), (ii) the neglect of other, arguably critical, manifestations of burnout such as cognitive impairment has been highlighted (Deligkaris et al., 2014; Schaufeli et al., 2020), (iii) the lack of proper cut-off scores (those proposed over time are now mainly outdated) established with representative samples and lack of proper nosological representation as a standalone diagnostic category have been highlighted as a severe impediment to its diagnostic use (Bianchi et al., 2013, 2015b, 2017b; Nadon et al., 2022; Schaufeli et al., 2020), (iv) some researchers have expressed concern about the inconsistent and arbitrary use of different factor structures (like one-, two-, or three-factor specifications) to represent burnout across studies, suggesting that these structures might be selected more to match researchers' objectives rather than to accurately reflect the true nature of burnout (Nadon et al., 2022; Worley et al., 2008), and (v) the MBI was never designed as a diagnostic tool (Maslach & Leiter, 2021). This is again exemplified by the fact that—according to the MBI manual (Maslach et al., 2017)—the MBI does not produce, and should not be used to produce, a single burnout score. However, rather than following these recommendations and relying on three distinct subscale scores, many studies have combined these scores to obtain a global estimate of burnout severity. This ignorance of formal recommendations illustrates the need for a single burnout score. Especially now that burnout has received recognition as a potential diagnostic category in some European countries (Lastovkova et al., 2018), it is important for epidemiologists and occupational health practitioners to be able to assess the prevalence of 'burnout' in and of itself, rather than as a combination of disparate components.

Beyond these operational considerations, an evolving body of research has also questioned, and investigated, the potential conceptual overlap between burnout and depression (e.g. Bianchi et al., 2015a, 2021; Schonfeld & Bianchi, 2016). According to the WHO (2019b), a depressive disorder is characterized by 'depressive mood (e.g. sad, irritable, empty) or loss of pleasure accompanied by other cognitive, behavioural, or neurovegetative symptoms that significantly affect the individual's ability to function'. Albeit typically seen as differing in terms of the context in which they occur (i.e. burnout is typically seen as work-specific whereas depression encompasses all spheres of life), both tend to spread out to all spheres of life and are highly correlated (Bianchi et al., 2015a; Glass & McKnight, 1996; Heinemann & Heinemann, 2017; Nadon et al., 2022). Ahola et al. (2014) showed that burnout and depressive symptoms develop in tandem over time through a person-centered approach. More precisely, their study initially identified three types of participants displaying varying levels (low, medium, high) of burnout and depressive symptoms, which evolved over time into four distinct trajectories. These trajectories included participants with consistently low or high levels of symptoms, as well as those experiencing increasing or decreasing symptoms over time. Based on results such as these, some have suggested that burnout may be nothing more than a depression emerging in the work context (Bianchi et al., 2017a, 2021; Nadon et al., 2022).

In contrast, burnout proponents typically argue that recasting burnout as depression would contribute to absolving organizations from their own role in the emergence of burnout (Epstein & Privitera, 2017) and would preclude further investigations of differences (Meier & Kim, 2022). Yet, the fact that both phenomena result from a complex biopsychosocial aetiology encompassing similar individual (e.g. biological and psychological factors) and social (e.g. higher levels of demands) characteristics, also suggests more similarities than differences (Bianchi et al., 2017a). However, alternative evidence also supports their distinctive aetiology (Koutsimani et al., 2019). For instance, recent evidence showcased the discriminant validity of both constructs (Tóth-Király et al., 2021; but also see Ahola et al., 2014; Hakanen & Schaufeli, 2012), demonstrating that: (i) whereas burnout is multidimensional, depression is best represented as unidimensional; (ii) both constructs share reciprocal associations over time, while remaining distinct; (iii) both constructs share well-differentiated associations with covariates in a way that was consistent with their contextual nature.

The current state of research thus leaves open the questions of whether burnout and depression are truly distinct states and whether burnout truly deserves consideration as a construct distinct from depression. Answering these questions seems to be further complicated by the various inadequacies associated with using the MBI as the gold-standard for burnout assessment. Resolving these issues is critical from a practical perspective given the emerging need for practicians to be able to differentially diagnose these two conditions as well as to support clinical and occupational research designed to uncover optimal, specific interventions to address both conditions, either jointly or separately. The present study addresses these issues by relying on an improved measure of burnout (i.e. the Burnout Assessment Tool [BAT]; Schaufeli et al., 2020) and state-of-the-art statistical modelling strategies (i.e. bifactor exploratory structural equation modeling [ESEM]; Morin et al., 2016a, 2016b, 2020), including healthy as well as patient samples.

Defining and operationalizing burnout

Over the years, various definitions of burnout have been presented. Recently, a 'harmonized' definition has been proposed which reduces burnout to merely exhaustion (Guseva Canu et al., 2021). However, Schaufeli (2021) responded that, despite the central role of exhaustion, burnout has always been conceptualized as encompassing more than just exhaustion, highlighting, for instance, that withdrawal (mental distancing) has been seen as a critical component of burnout since Freudenberger's (1974) early identification of this work-related phenomenon.

To address the various criticisms leveraged at the MBI and other instruments, Schaufeli et al. (2020) proposed the BAT as a novel, more comprehensive approach to burnout measurement. Rather than following procedures akin to those previously used to create burnout measures anchored in tradition and in early unstructured observations made by Freudenberger (1974) and Maslach (1976), Schaufeli et al. (2020) developed the BAT following a deductive quantitative methodology combined with an inductive approach based on interviews with Dutch and Flemish health practitioners with experience in working with burned-out employees. This method was made possible by the unique context of the Netherlands, where burnout is officially recognized as an occupational disease. As a result, Dutch professionals are uniquely experienced in categorizing psychologically distressed employees as suffering either from job strain, burnout, adaptation disorder, or depression.

The BAT relies on a definition of burnout as: 'a work-related state of exhaustion that occurs among employees, which is characterized by extreme tiredness, reduced ability to regulate cognitive and emotional processes, and mental distancing' (Schaufeli et al., 2020, p. 4). This definition encompasses four interrelated dimensions (exhaustion, mental distance, cognitive impairment, and emotional impairment) which can be combined into a single global severity score (Hadzibajramović et al., 2020, 2022). Recent research has supported the psychometric properties of the BAT as a robust measure of burnout that generalizes across countries and languages (De Beer et al., 2020; Schaufeli et al., 2020; Schaufeli & De Witte, 2023). However, despite their interest, these previous studies have failed to completely consider the construct-relevant psychometric multidimensionality likely to be present in BAT scores (Morin et al., 2016a, 2016b, 2020), as discussed below.

Construct-relevant psychometric multidimensionality and bifactor-ESEM

Modern developments in latent variable modeling have highlighted the need to account for two distinct sources of construct-relevant psychometric multidimensionality (i.e. when items shared a true association with more than one construct) in complex measurement instruments such as the BAT. Relative to confirmatory factor analytic (CFA) models, which assume that cross-loadings between items and non-target factors will be exactly zero, ESEM allows for the free estimation of the cross-loadings likely to occur when assessing conceptually related constructs due in part to the fallible nature of most questionnaire indicators (Morin et al., 2016a, 2016b, 2020).

Statistical research has shown that, whereas excluding cross-loadings from a model resulted in biased estimates of factor correlations and regressions, including unnecessary cross-loadings did not interfere with the ability to obtain accurate parameter estimates (Asparouhov et al., 2015; Mai et al., 2018; Morin et al., 2016a). Moreover, ESEM does not preclude the reliance on an a priori specification of the main indicators of each factor when implemented using target rotation, a confirmatory from of rotation procedure (Morin et al., 2020).

Beyond the assessment of conceptually related constructs, the BAT also assumes that ratings can be used to reflect both scores on the four specific subscales, as well as a global burnout score, in line with the formulation of burnout as a syndrome (Schaufeli et al., 2020). This second form of construct-relevant psychometric multidimensionality calls for bifactor models. In a bifactor model, ratings on all items included in an instrument are directly used to estimate a global factor (G-factor), as well as specific-factors (S-factors) reflecting the variance uniquely shared among all items associated to each subscale beyond that already explained by the global factor (Morin et al., 2016a, 2016b, 2020). The bifactor-ESEM framework combines both possibilities.

Research has supported the relevance of a bifactor (CFA or ESEM) representation of burnout as measured by multiple instruments in a variety of contexts (Armon et al., 2012; Barcza-Renner et al., 2016; Doherty et al., 2021; Hawrot & Koniewski, 2018; Isoard-Gautheur et al., 2018; Mészáros et al., 2014; Sandrin et al., 2022; Szigeti et al., 2017). However, beyond the ability to achieve a more accurate representation of burnout, the bifactor-ESEM framework also provides a way to empirically address the conceptual overlap between burnout and depression. Indeed, due to the way construct-relevant variance is separated in bifactor-ESEM, it becomes possible to directly assess whether any specificity remains associated with distinct specific dimensions once the variance explained by what they share (i.e. the G-factor) is taken out of these ratings (Arens & Morin, 2017, Morin et al., 2020).

For present purposes, the application of this framework to conceptually related measures of depression and burnout would make it possible to directly estimate a G-factor reflecting participants' overarching levels of psychological distress estimated from their ratings of both measures. Beyond this global factor, the strength (i.e. factor loadings, composite reliability) of the S-factors estimated as part of this model would also directly indicate whether something unique remains associated with each subscale, including the measure of depression. Thus, observing that depression ratings only contribute to define the psychological distress G-factor and remain associated with an 'empty' S-factor would argue for the overlapping nature of both constructs. In contrast, observing that the depression S-factor remains defined by satisfactory factor loadings (≥ .500) and associated with a satisfactory level of composite reliability would support the distinctive nature of both constructs.[1]

Bifactor-ESEM is thus specifically designed to help distinguish between the global aspects of these psychological states and their more specific manifestations. This approach is thus well-suited to investigations of the extent to which symptoms of burnout or depression both capture general feelings of psychological distress relative to distinctive, or unique, manifestations. Moreover, bifactor-ESEM, *via* cross-loadings, acknowledges that multiple survey items are likely to share associations with more than one construct. For example, a burnout item could also provide some insights into depression, and vice versa. This recognition of the complexity and overlap in human emotions and experiences allows for a more accurate and nuanced understanding of psychological distress—something that classical CFA is unable to do. For additional details on bifactor-ESEM, interested readers are referred to Morin (2023).

The current studies

The aim of the present series of two studies is to investigate the construct-relevant psychometric multidimensionality of burnout (measured using the BAT) and depression (measured using the depression subscale of the four-dimensional symptom questionnaire [4DSQ]) through the application of bifactor-ESEM analyses. In both studies, we rely on this approach to identify the optimal factor structure for this combination of measures by contrasting CFA, bifactor-CFA, ESEM, and bifactor-ESEM solution. This comparison will allow us to verify whether both measures contribute to the assessment of a general psychological distress G-factor, and whether each subscale included in these analyses retain a meaningful amount of specificity beyond the assessment of this G-factor. In Study 1, we also consider the cross-cultural (Belgium, Germany, Austria, Finland) and cross-linguistic (Dutch, German, and Finnish) generalizability of these conclusions by contrasting the results obtained among representative samples of participants recruited in four European countries.

In Study 2, we further assess the extent to which these conclusions generalize across four samples of patients recruited in the Netherlands based on their classification by professionals working for the Dutch occupational health authority: (i) job strain, (ii) burnout, (iii) depressive episode, or (iv) adaptation disorder (Verschuren, 2010). Whereas the former two categories reflect clinically significant feelings of psychological distress linked to the work area that vary in severity (burnout being a more severe clinical state than job strain), the latter two categories reflect clinically significant feelings of psychological distress that are not specific to the work context and vary also in severity (depressive episode being a more severe clinical state than adaptation disorder). The reliance on patient samples is novel and critically important for two reasons. First, whereas community samples typically include a majority of participants with relatively low scores on measures of burnout and depression, the consideration of patient samples makes it possible to test whether conclusions based on a relatively lower range of scores can generalize to higher scores. Second, contrasting these four patient samples makes it possible to test the discriminant validity of our ratings through tests of latent mean differences. Indeed, given that scores on the general psychological distress G-factors will be estimated using both instruments, we expect higher scores on this factor among the most clinically impaired patient samples (depressive episode and burnout) rather than among the less clinically impaired samples (adaptation disorders and job strain). Scores on the various burnout S-factors should themselves be higher in the burnout sample than in the depressive episode sample, whereas those on the depression S-factor should be higher in the depressive episode sample than in the burnout sample.

Study 1

Method

Participants and procedure

This study relies on a combined sample of 5199 participants (51.27% women, 16–79 years; M_{age} = 43.14; SD_{age}= 12.28) across all four countries. All data were collected with online questionnaires in either German, Dutch or Finnish languages. The first sample (data collected in December 2018) includes 1059 Austrian employees and is representative of the Austrian working population in terms of age (M_{age} = 42.98; SD_{age} = 13.32) and sex (49.90% women). The second sample (data collected in November 2017) includes 1500 Belgian employees and is representative of the Flemish working population in terms of age (M_{age}

= 40.90; SD_{age} = 11.60), sex (44.10% women), and economic sector as provided by STATBEL (http://statbel.fgov.be). The Belgian project was approved by the relevant research ethics committee (Reference number: G-2015 10 353) and the data collected by iVox. The third sample (data collected in December 2018) includes 1073 German employees and is representative of the German working population in terms of age (M_{age} = 41.79; SD_{age} = 13.14) and sex (48.50% women). The German and Austrian samples were both collected as part of the same project by Bilendi and approved by the research ethics committee of the University of Innsbruck (Certificate of good standing reference number: 64/2020). The fourth sample (data collected with an online survey between December 2020 and January 2021 by the Finnish Institute of Occupational Health) includes 1567 Finnish employees (M_{age} = 45.80; SD_{age} = 10.98; 59.50% women). The Finnish data collection was approved by the Ethical Review Committee of the Finnish Institute for Occupational Health (Reference number: 7/2019). For this sample, we relied on sampling weights based on age, gender, and residential area in our analyses to match the Finnish population distribution. Therefore, our samples were broadly representative of age and gender within each country.

Measures

Burnout was assessed with the original Burnout Assessment Tool (BAT-23; Schaufeli et al., 2020). This instrument relies on 23 items to measure the core of burnout: Exhaustion (8 items; e.g. 'When I get up in the morning, I lack the energy to start a new day at work'; α = .914), mental distance (5 items; e.g. 'I feel indifferent about my job'; α = .892), cognitive impairment (5 items; e.g. 'At work, I struggle to think clearly'; α = .909), and emotional impairment (5 items; e.g. 'At work, I may overreact unintentionally'; α = .885). All items were measured on a 5-point scale ranging from 1—'Strongly disagree' to 5—'Strongly agree' and can be used to obtain a global burnout score (α = .953). The BAT-23 has been shown to be invariant across European countries (De Beer et al., 2020).

Depression was assessed with the corresponding subscale from the Four-Dimensional Symptom Questionnaire (4DSQ; Terluin, 1994; Terluin et al., 2004; Kleinstäuber et al., 2021). In the Netherlands, the Royal Dutch Medical Association recommends this questionnaire for use by (occupational) health practitioners, including general practitioners, to distinguish mental health complaints reported by employees. Consequently, the 4DSQ is well-established in the Dutch occupational health system (Terluin et al., 2004). This depression subscale measures aspects of depressive cognitions, suicidal thoughts, and anhedonia symptoms forming a single factor (Kleinstäuber et al., 2021). The six items (e.g. 'Did you feel that you can't enjoy anything at all?'; α = .922) are scored from 1—'No' to 5—'Very often or constantly'. The subscale does not contain specific fatigue-related items, but research has shown that even when removing fatigue-related items from other depression scales had limited to no impact on the association between depression and burnout (Bianchi et al., 2021).

Analyses

All analyses were conducted using Mplus 8.8 (Muthén & Muthén, 2022) robust weighted least square estimator with mean and variance adjusted statistics (WLSMV) to account for the ordinal nature of the rating scales used in this study (Finney &

DiStefano, 2013). An important advantage of this estimator for psychometric investigations is that it provides a closer approximation of participants' true response process by modeling the exact thresholds at which their response change from one category to the other for each item rather than a single intercept for each item (e.g. Freund et al., 2013). Although WLSMV is slightly less efficient at handling missing responses than maximum likelihood-based estimators (Asparouhov & Muthén, 2010), this limitation is negligible in this study due to the very limited number of missing responses at the item level ($M = 0.06\%$).

In each country, we contrasted the following four alternative representations of participants' responses to the BAT-23 and to the depression subscale of the 4DSQ: CFA, bifactor-CFA, ESEM, and bifactor-ESEM. For the CFA solution, a five-factor model was specified in which all items were only allowed to represent their a priori factor, all factors were allowed to correlate with one another, and no cross-loading or correlated uniqueness was included. In the ESEM solution, the same five-factor model was specified using a confirmatory form of oblique rotation (i.e. target rotation; Morin et al., 2016a, 2020). This rotation procedure allowed us to explicitly indicate the key indicators of each factor (as in the CFA solution), while allowing all cross-loadings to be freely estimated but 'targeted' to be as close to zero as possible. The bifactor-CFA solution simply added a global factor (G-factor; psychological distress) to the previous CFA solution, allowing this G-factor to be defined by all items. The items thus retained their associations on their a priori factors (S-factors) which came to reflect the specificity explained by each subscale beyond that already explained by the G-factor (Morin et al., 2016a, 2020). In this solution, all factors were specified to be orthogonal (not correlated) according to typical bifactor specifications, which is a prerequisite for the interpretation of the G- and S-factors as substantively meaningful (Morin, 2023, Morin et al., 2020). Finally, the bifactor-ESEM solution combined the factor definition of the bifactor-CFA and the free estimation of all cross-loadings (targeted to be as close to zero as possible *via* an orthogonal bifactor target rotation procedure).

The optimal solution was then retained for formal tests of measurement invariance across countries, conducted in the following sequence (Millsap, 2011): (i) configural (same model), (ii) weak (equality of loadings), (iii) strong (equality of loadings and thresholds), (iv) strict (equality of loadings, thresholds, and uniquenesses); (v) latent variance-covariance (equality of loadings, thresholds, uniquenesses, and the latent variance-covariance matrix), and (vi) latent mean (equality of loadings, thresholds, uniquenesses, the latent variance-covariance matrix, and latent means) (see Millsap, 2011). To minimise the potential for human errors, the syntax used for the estimation of this model was generated using the code generator created specifically for multigroup invariance tests with (bifactor) ESEM models by De Beer and Morin (2022). This code generator automatically handles the calculation of chi-square difference tests for WLSMV using Mplus DIFFTEST function.

Given the known oversensitivity of the chi-square test of exact fit (and of chi-square difference tests) to sample size, minor misspecification and omitted variables, we rely on sample-size independent fit indices to assess model fit (Hu & Bentler, 1999; Marsh et al., 2005; Yu, 2002). More precisely, the comparative fit index (CFI) and Tucker-Lewis index (TLI) should show values of at least .90 to support acceptable fit, but ideally be above .95 to support excellent fit. Likewise, values lower or equal to .08 and .06

on the root mean error of approximation (RMSEA) were respectively taken to support acceptable and excellent fit. For tests of measurement invariance, decreases in CFI and TLI ≥ .01 and increases in RMSEA ≥ .015 between one model and the previous one in the sequence were used to reject the invariance hypothesis (Chen, 2007; Cheung & Rensvold, 2002). For our final models, we also report McDonald's (1970) omega reliability coefficients.

However, fit statistics alone are not sufficient to gauge the relative adequacy of the four models compared in this study (CFA, bifactor-CFA, ESEM, bifactor-ESEM), which also requires a clear comparison of the parameter estimates of each of the alternative models (Morin, 2023; Morin et al., 2016a, 2016b, 2020). The CFA and ESEM solution are first compared. In this comparison, beyond observing that the ESEM solution fits the data better, well-defined factors (i.e. high target loadings, satisfactory estimates of composite reliability), reduced factor correlations, and the presence of cross-loadings that do not detract from the proper interpretation of the factors can all be taken as evidence supporting the ESEM solution. The optimal solution (CFA or ESEM) is then compared with its bifactor counterpart. In this second comparison, beyond model fit, observing a well-defined G-factor and at least some well-defined S-factors (i.e. high target loadings, satisfactory estimates of composite reliability), in addition to slightly reduced cross-loadings can be taken as evidence supporting the bifactor solution. It is important to keep in mind that it is frequent for a subset of S-factors to retain only a limited amount of specificity, suggesting that the items used in the assessment of these S-factors mainly serve to define the G-factor, without retaining any specificity beyond their contribution to this global construct (Arens & Morin, 2017; Morin et al., 2020). As a result, a bifactor solution also provides a direct test of the extent to which each subscale is able to capture something qualitatively distinct from the G-factor (e.g. Arens & Morin, 2017).

Results

The goodness-of-fit of the alternative solutions is reported in Table 1. These results first show that all models achieved an acceptable fit to the data, that the ESEM and bifactor-CFA solutions had a similar fit to the data, that the CFA solution had the worst fit to the data, and that the bifactor-ESEM solution had a slightly higher fit to the data than all alternative solutions.

The factor loadings and uniqueness of the four solutions in the four countries are reported in supplemental Tables S1 (CFA and ESEM factor correlations and composite reliability), S2 (Austria), S3 (Belgium), S4 (Germany), and S5 (Finland) of the online supplements. Looking first at the CFA and ESEM solutions, both resulted in similarly well-defined factors: Austria (CFA: λ = .750 to .930, ω = .933–.973; ESEM: λ = .457–.973, ω = .901–.970); Belgium (CFA: λ = .684–.963, ω = .935–.971; ESEM: λ = .464–.986, ω = .929–.967); Germany (CFA: λ = .759–.958, ω = .915–.974; ESEM: λ = .475–.997, ω = .889–.969); and Finland (CFA: λ = .647–.959, ω = .915–.975; ESEM: λ = .406–.994, ω =.905–971). Table 3 presents the omega coefficients for all the factors in the studies. The ESEM solution revealed a variety of statistically significant cross-loadings, although none were large enough to call into question the clarity of the factor definition. Moreover, factor correlations were substantially reduced in ESEM relative to CFA,

Table 1. Fit statistics for the alternative measurement models estimated in study 1.

Model	χ^2 (df)	CFI	TLI	RMSEA	RMSEA 90%CI	CM	$\Delta\chi^2$ (*df*)	ΔCFI	ΔTLI	ΔRMSEA
Austria										
CFA	1939.444* (367)	.971	.968	.064	[.061, .066]		–	–	–	–
Bifactor-CFA	1356.105* (348)	.982	.979	.052	[.049, .055]					
ESEM	1050.030* (271)	.986	.979	.052	[.049, .055]					
Bifactor-ESEM	731.492* (247)	.991	.986	.043	[.039, .047]					
Belgium										
CFA	2549.283* (367)	.974	.971	.065	[.063, .068]		–	–	–	–
Bifactor-CFA	1496.437* (348)	.986	.984	.048	[.046, .051]					
ESEM	1336.080* (271)	.987	.981	.053	[.050, .056]					
Bifactor-ESEM	806.812* (247)	.993	.989	.040	[.037, .043]					
Germany										
CFA	2301.762* (367)	.963	.959	.070	[.067, .073]		–	–	–	–
Bifactor-CFA	1223.648* (348)	.983	.980	.048	[.045, .051]					
ESEM	1242.465* (271)	.981	.972	.058	[.055, .061]					
Bifactor-ESEM	891.166* (247)	.988	.980	.049	[.046, .053]					
Finland										
CFA	1387.666* (367)	.975	.973	.042	[.040, .044]		–	–	–	–
Bifactor-CFA	801.558* (348)	.989	.987	.029	[.026, .031]					
ESEM	902.156* (271)	.985	.977	.039	[.036, .041]					
Bifactor-ESEM	735.084* (247)	.988	.981	.036	[.033, .038]					
Measurement Invariance across Countries (Bifactor-ESEM)										
M1. Configural invariance	5168.472* (988)	.982	.970	.058	[.056, .059]	–	–	–	–	–
M2. Weak (λ) invariance	4383.801* (1402)	.987	.985	.041	[.039, .042]	M1	1233.021* (414)	+.005	.015	−.017
M3. Strong (λ, τ) invariance	4895.307* (1645)	.986	.986	.039	[.038, .041]	M2	951.081* (243)	−.001	+.001	−.002
M4. Strict (λ, τ, δ) invariance	5753.233* (1732)	.983	.984	.043	[.041, .044]	M3	821.247* (87)	−.003	−.002	+.004
M5. Latent var.-covar. (λ, τ, δ, ξ/φ) invariance	3366.073* (1795)	.993	.994	.026	[.025, .028]	M4	200.548* (63)	+.010	+.010	−.017
M6. Latent mean (λ, τ, δ, ξ/φ, η) invariance	3610.207* (1813)	.992	.993	.028	[.027, .029]	M5	146.978* (18)	−.001	−.001	+.002

Note: *$p < 0.01$; CFA: confirmatory factor analysis; ESEM: exploratory structural equation modeling; χ^2: robust weighed least square (WLSMV) chi-square; *df:* degrees of freedom; CFI: Comparative fit index; TLI: Tucker-Lewis index; RMSEA: Root mean square error of approximation; 90% CI = 90% confidence interval of the RMSEA; λ: factor loadings; τ: thresholds; δ: uniquenesses; ξ: factor variances; φ: factor covariances; η: factor means; CM: comparison model; $\Delta\chi^2$: change in χ^2; ΔCFI: change in CFI; ΔTLI: change in TLI; ΔRMSEA: change in RMSEA.

supporting the value of the ESEM solution: Austria (CFA: r = .583–.780; M_r = .702; ESEM= r = .514–.654; M_r = .580); Belgium (CFA: r = .512–.765; M_r = .649; ESEM= r = .472–.672; M_r = .572); Germany (CFA: r = .549–.762; M_r = .688; ESEM: r = .463–.657; M_r = .575); and Finland (CFA: r = .499–.739; M_r = .626; ESEM= r = .354–.672; M_r = .520). The correlations observed among the BAT components were stronger than those between the BAT components and the depression factor.

The ESEM solution was therefore retained and contrasted with its bifactor-ESEM counterpart. This solution resulted in a well-defined G-factor in each country: Austria (λ = .538–.812, ω = .985); Belgium (λ = .550–.818, ω = .984); Germany (λ = .507–.837, ω = .985); and Finland (λ = .474–.841, ω = .980). It also resulted in well-defined

Table 2. Standardised factor loadings (λ) and uniquenesses (δ) from the final bifactor-ESEM solution retained in study 1 (latent mean invariance).

Items	Global Factor (λ)	Exhaustion (λ)	Mental Distance (λ)	Cognitive Impairment (λ)	Emotional Impairment (λ)	Depression (λ)	δ
EX1	**.777**	**.372**	−.066	−.054	−.050	−.054	.245
EX2	**.740**	**.368**	.086	.036	*.014*	*.012*	.308
EX3	**.684**	**.458**	−.084	*.012*	.048	.031	.313
EX4	**.644**	**.481**	*−.012*	*−.006*	.050	.038	.350
EX5	**.712**	**.339**	.164	.064	−.032	.054	.343
EX6	**.728**	**.309**	.155	.188	*.018*	.037	.313
EX7	**.624**	**.412**	*−.001*	.130	.066	*.008*	.420
EX8	**.755**	**.414**	−.084	−.030	−.062	−.037	.246
MD1	**.742**	.042	**.538**	*.002*	−.044	−.042	.154
MD2	**.644**	.066	**.392**	.023	*.007*	.031	.426
MD3	**.762**	.048	**.501**	−.041	.034	*.011*	.163
MD4	**.663**	−.070	**.634**	.046	*.003*	*.006*	.152
MD5	**.680**	*−.023*	**.404**	*−.014*	.058	.077	.365
CC1	**.745**	*.000*	.035	**.478**	−.021	−.071	.209
CC2	**.753**	.056	*.007*	**.485**	.068	*−.001*	.190
CC3	**.626**	.050	*−.002*	**.604**	.063	*.003*	.237
CC4	**.729**	−.004	*−.010*	**.573**	−.027	−.071	.135
CC5	**.602**	.022	*.014*	**.556**	.161	.046	.300
EC1	**.652**	.063	.085	.084	**.587**	.111	.200
EC2	**.702**	.069	.091	.067	**.553**	.091	.176
EC3	**.677**	−.045	−.070	*.003*	**.361**	−.082	.397
EC4	**.770**	*.025*	.041	.052	**.402**	.064	.236
EC5	**.720**	−.109	−.147	−.024	**.533**	−.089	.155
DE1	**.772**	−.072	*.019*	−.120	−.104	**.462**	.160
DE2	**.697**	−.072	*−.010*	−.109	−.071	**.619**	.109
DE3	**.570**	.030	.016	.022	.029	**.802**	.030
DE4	**.763**	−.049	−.078	−.084	−.098	**.474**	.168
DE5	**.723**	−.048	−.099	−.089	−.059	**.499**	.205
DE6	**.557**	*.013*	*.005*	.039	.059	**.787**	.066

Note. Target (main) factor loadings are in bold; statistically non-significant parameters ($p \geq 0.05$) are in italics.

S-factors for all specific dimensions of both measures, supporting the idea that all dimensions retained meaningful specificity beyond the variance explained by the G-factor: Austria (λ = .220–.783, ω = .784–.950); Belgium (λ = .260–.813, ω = .761–.954); Germany (λ = .098–.839, ω = .765–.950); and Finland (λ = .229–.845, ω = .766–.961). Lastly, and further supporting this solution, cross-loadings were reduced relative to ESEM. The bifactor-ESEM solution was thus retained for interpretations and tests of measurement invariance.

The results from the tests of measurement invariance are reported near the bottom of Table 1. These results support the full invariance of this solution, and thus generalizability, across countries, as none of the alternative models resulted in the decrease in CFI or TLI higher than .10 or an increase in RMSEA greater than .015. The parameter estimates from the most invariant solution are reported in Table 2 (factor loadings and uniquenesses) and 3 (composite reliability). These results revealed a well-defined G-factor (λ = .557–.777, ω = .984), accompanied by similarly well-defined S-factors for exhaustion (λ = −0.109–.481, ω = .797), mental distance (λ = −0.147–.634, ω = .829), cognitive impairment (λ = −0.120–.604, ω = .872), emotional impairment (λ = −0.104–.587, ω = .836), and depression (λ = −0.089–.802, ω = .947). Most items had a stronger factor loading on the G-factor than on their a priori S-factor, with the exception of the depression suicidal ideation items: DE3 (λ = .802; 'That you would be better off if you were dead') and DE6 (λ = .787; 'Did you ever think 'If only I was dead'?').

Table 3. Composite reliability (omega) for the bifactor-ESEM measurement models.

	Global factor	Exhaustion	Mental distance	Cognitive impairment	Emotional impairment	Depression
Study 1						
Austria	.985	.789	.784	.829	.846	.950
Belgium	.984	.761	.761	.833	.830	.954
Germany	.985	.765	.859	.896	.848	.950
Finland	.980	.795	.843	.854	.802	.962
Latent mean invariance	.984	.797	.829	.872	.836	.947
Study 2						
Job strain	.975	.816	.789	.867	.848	.936
Burnout	.972	.828	.733	.886	.866	.937
Depressive episode	.972	.797	.722	.848	.849	.939
Adaptation disorder	.971	.808	.762	.866	.837	.935
Latent var.-covar. invariance	.972	.817	.763	.875	.847	.928

Supplemental Table S11 shows the violin plot distributions of the suicidal ideation items for both studies, which clearly indicate that these results cannot be dismissed as statistical artefacts as all scale response options were used by participants.

Study 2

Methods

A sample of employees ($N = 5791$; 53.70% women, 17–66 years; M_{age} = 39.54; SD_{age}= 11.11) who called in sick for psychological reasons at one of the largest Dutch Occupational Health Services (ArboNed) were asked to fill an online questionnaire in March 2020 and June 2021, about 6 weeks after calling in sick. The questionnaire was administered in Dutch. They received one of the following four diagnoses from ArboNed occupational physicians: burnout ($N = 749$; 12.93%; 52.20% women; M_{age} = 41.27; SD_{age}= 10.75), job strain ($N = 985$; 17.01%; 52.59% women; M_{age} = 40.59; SD_{age}= 11.53), depressive episode ($N = 768$; 13.26%; 51.17% women; M_{age} = 37.42; SD_{age}= 11.51) and adaptation disorder ($N = 3289$; 56.80%; 54.97% women; M_{age} = 39.33; SD_{age}= 10.87). These diagnoses are based on an officially sanctioned classification used in the Dutch occupational health system (CAS-codes; Dutch Institute for Social Insurance, 2002). These participants completed the same measures used in Study 1 (exhaustion α = .919), mental distance (α = .820), cognitive impairment (α = .920), emotional impairment (α = .888), global burnout (α = .944), and depression (α = .873). This data collection was approved by the Social and Societal Ethics Committee (SMEC) of KU Leuven (reference number: G-2015 10 353).

Results

Participants' responses were analysed following procedures identical to those used in Study 1. The goodness-of-fit of the alternative solutions are reported in Table 4. The results first show that all models had an acceptable fit to the data, that the CFA

Table 4. Fit statistics for the alternative measurement models estimated in study 2.

Model	χ^2 (*df*)	CFI	TLI	RMSEA	RMSEA 90%CI	CM	$\Delta\chi^2$ (*df*)	ΔCFI	ΔTLI	ΔRMSEA
Job strain										
CFA	1732.921* (367)	.971	.968	.062	[.059, .064]	–	–	–	–	–
Bifactor-CFA	1291.890* (348)	.980	.977	.053	[.049, .056]					
ESEM	1492.621* (271)	.974	.962	.068	[.064, .071]					
Bifactor-ESEM	1082.836* (247)	.983	.971	.059	[.055, .062]					
Burnout										
CFA	1560.896* (367)	.967	.963	.066	[.063, .069]	–	–	–	–	–
Bifactor-CFA	1179.633* (348)	.977	.973	.057	[.053, .060]					
ESEM	1261.386* (271)	.972	.959	.070	[.066, .074]					
Bifactor-ESEM	917.670* (247)	.981	.969	.060	[.056, .064]					
Depressive episode										
CFA	1806.432* (367)	.969	.965	.071	[.068, .075]	–	–	–	–	–
Bifactor-CFA	1039.341* (348)	.985	.982	.051	[.047, .054]					
ESEM	1527.680* (271)	.973	.959	.078	[.074, .082]					
Bifactor-ESEM	938.440* (247)	.985	.975	.060	[.056, .064]					
Adaptation disorder										
CFA	5778.420* (367)	.962	.958	.067	[.065, .069]	–	–	–	–	–
Bifactor-CFA	3628.254* (348)	.977	.973	.054	[.052, .055]					
ESEM	4298.896* (271)	.972	.958	.067	[.066, .069]					
Bifactor-ESEM	3035.319* (247)	.981	.968	.059	[.057, .061]					
Measurement Invariance across Patient Groups (Bifactor-ESEM)										
M1. Configural invariance	5880.134* (988)	.982	.970	.059	[.057, .060]	–	–	–	–	–
M2. Weak (λ) invariance	3698.316* (1402)	.992	.990	.034	[.032, .035]	M1	609.749* (414)	+.010	+.020	−.025
M3. Strong (λ, τ) invariance	3882.571* (1645)	.992	.992	.031	[.029, .032]	M2	393.257* (243)	.000	+.002	−.003
M4. Strict (λ, τ, δ) invariance	3632.810* (1732)	.993	.993	.028	[.026, .029]	M3	165.823* (87)	+.001	+.001	−.003
M5. Latent var.-covar. (λ, τ, δ, ξ/φ) invariance	2464.729* (1795)	.998	.998	.016	[.014, .018]	M4	92.869* (63)	+.005	+.005	−.012
M6. Latent mean (λ, τ, δ, ξ/φ, η) invariance	5268.271* (1813)	.987	.989	.036	[.035, .037]	M5	1078.138* (18)	−.011	−.009	+.020

Note. $^*p < 0.01$; CFA: confirmatory factor analysis; ESEM: exploratory structural equation modeling; χ^2: robust weighed least square (WLSMV) chi-square; *df:* degrees of freedom; CFI: Comparative fit index; TLI: Tucker-Lewis index; RMSEA: Root mean square error of approximation; 90% CI = 90% confidence interval of the RMSEA; λ: factor loadings; τ: thresholds; δ: uniquenesses; ξ: factor variances; φ: factor covariances; η: factor means; CM: comparison model; $\Delta\chi^2$: change in χ^2; ΔCFI: change in CFI; ΔTLI: change in TLI; ΔRMSEA: change in RMSEA.

solution had the worst fit to the data, and that the two bifactor solutions had the highest fit across samples. The highest fit was associated with the bifactor-ESEM solution in the job strain, burnout, and adaptation disorder samples, whereas both bifactor solutions had a similar fit in the depressive episode sample.

The factor loadings and uniqueness of the four solutions in the four samples are reported in supplemental Tables S6 (CFA and ESEM factor correlations and composite reliability), S7 (job strain), S8 (burnout), S9 (depressive episode), and S10 (adaptation disorder) of the online supplements. These results are similar to those reported in Study 1. When contrasting ESEM and CFA, both solutions resulted in similarly well-defined factors: job strain (CFA: λ = .630–.936, ω = .883–.945; ESEM: λ = .509–.986, ω = .878–.943); burnout (CFA: λ = .604–.980, ω = .843–.948; ESEM: λ = .509–.997, ω = .838–.944); depressive episode (CFA: λ = .600–.969, ω = .855–.945; ESEM: λ = .398–.993, ω = .806–.942); and adaptation disorder (CFA: λ = .593–.949, ω = .861–.946; ESEM: λ = .396–.997, ω =

.858–.939). The ESEM solution revealed a variety of statistically significant cross-loadings, although none were large enough to question the clarity of the factor definition. Factor correlations were substantially reduced in ESEM relative to CFA, supporting the value of the ESEM solution: job strain (CFA: r = .347–.733; M_r = .535; ESEM: r = .292–.626; M_r = .449); burnout (CFA: r = .289–.687; M_r = .521; ESEM= r = .156–.590; M_r = .412); depressive episode (CFA: r = .250–.773; M_r = .517; ESEM= r = .275–.682; M_r = .424); and adaptation disorder (CFA: r = .355–.733; M_r = .530; ESEM= r = .289–.629; M_r = .453). As in Study 1, correlations among BAT components were stronger than those between BAT components and the depression factor.

The ESEM solution was thus retained and contrasted with its bifactor-ESEM counterpart. This solution resulted in a well-defined G-factor in each patient sample: Job strain (λ = .234–.796, ω = .975); burnout (λ = .239–.766, ω = .972); depressive episode (λ = .151–.820, ω = .972); and adaptation disorder (λ = .195–.766, ω = .971). It also resulted in well-defined S-factors for all specific dimensions of both measures, supporting the idea that all dimensions retained meaningful specificity beyond the variance explained by the G-factor: job strain (λ = .143–.914, ω = .789–.936); burnout (λ = .223–.934, ω = .733–.937); depressive episode (λ = .185–.937, ω = .722–.939); and adaptation disorder (λ = .216–.936, ω = .762–.935). Further supporting the value of this solution, cross-loadings were also reduced relative to ESEM. Therefore, the bifactor-ESEM solution was retained for interpretations and tests of measurement invariance.

The results from the tests of measurement invariance are reported in the bottom of Table 4. These results confirmed the invariance of the factor loadings, response thresholds, item uniquenesses, and factor variances and covariances across all samples, supporting the generalizability of this factor structure and the lack of measurement biases across patient samples. The results also revealed latent mean differences across samples (i.e. ΔCFI = −0.011; ΔRMSEA = +.020), which we present in the next paragraph. The parameter estimates from the final retained model of latent variance-covariance invariance are reported in Tables 5 (factor loadings and uniquenesses) and 3 (composite reliability). These results revealed a well-defined G-factor (λ = .256–.762, ω = .972), accompanied by similarly well-defined S-factors for exhaustion (λ = .225–.535, ω = .817), mental distance (λ = .324–.664, ω = .763), cognitive impairment (λ = .509–.588, ω = .875), emotional impairment (λ = .490–.691, ω = .847), and depression (λ = .415–.944, ω = .928). Most items had a stronger factor loading on the G-factor than on their a priori S-factor, with the exception of mental distance items MD4 (λ = .664; 'I feel indifferent about my job') and MD5 (λ = .567; 'I'm cynical about what my work means to others'), emotional impairment item EC5 (λ = .691; 'At work I may overreact unintentionally'), and depression suicidal ideation items DE2 (λ = .792; 'That life is not worthwhile?'), DE3 (λ = .944; 'That you would be better off if you were dead') and DE6 (λ = .923; 'Did you ever think 'If only I was dead'?'). Considering the distribution of the suicidal ideation items for this study (see supplemental Table S11), one can clearly see that all scale response options were used but that the depressive episode patient group had a less skewed distribution compared to the country samples. That is, the depressive episode group had more differentiation of agreement with suicidal ideation items compared to the burnout group across both studies.

Table 5. Standardised factor loadings (λ) and uniquenesses (δ) from the final bifactor-ESEM solution retained in study 2 (latent variance-covariance invariance).

Items	Global factor (λ)	Exhaustion (λ)	Mental distance (λ)	Cognitive impairment (λ)	Emotional impairment (λ)	Depression (λ)	δ
EX1	**.723**	**.467**	.088	.025	*.008*	*.009*	.252
EX2	**.750**	**.401**	.054	.119	−.060	*−.015*	.255
EX3	**.653**	**.523**	*−.016*	−.027	.051	.029	.296
EX4	**.615**	**.485**	*−.006*	−.019	.042	.048	.382
EX5	**.713**	**.327**	.104	*.007*	−.076	.024	.367
EX6	**.709**	**.225**	*.011*	.191	−.063	−.028	.406
EX7	**.656**	**.416**	−.046	.092	*.009*	*.001*	.386
EX8	**.704**	**.535**	.017	*−.001*	.031	.033	.215
MD1	**.689**	.063	**.477**	*.001*	−.029	−.053	.290
MD2	**.441**	.058	**.324**	.043	.062	.061	.688
MD3	**.627**	.114	**.536**	−.085	.018	−.047	.297
MD4	**.495**	*−.017*	**.664**	.050	.031	.065	.307
MD5	**.442**	*.004*	**.567**	*.005*	.090	.110	.463
CC1	**.748**	−.047	*.004*	**.514**	−.076	−.070	.163
CC2	**.762**	.044	−.040	**.509**	*.006*	−.019	.156
CC3	**.628**	.095	.020	**.588**	.123	.058	.231
CC4	**.736**	.019	*−.003*	**.579**	*.002*	−.044	.122
CC5	**.558**	.029	.047	**.522**	.173	.060	.380
EC1	**.616**	−.039	−.048	*.002*	**.582**	*.015*	.277
EC2	**.633**	−.026	−.047	*−.008*	**.595**	−.039	.241
EC3	**.546**	.098	.106	.055	**.490**	*.021*	.438
EC4	**.593**	.037	.098	.085	**.552**	.078	.320
EC5	**.519**	*.009*	.043	.053	**.691**	.036	.248
DE1	**.563**	−.147	−.053	−.166	−.027	**.547**	.331
DE2	**.417**	−.076	−.024	−.088	*−.015*	**.792**	.184
DE3	**.256**	.040	.027	.034	*.010*	**.944**	.040
DE4	**.624**	−.217	−.204	−.248	−.117	**.460**	.235
DE5	**.554**	−.191	−.171	−.175	−.066	**.415**	.421
DE6	**.260**	.037	.022	.020	.030	**.923**	.078

Note. Target (main) factor loadings are in bold; statistically non-significant parameters ($p \geq .05$) are in italics.

The results related to the latent mean differences observed across samples in this final model are reported in Table 6. These results indicate that global levels of psychological distress (i.e. scores on the G-factor) were the highest in the burnout and depressive episode samples, which did not differ from one another, followed by the adaptation disorder sample, and were the lowest among the job strain sample. Specific levels of exhaustion were the highest in the burnout sample, followed by the adaption disorder sample, then by the job strain sample, and finally were the lowest in the depressive episode sample. Specific levels of mental distance were the highest in the job strain and depressive episode sample, which did not differ from one another, and were the lowest in the burnout and adaptation disorder sample, which did not differ from one another. Specific levels of cognitive impairment and emotional impairment were the highest in the burnout and adaptation disorder samples, which did not differ from one another, and were the lowest in the depressive episode and job strain sample, which did not differ from one another. Lastly, specific levels of depression were the highest in the depressive episode sample, followed by the burnout sample, and were the lowest in the job strain and adaptation disorder sample, which did not differ from one another.

General discussion

Across two studies, this research sought to contribute to our understanding of the similarities and differences between burnout and depression by investigating the

Table 6. Latent means (and standard errors in parentheses) from the final bifactor-ESEM solution retained in study 2 (latent variance-covariance invariance).

	Job strain	Burnout	Depressive episode	Adaptation disorder
Global Factor	.000	.380 (.052)**	.488 (.053)**	.144 (.039)**
Exhaustion	.000	.321 (.062)**	−.162 (.064)**	.085 (.045)
Mental Distance	.000	−.176 (.059)**	−.106 (.062)	−.147 (.044)**
Cognitive Impairment	.000	.126 (.060)*	.099 (.060)	.184 (.044)**
Emotional Impairment	.000	.124 (.057)*	−.091 (.061)	.096 (.042)*
Depression	.000	.131 (.069)	1.278 (.062)**	.014 (.052)
Global Factor	−.380 (.052)**	.000	.109 (.056)	−.235 (.043)**
Exhaustion	−.321 (.062)**	.000	−.483 (.068)**	−.236 (.053)**
Mental Distance	.176 (.059)**	.000	.070 (.064)	.029 (.048)
Cognitive Impairment	−.126 (.060)*	.000	−.028 (.064)	.058 (.050)
Emotional Impairment	−.124 (.057)*	.000	−.215 (.063)**	−.028 (.047)
Depression	−.131 (.069)	.000	1.147 (.067)**	−.117 (.058)*
Global Factor	−.488 (.053)**	−.109 (.056)	.000	−.343 (.045)**
Exhaustion	.162 (.064)*	.483 (.068)**	.000	.246 (.054)**
Mental Distance	.106 (.062)	−.070 (.064)	.000	−.042 (.052)
Cognitive Impairment	−.099 (.060)	.028 (.064)	.000	.084 (.050)
Emotional Impairment	.091 (.061)	.214 (.063)**	.000	.186 (.051)**
Depression	−1.278 (.062)*	−1.147 (.067)**	.000	−1.265 (.050)**
Global Factor	−.144 (.039)**	.235 (.043)**	.343 (.045)**	.000
Exhaustion	−.085 (.045)	.236 (.053)**	−.246 (.054)**	.000
Mental Distance	.147 (.044)**	−.029 (.048)	.042 (.052)	.000
Cognitive Impairment	−.184 (.044)**	−.058 (.050)	−.084 (.050)	.000
Emotional Impairment	−.096 (.042)*	.028 (.047)	−.186 (.051)**	.000
Depression	−.014 (.052)	.117 (.058)*	1.265 (.050)**	.000

Note. * $p < 0.05$; ** $p < 0.01$; Latent means are fixed to zero in one reference group for identification purposes, while the freely estimated means in the other samples directly expressed as differences from the referent group in standardized units. Statistically significant differences indicate that the mean in the target group is statistically different than those from the referent group (in which the means are fixed to 0).

construct-relevant multidimensionality present in ratings obtained on the BAT and on the depression subscale of the 4DSQ across four countries (study 1) and four distinct samples of patients (study 2). More precisely, by relying on the bifactor-ESEM framework, we could accurately disentangle the variance shared across both measures from that unique to each specific dimension of these instruments.

Both studies yielded almost identical results supporting the superiority of the bifactor-ESEM solution, as well as confirming our expectation (anchored in previous research; e.g. Doherty et al., 2021; Sandrin et al., 2022; Tóth-Király et al., 2021) that this solution would be the more suitable for these constructs. More precisely, these analyses revealed the presence of an underlying global psychological distress factor encompassing the variance shared among all indicators of burnout and depression. They also revealed that the four specific factors from the BAT (i.e. exhaustion, mental distance, cognitive impairment, and emotional impairment), as well as the specific factor capturing depression, all retained a meaningful level of specificity (reasonably large factor loadings and a satisfactory estimate of composite reliability) beyond the global factor. The fact that all these specific factors retained some meaningful level of specificity is incompatible with previous affirmations that the distinctive nature of both constructs is a simple artefact of the wording of the burnout items (contrary to depression items) as referring to work (see Maslach et al., 2001). These observations are also in line with recent research arguing that, beyond sharing a common core of psychological distress, burnout and depression represent conceptually distinct entities (e.g. Koutsimani et al., 2019; Meier & Kim, 2022; Schaufeli et al., 2020, Tóth-Király et al., 2021). More specifically, the identification of a strong global factor explains why previous studies (e.g. Chiu et al., 2015; Schonfeld & Bianchi, 2016; Thuynsma & De Beer, 2017) have revealed strong associations between burnout and depression. Yet, the presence of this common core, reflecting generic feelings of psychological distress, does not mean that both constructs are identical, as indicated by the presence of similarly strong specific factors reflecting the unique nature of both constructs. In plain language, our results show clear evidence that burnout and depression are distinct entities both characterized by the presence of strong psychological distress, and that banishing burnout to focus solely on depression is likely to be counterproductive.

Further supporting this interpretation, in the patient samples, we found that the burnout and depressive episode samples displayed a similar level of global psychological distress. In contrast, and supporting the discriminant validity of our specific factors, the burnout group displayed the highest levels of exhaustion, cognitive impairment, and emotional impairment, whereas the depressive episode group displayed the lowest levels on all these specific factors. Likewise, levels of depression were also significantly higher in the depression group than in the burnout episode group. The fact that burnout patients also displayed higher levels of depressive symptoms than the job strain and adaptation disorder samples is consistent with previous studies indicating that burnout sometimes tends to be accompanied by depressive symptoms (e.g. Bianchi et al., 2013; Chiu et al., 2015). Less expected was the observation that specific levels of mental distance were higher in the job strain and depressive episode sample than in the burnout sample, suggesting that mental distance could be a more important and reliable indicator of one's global levels of psychological distress than of burnout. Future studies will need to better understand this result.

Interestingly, in both studies (but even more strongly in the patient samples), the two depression items specific to suicidal ideation were found to load far more strongly on the specific depression factor than on the underlying global psychological distress factor. This suggests that suicidal ideation may represent a core difference between the burnout syndrome and depression, in line with similar conclusions reached by other studies (e.g. Deeb et al., 2018; Ernst et al., 2021), as well as with clinical recommendations to consider suicidal ideation in the differential diagnosis of depression and burnout (Hoogduin et al., 2001). Furthermore, the distribution patterns of suicidal ideation items in supplemental Table S11 reveal distinct trends between the burnout and depressive episode patient groups. Specifically, the burnout group exhibited more consistent patterns of suicidal ideation across each study, like that of the country plots, whereas the depressive episode patient group demonstrated distinct variability in agreement on suicidal ideation items compared to other groups. This divergence suggests that the unique characteristics of our data may provide insight into why some studies involving non-representative and/or subclinical samples yield varying degrees of overlap between burnout and depression. It appears that the composition of study samples, particularly in terms of employees who are (or are not) struggling, may also significantly influence these findings.

Moreover, in contrast to components of the MBI, which have been shown to correlate more strongly with depression than with one another (e.g. Schonfeld & Bianchi, 2016), the ESEM and CFA correlations obtained in both our studies support the value of the

BAT relative to the MBI, showing that BAT components correlate more strongly with one another than with the depression subscale. This pattern of association was particularly clear in the patient samples where the BAT components were only moderately correlated with depression, particularly in ESEM, which is known to capture better associations among constructs (Asparouhov et al., 2015; Mai et al., 2018). However, beyond our reliance on ESEM, it is also possible that relying on samples characterized by a broader range of burnout and depression experiences (i.e. patients relative to community samples) is necessary to fully capture this more nuanced pattern of associations.

Practical implications

This study clearly demonstrated that all components of BAT-assessed burnout, as well as the depression subscale of the 4DSQ, remained meaningful indicators of their respective constructs beyond their ability to capture a strong global psychological distress factor. These conclusions support the value of both measures. However, they also highlight the importance of researchers, clinicians, and occupational health practitioners carefully considering their purpose when selecting specific measurement instruments. In a notable development, Schaufeli et al. (2023) proposed a pooled international cut-off score for the identification of burnout utilizing the BAT in countries in which the psychometric validity of scores obtained on this instrument has been demonstrated, but emphasize that these should be regarded as preliminary pending future research in more diversified contexts. Importantly, although we did not specifically assess this structure as our goal was to specifically assess the distinctive nature of burnout and depression, it is important to keep in mind that the BAT itself has been shown to present a dual global (i.e. global levels of burnout) and specific (i.e. levels on the four subscales) (Schaufeli et al., 2020). This means that professionals using the BAT in practical contexts should be able to obtain both a global burnout score, while having access to scores on each specific dimension to obtain a richer assessment.

From the perspective of social, organizational and occupational psychology, burnout has been shown to be an important consideration for organizations. Over the years, an extensive body of research has established the value of modelling the effects of work-related conditions on burnout within theoretical frameworks such as the job demands-resources model (Bakker et al., 2023; Lesener et al., 2019). From a more clinical perspective, the novel nature of burnout as a potential diagnostic means that evidence is still lacking regarding how to handle it as a unique condition. Importantly, clear diagnostic criteria and representative norms are still missing to properly guide any potential clinical assessment of burnout (Brisson & Bianchi, 2017; Schaufeli, 2021). Clinicians facing manifestations of clinical distress in countries where burnout is not officially acknowledged as a condition may benefit from measures of job-related depression, such as the occupational depression inventory (ODI; Bianchi & Schonfeld, 2020) to identify cases categorically for referral to further clinical screening if required.

However, given that the level of psychological distress observed in our patient samples did not differ between the burnout and depressive episode groups, suggest that a third alternative is also viable; that of relying on a combined measure of burnout and depression such as we used in the present study. This combined measure makes it possible to consider both phenomena, as well as their common core, in a more comprehensive manner. Moreover, observing similar levels of psychological distress between these two subsamples suggest that burnout partly overlaps with the depressive spectrum. This would suggest that employees suffering from burnout could benefit from a reference to mental health professionals to be screened, and treated when appropriate, for the presence of depression or another mental health category. Indeed, at present, whereas specific protocols have been developed, tested, and validated support the clinical treatment of depression, similar interventions are still lacking in relation to burnout. Interestingly, our result suggests that those intervention protocols should account for the occasional presence of both conditions.

While interventions for burnout are typically focused on stress management and enhancing work-life balance, addressing depression, especially when accompanied by suicidal ideation, may require a more intensive approach. However, this distinction does not imply that individuals experiencing burnout complaints are free from the risk of experiencing suicidal thoughts. Consequently, when evaluating burnout risk referrals, it is important to screen for signs of suicidal ideation, ideally by a qualified professional, before determining the appropriate intervention (Hoogduin et al., 2001). In their review of research on mental health at work, Kelloway et al. (2023) differentiated between activities that can be conducted within the workplace and those, like treatment, which are usually conducted outside of it. They emphasized that occupational health professionals need to be aware of the limitations of what is appropriate to be addressed in the workplace.

Therefore, it is also crucial to consider the role of organizational infrastructure and the legal limitations that work psychology and human resource professionals face in this context. Not all work psychology professionals are legally permitted (licensed) to diagnose or assess for clinical conditions like depression (see Kelloway et al., 2023). Conversely, the assessment of burnout, recognized by the WHO as an occupational phenomenon rather than a medical condition, offers a more utilitarian avenue for organizations to screen a greater number of struggling employees.

Limitations and directions for future research

This study is not without limitations. First, the results are solely based on self-report measures which can suffer from biases such as social desirability, memory recall and limited self-awareness by respondents. While self-reported measures are predominant in occupational health psychology research (Spector, 2019), it is important to recognize their unique strength in capturing internal states such as cognitions and emotions; these aspects are inherently subjective and are most accurately assessed through self-report measures—offering valuable insights that might not be as effectively gauged through other methods (Spector, 2019). Nevertheless, the four groups of patients considered in Study 2 received an official diagnosis provided by independent occupational physicians. Furthermore, although these diagnoses were based on the CAS code system used in the Netherlands, this system may not be immediately applicable in other contexts. To alleviate these concerns, however, it is important to acknowledge that this CAS code system is largely based on the International Classification of Diseases (ICD) and that our results were largely in line with what would be expected based on these classifications. It might be interesting, for future studies, to consider asking occupational physicians to directly participate in rating the severity of patients' manifestations of burnout and depression as the main source of ratings. Beyond these considerations, our sole reliance on self-report measures means that part of the variance captured by the G-factor might reflect common method bias (CMB; Podsakoff et al., 2003), in addition to psychological distress. The only way to separate those two sources of 'global' variance would have been to incorporate covariates to the model, and to document the meaning of the G-factor through tests of criterion-related validity. In the present context, this limitation arguably remains minimal for two reasons. Firstly, our goal was not so much to document the scope and meaning of the G-factor, but rather to assess the extent to which each subdimension from both measures retained a meaningful level of specificity once everything that they had in common was taken into account. Importantly, the S-factors themselves are completely free from CMB, which gets completely absorbed by the G-factor. Secondly, investigations of common method bias (e.g. Podsakoff et al., 2003, 2012) have found that it rarely explains more than 25% of the variance as an upper bound—which is not negligible, but less than the variance explained by our G-factor. This explained variance can be calculated by squaring the factor loadings and reaches an average of 43% for our G-factor and 29% for our S-factors, leaving 26% of variance unique to the items (across studies and samples)—the rest being explained by the cross-loadings. This means that, even if we were to extract 25% of CMB from our G-factor, this would still leave 18% of variance attributable to psychological distress, providing even stronger support to our conclusions regarding the distinctiveness of both constructs. However, it would still be important for future studies to account for this methodological artefact when conducting tests of criterion-related validity seeking to establish the complete nomological network of the G- and S-factors identified in this study.

Second, predictive validity could not be assessed as no outcome measure was available in these samples. Factors important to the individual and organisation should thus be considered in future studies, such as turnover intention, actual turnover, performance, and organizational commitment. This would seem to be an important avenue to consider for future research seeking to expand upon the present results. Third, we relied on a variable-centered approach (i.e. relations among variables) to understand the overlapping and distinct nature of burnout and depression. The flip side of the coin, person-centered approaches (Morin et al., 2018), would rather consider this question by looking at subpopulations of employees displaying qualitatively distinct sets of psychological distress symptoms, and are likely to help us better understand when, and how, both conditions co-occur and the key drivers of this co-occurrence. Fourth, we relied on cross-sectional analyses, which are unable to clearly inform how each condition relates to the other over time, and the directionality of these associations. Although previous longitudinal studies have similarly documented the distinctive nature of burnout and depression using limited measures of both constructs (e.g. Tóth-Király et al., 2021), it would be highly interesting to expand upon these previous studies by considering more complete measures of burnout and depression (such as those used in the present study), while also considering the state and trait component of these associations (e.g. Hofmans et al., 2021); for instance *via* the application of random intercepts cross-lagged panel models (Hamaker et al., 2015) or latent curve models with structured residuals (Curran et al., 2014).

Fifth, although the fact that our results suggested that suicidal ideation could represent a core indicator of what is unique to depression relative to burnout, no additional information allowing us to further explore this unexpected observation was available in our datasets. It would thus seem important, for future studies, to move beyond the simple consideration of whether there is value in distinguishing between burnout and depression—indeed, we believe that this has been clearly established in ours and previous systematic, meta-analytic, studies (e.g. Koutsimani et al., 2019; Meier & Kim, 2022; Tóth-Király et al., 2021)—to more specifically consider how these two forms of psychological distress differ from one another.

Sixth, it would seem important to expand upon the current results through the consideration of the biological, neuropsychological, and cognitive underpinning of burnout and depression, as these underpinnings might also play a role in the differentiation between these two conditions as well as in the development of effective differential approaches to treatment. For example, a study on electrophysical (EEG) markers showed that significant differences exist in distinguishing burnout participants when conflicted/incongruent stimuli or erroneous reactions are being processed (see Golonka et al., 2018). Yet in a study on diurnal cortisol profiles, no significant differences were found between depression, burnout, and psychological distress but for that all three were related to associated increases in cortisol (Marchand et al., 2014).

Seventh, the depression subscale we used did not contain specific fatigue-related items. However, research indicates that the omission of these items from alternate depression scales minimally affects if at all, the associations between depression and burnout (e.g. Bianchi et al., 2021). Furthermore, we speculate that even if fatigue items were included in the measure, these would most likely have clustered with the general factor, with specific variance split between the exhaustion and depression. That is, we surmise that there would be no substantial impact on the overall findings of this study. In any case, this limitation highlights the need to systematically assess whether and how the present conclusions will generalize to other measures of burnout and depression.

Lastly, despite our reliance on multiple, large, representative, and clinical samples, our study remained limited to so called WEIRD samples (Western, Educated, Industrialized, Rich, Democratic; Henrich et al., 2010), highlighting the need for replication among diversified populations. Providing preliminary support to the generalizability of our results, a recent study conducted in Brazil and using similar methods found that even though BAT-assessed burnout shares some characteristics with depression, it can be clearly differentiated (de Amorim Macedo et al., 2023).

Conclusion

Our results contribute to the ongoing discussion about the differential nature of burnout and depression by demonstrating the value of considering these two states as meaningfully distinct, while sharing a common core of psychological distress. As is often the case in psychological research, the response thus does not seem to lie on an either (e.g. distinct) or (e.g. overlapping) continuum, but rather to represent a combination of both possibilities. This means that both states, despite their common core, also capture unique aspects. Beyond this theoretical discussion, decades of research have established that burnout does serve an important social and practical purpose, allowing for the identification of severe manifestations of psychological distress in the workplace that cannot be swept under the rug of conditions emerging primarily in the personal life of the employees. This extensive research evidence has resulted in the official recognition of burnout as a diagnosable, and insurable, condition in some European countries, and our results further support the idea that this recognition is anchored in a meaningfully distinct set of manifestations. Yet, given their overlap and the lack of efficient treatment strategies, it is reasonable to refer burned out employees for a clinical screening, and possible treatment, of a depressive condition. This recommendation is not anchored in the suggestion that both states are the same, simply in the need to offer optimal treatment to distressed employees, while we await the development of even better clinical strategies.

Note

1. Because item-level true score (i.e. reliable) variance is divided between two sets of factors (G and S) in a bifactor solution, it is typical for the S-factors to be more weakly defined than their first-order CFA or ESEM counterpart (Morin et al., 2020), leading to suggestions that composite reliability coefficients as low as .50 should still be considered acceptable for S-factors (e.g. Perreira et al., 2018).

Acknowledgements

The authors would like to acknowledge Robin Kok, PhD: HumanTotalCare, Research & Business Development, Utrecht, Netherlands, for his assistance with the study in the Netherlands.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The third and fourth authors were supported by funding from KU Leuven (C3-project C32/15/003). The authors affiliated to the Finnish Institute of Occupational Health acknowledge support from the Finnish Work Environment Fund. The last author was supported by a grant from the Social Science and Humanity Research Council of Canada (435-2018-0368) in the preparation of this paper.

Data availability statement

The data associated with this manuscript is available from the corresponding author upon reasonable request. Note the supplementary material that are available at: https://doi.org/10.1080/08870446.2024.2321358.

References

Ahola, K., Hakanen, J., Perhoniemi, R., & Mutanen, P. (2014). Relationship between burnout and depressive symptoms: A study using the person-centred approach. *Burnout Research*, *1*(1), 29–37. https://doi.org/10.1016/j.burn.2014.03.003

Arens, A. K., & Morin, A. J. S. (2017). Improved representation of the Self-Perception Profile for Children through bifactor exploratory structural equation modelling. *American Educational Research Journal*, *54*(1), 59–87. https://doi.org/10.3102/0002831216666490

Armon, G., Shirom, A., & Melamed, S. (2012). The big five personality factors as predictors of changes across time in burnout and its facets. *Journal of Personality*, *80*(2), 403–427. https://doi.org/10.1111/j.1467-6494.2011.00731.x

Asparouhov, T., & Muthén, B. (2010). Weighted least square estimation with missing data. https://www.statmodel.com/download/GstrucMissingRevision.pdf

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modelling with crossloadings and residual covariances. *Journal of Management*, *41*(6), 1561–1577. https://doi.org/10.1177/0149206315591075

Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. (2023). Job demands–resources theory: Ten years later. *Annual Review of Organizational Psychology and Organizational Behavior*, *10*(1), 25–53. https://doi.org/10.1146/annurev-orgpsych-120920-053933

Barcza-Renner, K., Eklund, R. C., Morin, A. J. S., & Habeeb, C. M. (2016). Controlling coaching behaviors and athlete burnout: Investigating the mediating roles of perfectionism and motivation. *Journal of Sport & Exercise Psychology*, *38*(1), 30–44. https://doi.org/10.1123/jsep.2015-0059

Bianchi, R., Boffy, C., Hingray, C., Truchot, D., & Laurent, E. (2013). Comparative symptomatology of burnout and depression. *Journal of Health Psychology*, *18*(6), 782–787. https://doi.org/10.1177/1359105313481079

Bianchi, R., & Schonfeld, I. S. (2020). The Occupational Depression Inventory: A new tool for clinicians and epidemiologists. *Journal of Psychosomatic Research*, *138*, 110249. https://doi.org/10.1016/j.jpsychores.2020.110249

Bianchi, R., Schonfeld, I. S., & Laurent, E. (2015a). Burnout–depression overlap: A review. *Clinical Psychology Review*, *36*, 28–41. https://doi.org/10.1016/j.cpr.2015.01.004

Bianchi, R., Schonfeld, I. S., & Laurent, E. (2015b). Burnout: Absence of binding diagnostic criteria hampers prevalence estimates. *International Journal of Nursing Studies*, *52*(3), 789–790. https://doi.org/10.1016/j.ijnurstu.2014.12.008

Bianchi, R., Schonfeld, I. S., & Laurent, E. (2017a). Burnout or depression: Both individual and social issue. *The Lancet*, *390*(10091), 230. https://doi.org/10.1016/S0140-6736(17)31606-9

Bianchi, R., Schonfeld, I. S., & Laurent, E. (2017b). Burnout syndrome'–from nosological indeterminacy to epidemiological nonsense. *BJPsych Bulletin*, *41*(6), 367–368. https://doi.org/10.1192/pb.41.6.367

Bianchi, R., Verkuilen, J., Schonfeld, I. S., Hakanen, J. J., Jansson-Fröjmark, M., Manzano-García, G., Laurent, E., & 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

Bresó, E., Salanova, M., & Schaufeli, W. B. (2007). In search of the "third dimension" of burnout: Efficacy or inefficacy? *Applied Psychology*, *56*(3), 460–478. https://doi.org/10.1111/j.1464-0597.2007.00290.x

Brisson, R., & Bianchi, R. (2017). Stranger things: On the upside down world of burnout research. *Academic Psychiatry: The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, *41*(2), 200–201. https://doi.org/10.1007/s40596-016-0619-7

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

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chiu, L. Y., Stewart, K., Woo, C., Yatham, L. N., & Lam, R. W. (2015). The relationship between burnout and depressive symptoms in patients with depressive disorders. *Journal of Affective Disorders*, *172*, 361–366. https://doi.org/10.1016/j.jad.2014.10.029

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology*, *82*(5), 879–894. https://doi.org/10.1037/a0035297

de Amorim Macedo, M. J. , de Freitas, C. P. P., Bermudez, M. B., Vazquez, A. C. S., Salum, G. A., & Dreher, C. B. (2023). The shared and dissociable aspects of burnout, depression, anxiety, and irritability in health professionals during COVID-19 pandemic: A latent and network analysis. *Journal of Psychiatric Research*, *166*, 40–48. https://doi.org/10.1016/j.jpsychires.2023.09.005

De Beer, L. T., & Bianchi, R. (2019). Confirmatory Factor Analysis of the Maslach Burnout Inventory. *European Journal of Psychological Assessment*, *35*(2), 217–224. https://doi.org/10.1027/1015-5759/a000392

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

Deeb, G. R., Braun, S., Carrico, C., Kinser, P., Laskin, D., & Golob Deeb, J. (2018). Burnout, depression and suicidal ideation in dental and dental hygiene students. *European Journal of Dental Education: Official Journal of the Association for Dental Education in Europe*, *22*(1), e70–e74. https://doi.org/10.1111/eje.12259

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

Doherty, A., Mallett, J., Leiter, M., & McFadden, P. (2021). Measuring burnout in social work: Factorial validity of the Maslach Burnout Inventory-Human Services Survey. *European Journal of Psychological Assessment*, *37*(1), 6–14. https://doi.org/10.1027/1015-5759/a000568

Dutch Institute for Social Insurance. (2002). CAS: Classificaties voor Arbo en SV. Classificatie van klachten, ziekten en oorzaken voor bedrijfs- en verzekeringsartsen [CAS: Classification of symptoms, diseases and causes for occupational and insurance physicians]. https://www.steungroep.nl/images/her_keuring_WIA_of_WAO/Wetten_en_regels_bij_her_keuring/CAS_Classificaties_voor_Arbo_en_SV_UWV_2002.pdf

Epstein, R. M., & Privitera, M. R. (2017). Physician burnout is better conceptualised as depression–authors' reply. *The Lancet*, *389*(10077), 1398. https://doi.org/10.1016/S0140-6736(17)30898-X

Ernst, J., Jordan, K. D., Weilenmann, S., Sazpinar, O., Gehrke, S., Paolercio, F., Petry, H., Pfaltz, M. C., Méan, M., Aebischer, O., Gachoud, D., Morina, N., von Känel, R., & Spiller, T. R. (2021). Burnout, depression and anxiety among Swiss medical students–A network analysis. *Journal of Psychiatric Research*, *143*, 196–201. https://doi.org/10.1016/j.jpsychires.2021.09.017

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 439–492). Information Age Publishing.

Freudenberger, H. J. (1974). Staff burn-out. *Journal of Social Issues*, *30*(1), 159–165. https://doi.org/10.1111/j.1540-4560.1974.tb00706.x

Freund, P. A., Tietjens, M., & Strauss, B. (2013). Using rating scales for the assessment of physical self-concept: Why the number of response categories matters. *Measurement in Physical Education and Exercise Science*, *17*(4), 249–263. https://doi.org/10.1080/1091367X.2013.807265

Galanti, T., Guidetti, G., Mazzei, E., Zappalà, S., & Toscano, F. (2021). Work from home during the COVID-19 outbreak: The impact on employees' remote work productivity, engagement, and stress. *Journal of Occupational and Environmental Medicine*, *63*(7), e426–e432. https://doi.org/10.1097/JOM.0000000000002236

Glass, D. C., & McKnight, J. D. (1996). Perceived control, depressive symptomatology, and professional burnout: A review of the evidence. *Psychology & Health*, *11*(1), 23–48. https://doi.org/10.1080/08870449608401975

Golonka, K., Mojsa-Kaja, J., Marek, T., & Gawlowska, M. (2018). Stimulus, response and feedback processing in burnout–An EEG study. *International Journal of Psychophysiology: Official Journal *of the International Organization of Psychophysiology*, *134*, 86–94. https://doi.org/10.1016/j.ijpsycho.2018.10.009

Guseva Canu, I., Marca, S. C., Dell'Oro, F., Balázs, Á., Bergamaschi, E., Besse, C., Bianchi, R., Bislimovska, J., Koscec Bjelajac, A., Bugge, M., Busneag, C. I., Çağlayan, Ç., Cernițanu, M., Costa Pereira, C., Dernovšček Hafner, N., Droz, N., Eglite, M., Godderis, L., Gündel, H., … Wahlen, A. (2021). Harmonized definition of occupational burnout: A systematic review, semantic analysis, and Delphi consensus in 29 countries. *Scandinavian Journal of Work, Environment & Health*, *47*(2), 95–107. https://doi.org/10.5271/sjweh.3935

Hakanen, J. J., & Schaufeli, W. B. (2012). Do burnout and work engagement predict depressive symptoms and life satisfaction? A three-wave seven-year prospective study. *Journal of Affective Disorders*, *141*(2–3), 415–424. https://doi.org/10.1016/j.jad.2012.02.043

Hadžibajramović, E., Schaufeli, W., & De Witte, H. (2020). A Rasch analysis of the Burnout Assessment Tool (BAT). *PloS One*, *15*(11), e0242241. https://doi.org/10.1371/journal.pone.0242241

Hadžibajramović, E., Schaufeli, W., & De Witte, H. (2022). Shortening of the Burnout Assessment Tool (BAT) – from 23 to 12 items using content and Rasch analysis. *BMC Public Health*, *22*(1), 560. https://doi.org/10.1186/s12889-022-12946-y

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Hawrot, A., & Koniewski, M. (2018). Factor structure of the Maslach Burnout Inventory–Educators Survey in a Polish-speaking sample. *Journal of Career Assessment*, *26*(3), 515–530. https://doi.org/10.1177/1069072717714545

Heinemann, L. V., & Heinemann, T. (2017). Burnout research: Emergence and scientific investigation of a contested diagnosis. *SAGE Open*, *7*Advance online publication. (1), 215824401769715. https://doi.org/10.1177/2158244017697154

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29. https://doi.org/10.1038/466029a

Hofmans, J., Morin, A. J. S., Breitsohl, H., Ceulemans, E., Chénard-Poirier, L. A., Driver, C. C., Fernet, C., Gagné, M., Gillet, N., González-Romá, V., Grimm, K. J., Hamaker, E. L., Hau, K.-T., Houle, S. A., Howard, J. L., Kline, R. B., Kuijpers, E., Leyens, T., Litalien, D., … Wille, B. (2021). The baby and the bathwater: On the need for substantive-methodological synergy in organizational research. *Industrial and Organizational Psychology*, *14*(4), 497–504. https://doi.org/10.1017/iop.2021.111

Hoogduin, C., Schaufeli, W. B., Schaap, C., & Bakker, A. (2001). *Behandelingsstrategieën bij burn-out* [The treatment of burnout]. Bohn Stafleu Van Loghum.

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

Isoard-Gautheur, S., Martinent, G., Guillet-Descas, E., Trouilloud, D., Cece, V., & Mette, A. (2018). Development and evaluation of the psychometric properties of a new measure of athlete burnout: The Athlete Burnout Scale. *International Journal of Stress Management*, *25*(S1), 108–123. https://doi.org/10.1037/str0000083

Kelloway, E. K., Dimoff, J. K., & Gilbert, S. (2023). Mental health in the workplace. *Annual Review of Organizational Psychology and Organizational Behavior*, *10*(1), 363–387. https://doi.org/10.1146/annurev-orgpsych-120920-050527

Kleinstäuber, M., Exner, A., Lambert, M. J., & Terluin, B. (2021). Validation of the Four-Dimensional Symptom Questionnaire (4DSQ) in a mental health setting. *Psychology, Health & Medicine*, *26*(sup1), 1–19. https://doi.org/10.1080/13548506.2021.1883685

Koutsimani, P., Montgomery, A., & Georganta, K. (2019). The relationship between burnout, depression, and anxiety: A systematic review and meta-analysis. *Frontiers in Psychology*, *10*, 284. https://doi.org/10.3389/fpsyg.2019.00284

Lastovkova, A., Carder, M., Rasmussen, H. M., Sjoberg, L., Groene, G. J. D., Sauni, R., Vevoda, J., Vevodova, S., Lasfargues, G., Svartengren, M., Varga, M., Colosio, C., & Pelclova, D. (2018). Burnout syndrome as an occupational disease in the European Union: An exploratory study. *Industrial Health*, *56*(2), 160–165. https://doi.org/10.2486/indhealth.2017-0132

Lesener, T., Gusy, B., & Wolter, C. (2019). The job demands-resources model: A meta-analytic review of longitudinal studies. *Work & Stress*, *33*(1), 76–103. https://doi.org/10.1080/02678373.2018.1529065
Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing exploratory structural equation modeling and existing approaches for multiple regression with latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 737–749. https://doi.org/10.1080/10705511.2018.1444993
Marchand, A., Durand, P., Juster, R. P., & Lupien, S. J. (2014). Workers' psychological distress, depression, and burnout symptoms: Associations with diurnal cortisol profiles. *Scandinavian Journal of Work, Environment & Health*, *40*(3), 305–314. https://www.jstor.org/stable/43188021 https://doi.org/10.5271/sjweh.3417
Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics* (pp. 275–340). Erlbaum.
Maslach, C. (1976). Burned-out. *Human Behavior*, *5*, 16–22.
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. (2016). Understanding the burnout experience: recent research and its implications for psychiatry. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *15*(2), 103–111. https://doi.org/10.1002/wps.20311 27265691
Maslach, C., & Leiter, M. P. (2021, March 19). How to measure burnout accurately and ethically. *Harvard Business Review*. https://hbr.org/2021/03/how-to-measure-burnout-accurately-and-ethically
Maslach, C., Leiter, M. P., & Jackson, S. E. (2017). *Maslach burnout inventory manual* (4th ed.). Mind Garden.
Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, *52*(1), 397–422. https://doi.org/10.1146/annurev.psych.52.1.397
McDonald, R. P. (1970). Theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*(1), 1–21. https://doi.org/10.1111/j.2044-8317.1970.tb00432.x
Meier, S. T., & Kim, S. (2022). Meta-regression analyses of relationships between burnout and depression with sampling and measurement methodological moderators. *Journal of Occupational Health Psychology*, *27*(2), 195–206. https://doi.org/10.1037/ocp0000273
Mészáros, V., Adám, 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: Journal of the International Society for the Investigation of Stress*, *30*(1), 82–88. https://doi.org/10.1002/smi.2481
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. (2016a). 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. https://doi.org/10.1080/10705511.2014.961800
Morin, A. J. S., Arens, A. K., Tran, A., & Caci, H. (2016b). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the composite scale of morningness. *International Journal of Methods in Psychiatric Research*, *25*(4), 277–288. https://doi.org/10.1002/mpr.1485
Morin, A. J. S., Bujacz, A., & Gagné, M. (2018). Person-centered methodologies in the organizational sciences. *Organizational Research Methods*, *21*(4), 803–813. https://doi.org/10.1177/1094428118773856
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. (2022). *Mplus user's guide, version 8.8.* Muthén & Muthén.

Nadon, L., De Beer, L. T., & Morin, A. J. S. (2022). Should burnout be conceptualized as a mental disorder? *Behavioral Sciences (Basel, Switzerland)*, *12*(3), 82. https://doi.org/10.3390/bs12030082

Perreira, T. A., Morin, A. J. S., Hebert, M., Gillet, N., Houle, S. A., & Berta, W. (2018). The short form of the Workplace Affective Commitment Multidimensional Questionnaire (WACMQ-S): A bifactor-ESEM approach among healthcare professionals. *Journal of Vocational Behavior*, *106*, 62–83. https://doi.org/10.1016/j.jvb.2017.12.004

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science reserch and recommendations on how to control it. *Annual Review of Psychology*, *63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Sandrin, E., Morin, A. J. S., Fernet, C., & Gillet, N. (2022). Complementary variable- and person-centered approaches to the dimensionality of burnout among fire station workers. *Anxiety, Stress, and Coping*, *35*(4), 440–457. https://doi.org/10.1080/10615806.2021.1959917

Schaufeli, W. B. (2021). The burnout enigma solved? *Scandinavian Journal of Work, Environment & Health*, *47*(3), 169–170. https://doi.org/10.5271/sjweh.3950

Schaufeli, W., & De Witte, H. (2023). Burnout Assessment Tool (BAT). In C. U. Krägeloh, M. Alyami, & O. N. Medvedev (Eds.), *International handbook of behavioral health assessment*. Springer. https://doi.org/10.1007/978-3-030-89738-3_54-1

Schaufeli, W. B., De Witte, H., Hakanen, J. J., Kaltiainen, J., & Kok, R. (2023). How to assess severe burnout?: Cutoff points for the Burnout Assessment Tool (BAT) based on three European samples. *Scandinavian Journal of Work, Environment & Health*, *49*(4), 293–302. https://doi.org/10.5271/sjweh.4093

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), 9495. https://doi.org/10.3390/ijerph17249495

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

Schonfeld, I. S., & Bianchi, R. (2016). Burnout and depression: Two entities or one? *Journal of Clinical Psychology*, *72*(1), 22–37. https://doi.org/10.1002/jclp.22229

Spector, P. E. (2019). Do not cross me: Optimizing the use of cross-sectional designs. *Journal of Business and Psychology*, *34*(2), 125–137. https://doi.org/10.1007/s10869-018-09613-8

Szigeti, R., Balázs, N., Bikfalvi, R., & Urbán, R. (2017). Burnout and depressive symptoms in teachers: Factor structure and construct validity of the Maslach Burnout Inventory-Educators Survey among elementary and secondary school teachers in Hungary. *Stress and Health: Journal of the International Society for the Investigation of Stress*, *33*(5), 530–539. https://doi.org/10.1002/smi.2737

Terluin, B. (1994). *Overspanning onderbouwd. Een onderzoek naar de diagnose surmenage in de huisartspraktijk*. Universiteit Utrecht.

Terluin, B., Rhenen, W. V., Schaufeli, W. B., & De Haan, M. (2004). The Four-Dimensional Symptom Questionnaire (4DSQ): Measuring distress and other mental health problems in a working population. *Work & Stress*, *18*(3), 187–207. https://doi.org/10.1080/0267837042000297535

Thuynsma, C., & de Beer, L. T. (2017). Burnout, depressive symptoms, job demands and satisfaction with life: Discriminant validity and explained variance. *South African Journal of Psychology*, *47*(1), 46–59. https://doi.org/10.1177/0081246316638564

Tóth-Király, I., Morin, A. J. S., & Salmela-Aro, K. (2021). Reciprocal associations between burnout and depression: An eight-year longitudinal study. *Applied Psychology*, *70*(4), 1691–1727. https://doi.org/10.1111/apps.12295

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

Weel, A. (2021). Beroepsziekten in cijfers 2020 [Occupational diseases in numbers 2020]. *Tijdschrift Voor Bedrijfs- en Verzekeringsgeneeskunde*, *29*(3), 43–43. https://doi.org/10.1007/s12498-021-1343-0

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

World Health Organization [WHO]. (2019b). Depressive disorders. In *International statistical classification of diseases and related health problems* (11th ed.). WHO. https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/1563440232

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. https://doi.org/10.1177/0013164408315268

Yu, C. Y. (2002). *Evaluating cut-off criteria of model fit indices for latent variable models with binary and continuous outcomes*. University of California.