Systematic Review

Maslach Burnout Inventory – General Survey

A Systematic Review and Meta-Analysis of Measurement Properties

Leon T. De Beer[1,2], Leoni van der Vaart[1,3], Maximiliano Escaffi-Schwarz[4], Hans De Witte[5,3], and Wilmar B. Schaufeli[5,6]

[1]Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway
[2]WorkWell Research Unit, North-West University, Potchefstroom, South Africa
[3]Optentia Research Unit, North-West University, Vanderbiljpark, South Africa
[4]Facultad de Administración y Economía, Universidad Diego Portales, Santiago, Chile
[5]Research Unit Occupational & Organizational Psychology and Professional Learning, Leuven, KU Leuven, Belgium
[6]Department of Psychology, Utrecht University, Utrecht, The Netherlands

Abstract: This study sought to investigate the measurement properties of a "gold standard" instrument for measuring burnout, the Maslach Burnout Inventory – General Survey (MBI-GS), with a systematic literature review and meta-analysis of studies that had as the primary aim its psychometric validation. The search spanned from January 1996 to December 2022 using the following databases: Web of Science, Scopus, PubMed, APA PsychINFO, ScienceDirect, and ProQuest Dissertations and Theses Global. Our search identified 35 eligible studies for inclusion in the systematic review. Of these, 17 were included in the meta-analysis. The meta-analysis of the original 16-item version studies supported a three-dimensional representation of burnout with modest internal consistencies. However, similar to the narrative findings, it also indicated the viability of a two-factor solution. Consequently, the structural validity of the MBI-GS remains unclear, and so does its cross-cultural validity. The criterion validity of the cynicism and personal efficacy scales also raised a few questions. Overall, the findings provided some support for the adequacy of the measurement properties of the MBI-GS as a research tool measuring exhaustion, cynicism, and professional efficacy. However, we also identified challenges and practices of which researchers should be cognisant and that they should consider in future burnout projects.

Keywords: MBI-GS, burnout scale, employee health, employee well-being, exhaustion

Since its first appearance in a publication at the end of the 1960s (Bradley, 1969), "burnout" has undeniably become ingrained in the modern work zeitgeist. While the term is now more commonly attributed to Freudenberger (1974), nearly 50 years have passed (see Schaufeli, 2023). Since then, burnout has not only received "pop psychology" status but has also become a research priority for several types of psychologists and clinicians. Its popularity is easily demonstrated by a casual topic or title search on the Web of Science, which produces more than 41,250 hits. Google Scholar delivers 1,580,000 results. Burnout was also recently recognized by the World Health Organization (2019) as an "occupational phenomenon" (para. 1) in the ICD-11 and defined as "… a syndrome conceptualized as resulting from chronic workplace stress that has not been successfully managed" (para. 4).

The work of Maslach and colleagues has popularised burnout over the last few decades (Bianchi et al., 2022; Schaufeli, 2017). Maslach and Jackson (1981) published the first iteration of the research tool, the Maslach Burnout Inventory (MBI; 22 items), exclusively for use with human services workers. At that time, it was thought that burnout occurred due to interactions with people (or patients). This first version of the MBI consisted of three factors: emotional exhaustion (EE; "… feelings of being emotionally overextended and exhausted by one's work"; p. 101), depersonalization (DP; "… unfeeling and impersonal response towards recipients of one's care or service"; p. 101), and personal accomplishment (PA; "… feelings of

competence and successful achievement in one's work with people"; p. 101).

However, as research on burnout progressed, it became increasingly evident that the phenomenon was not limited to human services work, but also manifested itself in other occupations and sectors (Demerouti et al., 2001; Halbesleben & Demerouti, 2005; Pines & Aronson, 1988). Today, burnout is considered by most to be a common phenomenon in any occupation or work context. Practically, this created a problem, as the original MBI was designed for human services work, and the items were tailored for that context. Consequently, adapted versions of the MBI were created with items more suitable for specific areas of investigation. The original MBI became known as the MBI-Human Services Survey (MBI-HSS). In addition, a parallel version for teachers and educators, the MBI-Educators Survey (MBI-ES), was created (Maslach & Jackson, 1986). The MBI-General Survey (MBI-GS) was developed for use in the more general work context, suitable for any occupation or work context (Schaufeli et al., 1996). Due to these advancements, and despite the creation of other competing burnout measures, the MBI became the "gold standard" of burnout measurement.

The MBI-GS comprises 16 items and three dimensions: exhaustion (more generic than the original), cynicism (reflecting indifference and/or a distant attitude toward work in general), and professional efficacy (a broader scope than the original PA, expectations of continued effectiveness at work) (Schaufeli et al., 1996; Schutte et al., 2000). Five items are related to exhaustion, five are associated with cynicism, and six are linked to professional efficacy. The "exhaustion" items were derived from some of the original MBI-HSS items measuring "emotional exhaustion." The five "cynicism" items were novel, replacing "depersonalization" in the general survey. The new "professional efficacy" dimension had one item adapted from the MBI-HSS element of "personal accomplishment," with the remaining five being newly devised (Schaufeli et al., 1996).

Previous Systematic Reviews and Meta-Analyses

Although no specific systematic reviews or meta-analyses on the measurement properties of the MBI-GS were identified, we found a relevant review focusing on the measurement properties of the original MBI. In their systematic literature review and meta-analysis, Worley and colleagues (2008) investigated the factor structure of the MBI across 45 single factor-analytic studies. Mixed results were evident. Their results supported the three-factor model (EE, DP, PA) of burnout, even more so when *orthogonal* rotations were used. There was also support for the *correlated* three-factor model in the confirmatory factor analyses, despite discrepancies in model construction. However, an alternative two-factor solution (EE + DP, PA) was viable as well, with the combined EE + DP factor interpreted as the core of burnout. Worley and colleagues (2008) did not include the MBI-GS in their investigation. Our literature search indicated that no one had taken up this task yet, leaving a substantive gap regarding the psychometric properties of the MBI-GS – which has arguably become the flag-bearer for burnout research over the past 25+ years.

The Current Study

A recent call to action underlined the importance of new systematic reviews and meta-analytic studies in the psychological assessment sphere in which burnout measures were cited as an example (see Iliescu et al., 2022). In heeding this call, we aimed to review the measurement properties (i.e., validity and internal consistency) of the MBI-GS and the evidence and study quality behind the MBI-GS. More specifically, we aimed to evaluate the construct-relevant validity (i.e., structural validity), measurement invariance across conditions (i.e., cross-cultural validity), internal consistency or reliability of the most optimal factor-analytic solution, and criterion validity using relevant work-related outcomes (e.g., mental health disorders/problems).

A review of the measurement properties of the MBI-GS is essential when one considers the conflicting results from single studies using this instrument. These discrepancies center around the questions above regarding its optimal factor solution. Is burnout (as measured by the MBI-GS) best represented by the proposed three-factor solution (Bakker et al., 2002), or is a two-factor solution (Pando Moreno et al., 2015) more suitable? Is there sufficient evidence for its internal consistency, specifically for the newly developed cynicism scale (Bria et al., 2014)? Should some cynicism items be removed (Chirkowska-Smolak & Kleka, 2011)? Should we change the valence of professional efficacy (Schaufeli & Salanova, 2007; Taris et al., 1999), and is professional efficacy a valid component of burnout (Qiao & Schaufeli, 2011)? Do all three factors play a role in mental health disorders/problems, or are they differentially related to mental health outcomes (Roelofs et al., 2005; Xanthopoulou et al., 2012)? By meta-analyzing the measurement properties, we should also be able to answer these questions and provide insights into whether it is warranted to adapt the MBI-GS by changing the item content (e.g., Langballe et al., 2006).

Evidence for the cross-cultural validity of psychological instruments is also becoming increasingly important to ensure that concepts, models, and scores can be fairly used without bias (Meuleman et al., 2023). Consequently, this raises the question of whether the MBI-GS is equivalent across countries, especially when comparing WEIRD

(Western, educated, industrial, rich, democratic) with non-WEIRD countries (Beyebach et al., 2021; Henrich, 2020; Henrich et al., 2010). In achieving its aims, this study sought to provide a snapshot of the status quo and encourage future action on the MBI-GS.

Method

In conducting the review, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (Page et al., 2021). The review was pre-registered on the international prospective register of systematic reviews (PROSPERO) to avoid duplication and to minimize reporting bias. This registration is available at: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=393159. All supplementary materials are available online at the Open Science Framework: https://doi.org/10.17605/OSF.IO/N9V5T.

Data Sources and Search

Based on the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) methodology for systematic reviews of patient-reported outcome measures (PROMs) guidelines (Mokkink, de Vet, et al., 2018; Prinsen et al., 2016), the search and selection processes were divided into four distinct phases and actioned by the research team. Firstly, we formulated the aim of the review using the PICO framework: P – Population (i.e., employees/organizations/[non-student] workers), I – Instruments (i.e., MBI-GS), C – Construct (i.e., burnout), O – Outcomes (i.e., measurement properties/reliability [internal consistency]/validity [structural validity, hypothesis testing, cross-cultural validity, and criterion validity]). The PICO framework is typically used in reviews of the psychometric properties of measuring instruments (Aromataris & Munn, 2020).

Secondly, eligibility criteria (in line with the PICO framework) were formulated. Studies were included in the review if they (a) focused on developing and validating the MBI-GS among employees. Studies were excluded if they included student samples (as "student employees") and if they used the MBI-GS as an outcome measure. Studies were also excluded if they focused on the discriminant validity of burnout (e.g., differentiation between burnout and depression or burnout and work engagement), as the primary focus of these studies was not the development and/or validation of the MBI-GS. Studies were included if they (b) reported original empirical work using a quantitative approach. Qualitative and theoretical studies were excluded. The review included studies that (c) were published between 1996 and December 2022, as the MBI-GS was introduced in 1996. No restrictions were placed on language. It included studies that (d) were published in academic journals, scientific books, conference proceedings, and theses/dissertations. Popular press articles were excluded. Studies were included in the review if they (e) used factor-analytic methods.

Thirdly, based on the search terms, the first and second authors independently conducted a comprehensive systematic literature search between January and March 2023. This search was performed using the following databases: Web of Science, Scopus, PubMed, APA PsychINFO, ScienceDirect, and ProQuest Dissertations and Theses Global. The search query was as follows:

> "Maslach Burnout Inventory – General Survey" OR MBI-GS OR MBI* AND work* OR employee* OR organi?ation* AND burn*out AND valid* OR psychometric OR "cross-cultur*" OR properties OR measurement.

Basic manuscript details (i.e., database, item type, publication year, author(s), source title, publication title, volume, issue, pages, DOI, and abstract) were captured in Microsoft Excel®. Each author's dataset was screened and cleaned, and duplicates were removed independently using the Systematic Review Accelerator (SRA) app (Clark et al., 2020). The two datasets were exported to Microsoft Excel® spreadsheets. The combined dataset contained 3,741 records. The two authors' lists were compared to ensure replicability and duplicates (n = 2,450) were removed using the SRA app (Clark et al., 2020). Afterward, a total of 1,291 records remained.

In the fourth place, the titles of the manuscripts were screened and assessed against the eligibility criteria. Eligibility was screened by each of the researchers independently. Once the independent screening had been completed, the first and second authors compared the results. Cohen's kappa (κ) value was calculated with the irr package in R to quantify the interrater agreement (Gamer et al., 2019). For this step, the results showed an "almost perfect agreement" (κ = 0.939). Where differences occurred, the reasoning was discussed, and another author's opinion was sought as a tiebreaker. The first and second authors reviewed the full papers (n = 67) for inclusion/exclusion. The full texts of those papers identified via abstract screening were then read and assessed against the eligibility criteria. A similar process to the previous step was employed, where two researchers screened each text (Cohen's κ = 0.901; "almost perfect agreement"). A backward search (i.e., checking the references of the 22 studies deemed eligible) was done to see whether any additional papers needed to be included. An additional 34 items were

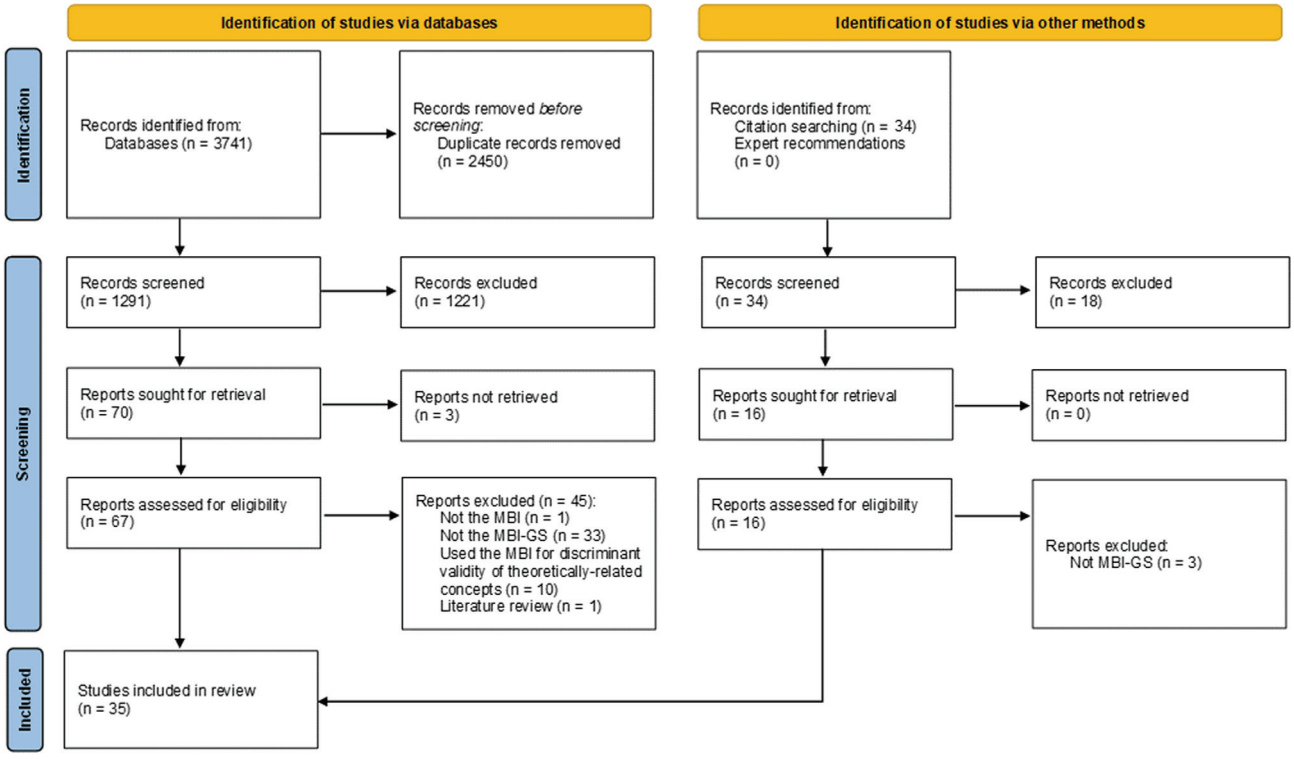

Figure 1. Flow chart outlining the search, screening, and selection process. Adapted from the PRISMA preferred reporting for systematic reviews (Page et al., 2021).

identified, of which 13 were eventually included in this study. The authors discussed each manuscript to collate a final list of manuscripts.

The final list of included manuscripts (n = 35) was then collated and circulated to two prominent academics within the field of occupational health psychology to determine whether any important records had somehow been missed or had to be included in their opinion. They had to have at least 10 years of academic experience related to burnout, to have served on the editorial boards of prominent occupational health psychology journals, to have published at least 50 papers/chapters on occupational health psychology, and to have a minimum Google Scholar h-index of 25. This step yielded no additional publications when compared strictly to our eligibility criteria. We also contacted the Occupational Health Psychology fraternity via a well-known mailing list (OHPLIST@lists.apa.org) to request unpublished manuscripts and/or datasets aligned with our aims. Again, no additional documents were recommended or received from this request. Figure 1 provides an overview of the complete search, screening, and selection process.

Quality Assessment and Data Extraction

The COSMIN risk of bias checklist was used to evaluate the quality of the included studies (Mokkink, de Vet, et al., 2018). Each study was rated as very good, adequate, doubtful, or inadequate (see the online supplementary material). This step was performed independently by the first and second authors. A third author resolved differences. No studies were excluded, as the results of studies deemed "inadequate quality" were still in line with those from good-quality studies (see the recommendation by Mokkink, Prisen, et al., 2018). A customized data extraction sheet containing the following information was developed for this review (see Table S1 in the supplementary material): bibliometric study characteristics (i.e., sample description, sample size, age, gender, country/language of administration, properties assessed, and analytical methods, and results). A codebook was also developed for the meta-analysis, where we coded the following information: sample characteristics (i.e., sample size, age, work/organizational/position tenure, language, country, and industry), means, internal consistency coefficients, exploratory factor analysis (EFA)/confirmatory factor analysis (CFA) factor loadings, and correlations. The first and second authors coded the data, and the third author-verified it before performing the analyses.

Data Synthesis and Analysis

Next, the measurement properties of every single study were rated against the criteria for good measurement

properties. Each result (see Table S1 in the supplementary material, De Beer, 2023) was rated as sufficient (+), insufficient (−), or indeterminate (?) (Prinsen et al., 2016). The results were qualitatively summarized (using all eligible studies) and quantitatively pooled (using only those studies that included all 16 items in their analyses). The quantitative summary focused on four key criteria: factor structure, correlation between factors (i.e., structural validity), internal consistency, and criterion validity. All quantitative analyses were conducted in R v4.2.2 (R Core Team, 2021) with the following packages: metafor (Viechtbauer, 2010) for univariate meta-analytical procedures, psych (Revelle, 2023) and GPArotation (Bernaards & Jennrich, 2005) for the meta-analytical EFA, and metaSEM (Cheung, 2015b) for the two-stage structural equation modeling (TSSEM) technique.

We used the procedure described by Shafer (2005) and Worley and colleagues (2008) to assess the factor structure. This procedure involves counting the number of times two items exhibited strong factor loadings on the same factor across all samples, creating a co-occurrence matrix per study. Then, one sums all the co-occurrence matrices into a pooled matrix to use as input for an EFA. Items were considered to load together when two items had a factor loading of at least .40 in the same factor, and a sample-weighted co-occurrence matrix was used to assign more weight to studies with larger sample sizes (Shafer, 2005). After creating the sample-weighted co-occurrence matrix, parallel analysis was used to decide how many factors to extract, followed by principal component analysis (PCA) with oblimin rotation. Co-occurrence matrices often have high multicollinearity, so PCA is preferred over factor analysis procedures that require inverting the matrix, resulting in computational problems when collinearity is high (Shafer, 2005). Results from a PCA using this method are interpreted in the same way as running the same analysis using primary data.

We used univariate meta-analytic techniques with a random effects model to assess the correlation between factors. We focused on the reliability-corrected correlation and followed Siegel and colleagues' (2022) recommendations to use different techniques to assess publication bias. Specifically, we used the trim and fill procedure using the R0 estimator that provides a significance test, Egger's regression using the standard error as a predictor, and a fail-safe number (FSN).

To assess the reliability, we followed the recommendations by Rodriguez and Maeda (2006) and Greco and colleagues (2018), who suggest using a raw alpha weighted by variance precision with a random effects model, as this produces the most unbiased estimate. Here, we had sufficient data to explore how reliability performed in different samples, such as validations in Spanish (k = 4) versus Dutch (k = 5) versus other languages (k = 11) and WEIRD (k = 14) versus non-WEIRD (k = 7) samples.

Finally, we employed a multivariate approach to investigate the relationship between burnout factors and criterion variables (i.e., mental health disorders/problems). This approach is generally preferred over univariate methods, as it allows us to control for the covariance between factors when predicting their impact on criterion variables. Specifically, we utilized the TSSEM technique Cheung (2015a) developed, which uses a sample-weighted pooled correlation matrix. This matrix is then used to run an SEM model, and in this study, we used the reliability-corrected correlations to create the pooled correlation matrix. The TSSEM approach offers several advantages compared to other multivariate approaches such as general linear modeling (GLM). A very salient strength is the use of sample-weighted paths to estimate standard errors and significance tests, as opposed to the harmonic mean typically used in GLM. Consequently, the TSSEM approach provides more accurate estimates of effect sizes, standard errors, and significance tests.

It is important to mention that, to conduct TSSEM analyses, we needed to transform the data. While most studies measured exhaustion, cynicism, and professional efficacy, only a few measured the same criterion variable. Therefore, we recoded all variables related to mental health disorders/problems (e.g., anxiety, depression, stress, somatization) into a single label to increase the sample size. In cases where a single study provided multiple correlations for the same variable, we aggregated them by transforming them into Fisher's Z values and then converting them back to correlation values. This step was necessary because correlations do not follow a linear distribution and should not be summed or averaged. As a drawback, however, creating a unified label reduced the specificity of the effect.

Lastly, the ratings of each measurement property in the meta-analysis were accompanied by a grading of the quality of evidence (see the online supplementary material, De Beer, 2023). Using the modified GRADE approach, one can assess the trustworthiness of the pooled results (Prinsen et al., 2016).

Results

When reporting the results, we first provide a qualitative summary (i.e., the systematic review component), after which we provide the pooled results from the statistical analyses (i.e., the meta-analytic component). The only exception is cross-cultural validity, as we could not perform a meta-analysis for this measurement property. As many selected studies either excluded item(s) from the start or

early in their analyses, the qualitative summaries provide a more nuanced overview of any challenges experienced with the psychometric properties of the instrument. However, because the paper aimed to evaluate the psychometric properties of the MBI-GS, it is more appropriate to cast an overall judgment based on studies that included all 16 items – as this is how the standard instrument is provided to users.

Study Characteristics

Our qualitative summary of the studies is based on 43 independent samples. Seven of the 35 studies used more than one sample: two (Leiter & Schaufeli, 1996; Roelofs et al., 2005; Shirom & Melamed, 2006; Taris et al., 1999; Vanheule et al., 2012; Xanthopoulou et al., 2012) or three (García et al., 2020). Three studies (Chen et al., 2014; Mäkikangas et al., 2011; Richardsen & Martinussen, 2005) were longitudinal; therefore, their samples were paired. The countries were represented by several independent samples: Argentina (n = 1), Belgium (n = 2), Bolivia (n = 1), Ecuador (n = 1), Canada (n = 2), China (n = 2), Colombia (n = 2), Costa Rica (n = 1), the Dominican Republic (n = 1), Finland (n = 2), France (n = 1), Greece (n = 2), Japan (n = 1), Malaysia (n = 1), Mexico (n = 1), Norway (n = 2), Peru (n = 2), Poland (n = 1), Romania (n = 1), South Africa (n = 2), Spain (n = 3), the Netherlands (n = 6), Venezuela (n = 2), and Vietnam (n = 1). A noteworthy observation is the lack of academic publications of validation studies from the United States. The MBI-GS was translated into Chinese, Cuban, Dutch, English, Finnish, French, Greek, Japanese, Malay, Norwegian, Polish, Portuguese, Romanian, Spanish, and Vietnamese to enable global use. Despite its many translations, the Dutch (n = 7) and Spanish (n = 9) translations were most frequently used in these studies. The sample sizes ranged from 73 to 9,055. The participants' mean ages ranged from 26 to 51, whereas their mean work experience and organizational and position tenure ranged 3.14–16, 1.69–11, and 2.45–8.6 years, respectively. In almost half of the studies (i.e., 49%), there was a higher percentage of females than males.

Our quantitative pooling of results is based on 22 independent samples. Here we included Bedoya and Agudelo (2016), Chirkowska-Smolak and Kleka (2011), and Nguyen and colleagues (2018), as their exploratory factor analyses were conducted with all 16 items. A total of 16,464 participants were included in the analyses, of whom only 6,474 (40%) were females. Participants represented the following countries: Canada (n = 3,729), China (n = 748), Colombia (n = 1,093), Finland (n = 155), France (n = 1,312), Greece (n = 232), Japan (n = 285), Malaysia (n = 123), Norway (n = 694), Poland (n = 1,148), Spain (n = 262), the Netherlands (n = 3,550), Venezuela (n = 233), Vietnam (n = 430), and a mixed sample from South America (n = 2,470). Their mean age was 36.30 (SD = 5.64), and their mean work experience and organizational and position tenure were 8.63 (SD = 4.73), 7.40 (SD = 5.00), and 6.88 (SD = 2.68), respectively.

Structural Validity

Our systematic review indicated that the varimax rotation was used in many studies (e.g., Chirkowska-Smolak & Kleka, 2011; Chen et al., 2014; Mojsa et al., 2006) that conducted an EFA. This was somewhat surprising, given that the three components of the MBI were expected to correlate because burnout was conceived as a syndrome (i.e., a set of symptoms that co-occurred and referred to the same underlying condition). However, direct oblimin (Bocéréan et al., 2019; Le Roux, 2005) and promax (Bedoya & Agudelo, 2016; Nguyen et al., 2018) rotation methods were also used. In the end, most studies opted for or expected a three-factor solution (e.g., Millán de Lange & D'Aubeterre López, 2012; Moreno-Jiménez et al., 2001; Schuster et al., 2015; Tómas et al., 2016). There were some exceptions where a two- (Pando Moreno et al., 2015; Spontón et al., 2019) or even a four-factor (Gil-Monte, 2002) solution fitted the data better or authors only measured EX and CY (Qiao & Schaufeli, 2011), citing earlier issues with PE.

The three-factor solution was often marred by challenges. In 51% of the included studies, the three-factor model only fitted after model modifications (i.e., deleting items and/or correlating error variances). Of these, no fewer than six studies removed Item 13 ("… do my job and not be bothered") from the cynicism scale to improve model fit (e.g., Bria et al., 2014; Storm & Rothmann, 2003; Xanthopoulou et al., 2012), whereas three removed it before administration (Kleijweg et al., 2013; Le Roux, 2005; Vanheule et al., 2012).

Regarding the meta-analytical EFA results, Table 1 presents the raw and sample-weighted co-occurrence matrix. The parallel analysis (Figure 2) suggested that three factors had to be extracted. However, we considered that the cut-off point was close to two factors being the ideal model. Hence, we also estimated a two-factor model using PCA with oblimin rotation. Results for this model can be observed in the supplementary material (De Beer, 2023, Table S5). The three-factor model (Table 2) showed that all items loaded more strongly on their factor. Similar results were found with oblimin rotation using the proportion co-occurrence matrix (Table S2) and varimax rotation using the sample-weighted (Table S3) and proportion (Table S4) co-occurrence matrices. Furthermore, the three factors explained 90% of the variance in the dataset. However, there were two important cross-loadings: item MG6 ("… burned out … work"), an exhaustion item that loaded

Table 1. Co-occurrence matrix

		MG1	MG2	MG3	MG4	MG6	MG5	MG7	MG10	MG11	MG12	MG16	MG8	MG9	MG13	MG14	MG15
		exh1	exh2	exh3	exh4	exh5	effic1	effic2	effic3	effic4	effic5	effic6	cyn1	cyn2	cyn3	cyn4	cyn5
MG1	exh1	**10**	10	10	9	9	0	0	0	1	0	0	3	4	4	2	3
MG2	exh2	1.00	**10**	10	9	9	0	0	0	1	0	0	3	4	3	2	3
MG3	exh3	1.00	1.00	**10**	9	9	0	0	0	1	0	0	4	5	5	3	4
MG4	exh4	.98	.98	.98	**10**	8	0	0	0	0	0	0	5	6	4	4	3
MG6	exh5	.95	.95	.95	.93	**10**	0	0	0	1	0	0	7	8	8	6	7
MG5	effic1	.00	.00	.00	.00	.00	**10**	10	10	9	10	10	0	0	0	0	0
MG7	effic2	.00	.00	.00	.00	.00	1.00	**10**	10	9	10	10	0	0	0	0	0
MG10	effic3	.00	.00	.00	.00	.00	1.00	1.00	**10**	9	10	10	0	0	0	0	0
MG11	effic4	.02	.02	.02	.00	.02	.98	.98	.98	**10**	9	9	0	0	1	0	1
MG12	effic5	.00	.00	.00	.00	.00	1.00	1.00	1.00	.98	**10**	10	0	0	0	0	0
MG16	effic6	.00	.00	.00	.00	.00	1.00	1.00	1.00	.98	1.00	**10**	0	0	0	0	0
MG8	cyn1	.30	.30	.33	.35	.56	.00	00	.00	.00	.00	.00	**10**	9	7	7	6
MG9	cyn2	.70	.70	.73	.75	.96	.00	.00	.00	.00	.00	.00	.60	**10**	8	8	7
MG13	cyn3	.70	.49	.73	.71	.96	.00	.00	.00	.02	.00	.00	.56	.96	**10**	8	9
MG14	cyn4	.61	.61	.64	.66	.87	.00	.00	.00	.00	.00	.00	.51	.91	.91	**10**	9
MG15	cyn5	.63	.63	.66	.64	.85	.00	.00	.00	.02	.00	.00	.44	.84	.88	.93	**10**

Note. MG=Maslach Burnout Inventory-General Survey; cyn=cynicism; exh=exhaustion; effic=professional efficacy. Values above the diagonal are the raw values, while values below the diagonal are sample-weighted rounded to the second decimal. Values in bold represent the diagonal.

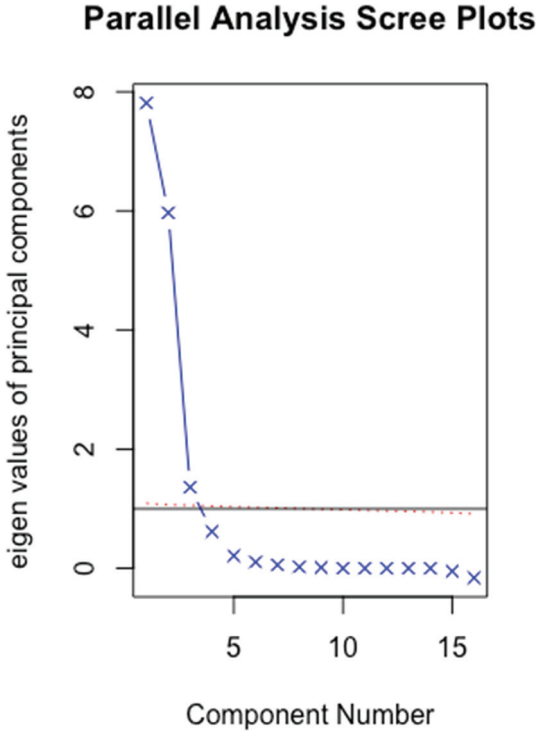

Figure 2. Parallel analysis for the sample-weighted co-occurrence matrix.

Table 2. Principal component analysis with oblimin rotation using the sample-weighted co-occurrence matrix

Item	Exhaustion	Cynicism	Professional efficacy
MG1 (exh1)	0.99		
MG2 (exh2)	1.04		
MG3 (exh3)	0.96		
MG4 (exh4)	0.93		
MG6 (exh5)	0.59	0.54	
MG8 (cyn1)	−0.31	0.90	
MG9 (cyn2)	0.18	0.84	
MG13 (cyn3)		0.92	
MG14 (cyn4)		0.91	
MG15 (cyn5)	0.13	0.84	
MG5 (effic1)			1.00
MG7 (effic2)			1.00
MG10 (effic3)			1.00
MG11 (effic4)			0.99
MG12 (effic5)			1.00
MG16 (effic6)			1.00
Proportion variance explained	0.27	0.26	0.37

Note. MG = Maslach Burnout Inventory-General Survey; cyn = cynicism; exh = exhaustion; effic = professional efficacy.

strongly on cynicism, and MG8 ("… less interested … since I started …"), a cynicism item that had a moderate and negative factor loading on exhaustion. Moreover, the two-factor model suggested that the two factors were professional efficacy and exhaustion/cynicism (i.e., all items of exhaustion and cynicism loaded onto the same factor).

Although the associations between the three dimensions were most often significant and in the expected direction, there were some notable exceptions. For example,

Table 3. Meta-analytic correlations between the three factors of burnout

Variable 1	Variable 2	Sample	*K*	*N*	*r*	r_c	CI_L	CI_U	I^2	*Q* (*df*)
Exhaustion	Cynicism	All	17	8,157	.63	.75**	.66	.84	99.9	1349.89 (16)**
Exhaustion	Professional efficacy	All	16	7,924	−.27	−.33**	−.41	−.24	93.27	196.71 (15)**
Cynicism	Professional efficacy	All	17	11,237	−.44	−.55**	−.66	−.44	99.52	4012.48 (16)**
Exhaustion	Cynicism	WEIRD	10	5,715	.59	.70**	.56	.84	99.93	1064.61 (9)**
Exhaustion	Professional efficacy	WEIRD	10	5,715	−.25	−.31**	−.39	−.22	89.62	94.71 (9)**
Cynicism	Professional efficacy	WEIRD	10	8,795	−.42	−.54**	−.63	−.45	97.25	324.05 (9)**
Exhaustion	Cynicism	Non-WEIRD	7	2,442	.69	.81**	.72	.91	99.31	213.34 (6)**
Exhaustion	Professional efficacy	Non-WEIRD	6	2,209	−.29	−.35**	−.53	−.18	95.62	73.08 (5)**
Cynicism	Professional efficacy	Non-WEIRD	7	2,442	−.46	−.56**	−.8	−.32	99.65	998.05 (6)**

Notes. *K* = number of studies included; *N* = total sample size across *K*; *r* = correlation; r_c = corrected correlation; CI_L = lower confidence interval, CI_U = upper confidence interval; I^2 = *I*-squared statistic (heterogeneity statistic); *Q* = *Q* statistic; *df* = degrees of freedom. **$p < .01$.

Table 4. Publication bias analysis of the correlations between the three burnout factors

Variable 1	Variable 2	Sample	Samples imputed right	*p*-value $R0_{right}$	Samples imputed left	*p*-value $R0_{left}$	Egger's regression estimate [CI_L, CI_U]	*t*-value Egger's regression	*p*-value Egger's regression	FSN
Exhaustion	Cynicism	All	2	.13	0	.50	.99 [.95, 1.02]	51.9647	0	1,239,551
Exhaustion	Professional efficacy	All	0	.50	0	.50	−.37 [−.62, −.12]	−2.8621	.004	4,644
Cynicism	Professional efficacy	All	0	.50	0	.50	−.82 [−.98, −.66]	−9.9652	0	311,472
Exhaustion	Cynicism	WEIRD	1	.30	0	.50	.98 [.91, 1.06]	25.3379	0	549,621
Exhaustion	Professional efficacy	WEIRD	0	.50	0	.50	−.22 [−.47, .04]	−1.6501	.01	1,486
Cynicism	Professional efficacy	WEIRD	2	.13	2	.13	−.64 [−.85, −.44]	−6.2704	0	19,395
Exhaustion	Cynicism	Non-WEIRD	0	.50	0	.50	.99 [.95, 1.03]	5.8603	0	138,369
Exhaustion	Professional efficacy	Non-WEIRD	0	.50	0	.50	−.69 [−1.23, −.15]	−2.4974	.013	870
Cynicism	Professional efficacy	Non-WEIRD	0	.50	0	.50	−.96 [−1.14, −.78]	−1.2459	0	175,398

Notes. Samples imputed with trim-and-fill R0 estimator; Egger's regression using the standard error as predictor value. *p*-values for Egger's regression are significant when $p < .001$. FSN = Fail-Safe-Number.

EX and PE were unrelated in five studies (Bocéréan et al., 2019; Chirkowska-Smolak & Kleka, 2011; Kitaoka-Higashiguchi et al., 2004; Wang et al., 2019; Xanthopoulou et al., 2012), whereas CY and PE were unrelated in three studies (Kitaoka-Higashiguchi et al., 2004; Wang et al., 2019; Xanthopoulou et al., 2012). For this reason, some researchers (e.g., Qiao & Schaufeli, 2011) opted to measure burnout without PE or concluded that PE needed to be removed entirely (Spontón et al., 2019). This pointed to the fact that MBI-defined burnout was not always represented as a syndrome as it ought to be but as a set of eruptively independent symptoms.

Table 3 presents the results of the meta-correlations between the three factors. As can be seen, the correlation between exhaustion and cynicism was extremely high (r_c = .75), while the other correlations were at moderate levels (r_c = −.33 for exhaustion and professional efficacy, and r_c = −.55 for cynicism and professional efficacy). The publication bias analysis suggested that there was a chance that these correlations suffered from publication bias (see Table 4). Specifically, the correlations between exhaustion and cynicism and cynicism and professional efficacy showed an Egger's regression *p*-value lower than .001, which indicated a funnel plot asymmetry. Examination of the funnel plots (see Figures 3 and 4) revealed that larger studies systematically reported larger correlations, while studies with smaller samples systematically reported lower correlations. However, trim and fill and FSN suggested no publication bias in the results. Thus, overall, it could be concluded that there was a moderate danger of publication bias in the correlations.

Internal Consistency and Reliability

Using the COSMIN criteria for rating measurement properties (Mokkink, Prisen, et al., 2018), approximately one-third (i.e., 34%) of studies in our systematic review had a "sufficient" internal consistency rating (i.e., they reported alpha

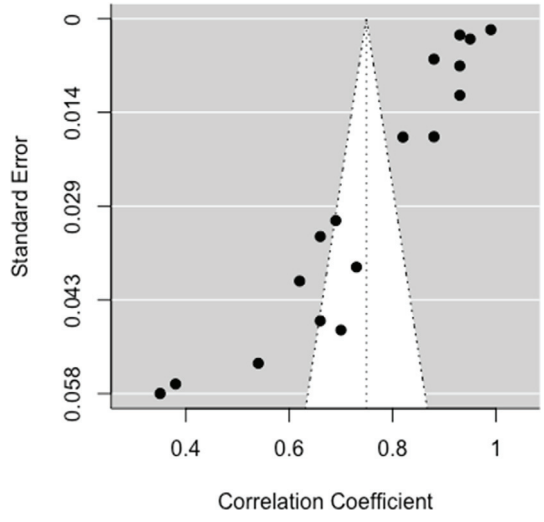

Figure 3. Funnel plot for the Exhaustion – Cynicism correlation.

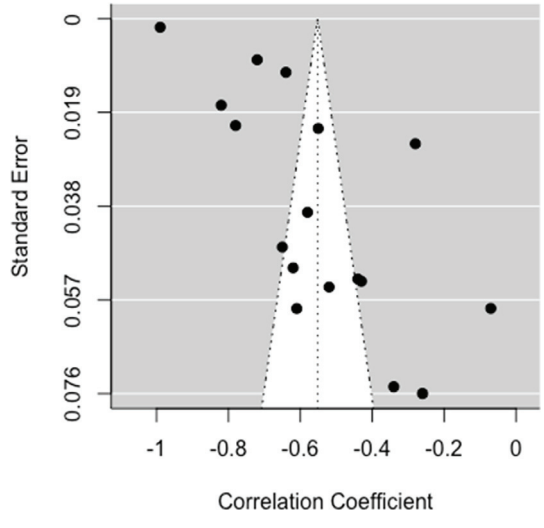

Figure 4. Funnel plot for the Cynicism – Professional Efficacy correlation.

coefficients exceeding .70). Of these, one study (Qiao & Schaufeli, 2011) measured only EX and CY, and three (Kleijweg et al., 2013; Le Roux, 2005; Vanheule et al., 2012) excluded Item 13 before administration. Almost half (i.e., 49%) of the studies had an "indeterminate" internal consistency rating, which stemmed from low evidence of structural validity (i.e., items were deleted, or error variances were correlated). When further analyzing the studies that received a "sufficient" internal consistency rating on an (independent) sample level, some observations were worth noting. For exhaustion, 7% of alpha coefficients were above .70, 71% above .80, and 21% above .90. For cynicism, 57% of alpha coefficients were above .70, whereas 43% were above .80. For professional (in)efficacy, 54% of α coefficients were above .70, whereas 38% were above .80. Relying on the stricter criterion of .80 (Lance et al., 2006; Nunnally, 1978), less than half of these studies provided evidence for the internal consistency of the cynicism and professional (in)efficacy dimensions.

Fourteen percent of the studies had an "insufficient" internal consistency rating. The problem appeared to lie mainly with the CY scale (see Bria et al., 2014; Chirkowska-Smolak & Kleka, 2011), but there were also two instances in which the internal consistency of the PE scale was borderline (i.e., α = .69) (Bedoya & Agudelo, 2016; Taris et al., 1999). Only three studies (Chen et al., 2014; Mäkikangas et al., 2011; Richardsen & Martinussen, 2005) were longitudinal, of which two provided information regarding reliability. In general, the reliability of the instrument was sufficient, with ICC values ranging between .80 and .88 for a two-week interval (Chen et al., 2014) and stability coefficients ranging between .58 and .72 over a period of 6 months (Richardsen & Martinussen, 2005).

The meta-analytic internal consistency (or reliability) results can be seen in Table 5. As the results show, exhaustion had the highest reliability levels (α = .85), while cynicism and professional efficacy had lower but similar, levels (α = .76 and α = .76, respectively). Importantly, in all cases, reliabilities achieved appropriate levels (i.e., α > .70), and there were no significant differences between different categories (i.e., language or [non-]WEIRD classification). Overall, across all studies included in the meta-analysis, reliability levels were generally acceptable when relying on a cut-off of .70, but modest when relying on a cut-off of .80 (Lance et al., 2006; Nunnally, 1978). The only exception was EX, which exceeded the cut-off of .80. When considering the lower confidence interval of the analyses, the CY and PE values of some samples dipped below .70, but never below .62.

Cross-Cultural Validity

Importantly, it was observed that most of the studies included in the systematic review focused more on testing equivalence across different occupational groups (e.g., Bakker et al., 2002; Richardsen & Martinussen, 2005; Shirom & Melamed, 2006; Vanheule et al., 2012) rather than across countries. Although most samples came from non-WEIRD (we relied on Beyebach and colleagues, 2021, to categorize countries) countries, equivalence testing for WEIRD versus non-WEIRD countries was absent. Of those that focused on countries, the following results were reported: scalar invariance in South America (i.e., non-WEIRD; García et al., 2020), configural invariance for Finland, Sweden, and the Netherlands (i.e., WEIRD) (Schutte et al., 2000), and metric invariance for Greece and the Netherlands (i.e., WEIRD) (Xanthopoulou et al., 2012). It is evident that the results from the measurement

Table 5. Pooled internal consistencies across different samples

Variables	Full sample	Spanish sample	Dutch sample	WEIRD sample	Non-WEIRD sample
Exhaustion	.85 [.80, .90]	.86 [.78, .94]	.76 [.66, .86]	.89 [.81, .97]	.85 [.78, .92]
Cynicism	.76 [.70, .82]	.72 [.62, .83]	.76 [.72, .80]	.79 [.70, .87]	.78 [.71, .84]
Professional efficacy	.76 [.70, .81]	.79 [.65, .93]	.76 [.69, .83]	.76 [.69, .83]	.76 [.69, .82]

Notes. Confidence intervals are reported in brackets.

invariance tests for equivalence across the studies were inconsistent and varied widely. This variation may be attributed to the fact that these studies were conducted over the last several decades, during which understanding (e.g., Putnick & Bornstein, 2016) and methodological checklists (e.g., van de Schoot et al., 2012) for testing for invariance evolved significantly. Overall, the evidence for equivalence was somewhat weak, and it would be irresponsible to engage in cross-country comparison based on the included studies, except perhaps within South America.

Criterion Validity

The MBI-GS is often used as a criterion measure in studies that validate other burnout instruments (De Beer, Schaufeli, & Bakker, 2022; De Beer, Schaufeli, & De Witte, 2022; Redelinghuys & Morgan, 2023). Mokkink, Prisen, and colleagues (2018, p. 12) define criterion validity as "the degree to which the scores of a PROM are an adequate reflection of a 'gold standard'". Therefore, criterion validity in the context of the present study was to evaluate its association with other burnout instruments (e.g., the Burnout Measure [BM], the Shirom-Melamed Burnout Measure [SMBM], and the Oldenburg Burnout Inventory [OLBI]). Here, results from our narrative analysis indicated that the associations between the burnout dimensions, as measured by the MBI-GS, and the dimensions from the BM (r = −.57 to .86), SMBM (r = .28 to .87), and OLBI (r = −.58 to .85) were most often significant and in the expected directions (Demerouti et al., 2003; Qiao & Schaufeli, 2011; Shirom & Melamed, 2006). These results provided some evidence for the criterion validity of the MBI-GS, even though it should be noted that the other instruments (particularly the BM and SMBM) only measure EX, and none of them (including the OLBI) measures PE (or professional inefficacy).

In our study, the criterion validity also extended to the associations between the MBI-GS and important work-related antecedents and outcomes. Studies in the systematic review demonstrated that EX and CY showed significant positive associations with job characteristics (or demands) such as workload (Bria et al., 2014; Shirom & Melamed, 2006), emotional demands, and negative work-home interference. PE was positively associated with negative work-home interference (Bria et al., 2014). EX and CY were associated negatively with job resources such as job control, whereas the opposite was true for PE (Shirom & Melamed, 2006).

Significant associations were also reported for "outcomes" of burnout. For example, EX and CY correlated negatively with dimensions of work engagement, whereas PE correlated positively with them (Schaufeli et al., 2002; Xanthopoulou et al., 2012). EX and CY also correlated positively with mental health disorders/problems (e.g., stress, anxiety, depression, agoraphobia), whereas PE correlated negatively (Bocéréan et al., 2019; Roelofs et al., 2005). Similar patterns were observed for job satisfaction (Bravo et al., 2021) and turnover intention (Wang et al., 2019). Even though there were some exceptions (e.g., EX being unrelated to work engagement and differential mental health outcomes for the three dimensions) (Roelofs et al., 2005; Xanthopoulou et al., 2012), the majority of studies in the systematic review supported the criterion validity of the MBI-GS.

However, meta-analytic exploration of the relationships (see Table 6) showed that only exhaustion had a significant and positive relationship with mental health disorders/problems (b = .50, p < .05). Professional efficacy had a marginally non-significant relationship with mental health disorders/problems, as its confidence interval nearly reached the threshold of statistical significance by thousandths of a decimal (b = −.27 [−.55, .003]). The lack of significance here might be due to high dispersion in the results and could indicate the presence of moderators.

Discussion

The central purpose of our review was to evaluate the measurement properties (i.e., validity, internal consistency, and reliability) of the MBI-GS and to critically assess the evidence and quality of studies related to this inventory – which has not been done to date. Specifically, our review focused on research primarily aimed at validating the MBI-GS.

The prevailing body of literature provided moderate evidence for a three-dimensional construct exhibiting modest internal consistency among employees. Our internal consistency results are aligned with a meta-analysis conducted on the Maslach Burnout Inventory (HSS and ESS;

Table 6. Meta-SEM results for the three burnout factors on mental health disorders/problems

	Estimate (*SE*)	CI_L	CI_U
Exhaustion	.50 (.26)*	.002	1.004
Cynicism	−.12 (.38)	−.87	0.62
Professional Efficacy	−.27 (.14)	−.55	0.003

Notes. CI_L = lower confidence interval, CI_U = upper confidence interval. *$p < .05$

Wheeler et al., 2011). Specifically, we found that all dimensions generally scored above .70 and that, among the three dimensions, exhaustion was more likely to show values above .80.

Furthermore, some caveats were identified that warrant presentation. The review unveiled common practices such as correlating item residuals and/or removing items, notably Item 13, to achieve model fit. For this reason, some burnout researchers resorted to a 15-item version (Bria et al., 2014; Storm & Rothmann, 2003; Xanthopoulou et al., 2012) or used such a version from the start (Kleijweg et al., 2013; Le Roux, 2005; Vanheule et al., 2012). This can project an illusion of the original MBI-GS 16-item version being utilized when, in reality, an adapted version is being used that deviates from the official MBI-GS manual's guidelines (see Schaufeli et al., 1996). These practices, furthermore, translate into sample-specific solutions, which we could not replicate in the meta-analysis. However, our review also detected evidence of a two-factor model fitting the data (Pando Moreno et al., 2015; Spontón et al., 2019), and the meta-analytic EFA suggested that, although a three-factor solution seemed to fit the data better, a two-factor solution was also plausible. EX (i.e., the energetic component) and CY (i.e., the motivation component) would cluster together in this instance, and PE would be a distinct factor (Schaufeli & Taris, 2005).

Moderate evidence existed for the configural (cross-cultural) validity of the three-factor solution. Three studies (García et al., 2020; Schutte et al., 2000; Xanthopoulou et al., 2012) supported replicating the three-factor solution at the country level. However, this was only possible with model modification at the structural validity phase. None of the included studies evaluated invariance between WEIRD and non-WEIRD countries. Among WEIRD countries, the results were mixed. One study reported only configural invariance – the same factor structure across countries (Schutte et al., 2000) – whereas another reported weak invariance (i.e., the same factor structure and factor loadings) (Xanthopoulou et al., 2012). However, no studies systematically used the procedure of increasingly constrained parameters as has become the norm: configural, metric, scalar, and strict models (see Putnick & Bornstein, 2016).

The MBI manual unequivocally instructs against calculating a total burnout score or diagnosing any medical condition (Maslach & Leiter, 2021). However, an overall score is also needed because burnout is characterized as a "syndrome" (World Health Organization, 2019). In our review, second-order models, which were not superior fitting to the first-order three-factor models, were attempted to model a total score for the MBI-GS (Vanheule et al., 2012; Wang et al., 2019). Intriguingly, no studies with bifactor models or (bifactor) exploratory structural equation modeling (ESEM) techniques (Morin, 2023) met our inclusion criteria. This could be due to the age of the MBI-GS and the fact that these techniques might still be perceived as too complex to implement and have only become popular more recently. ESEM, specifically bifactor-ESEM (BESEM), is ideally suited to capture the "syndrome" element in a global factor and construct-relevant multidimensionality in its non-target cross-loadings, with the importance of the underlying components being captured in the specific factors (Morin et al., 2020). Therefore, BESEM can model the MBI-GS as an overall score – a global syndrome score – from the three underlying components, but also consider their unique relevance in the presence of the global factor (see Morin, 2023).

The BESEM framework could, furthermore, help determine and illustrate the role of PE in burnout. For many years, there has been debate about the conceptualization and measurement of burnout, in which the legitimacy of PE (and inefficacy, by extension) as a core component of the burnout syndrome has been questioned (e.g., Schaufeli & Taris, 2005). This state of affairs has led to situations where groups of researchers have attempted to redefine burnout as exhaustion only (Guseva-Canu et al., 2021) or conceptualize it as containing similar (i.e., EX and mental distance instead of CY), but also additional components (i.e., cognitive and emotional impairment) in the syndrome (Schaufeli et al., 2020; Schaufeli, 2021).

Similarly, in our review, some researchers opted for a two-factor solution, in which PE was already omitted at the administration phase (Qiao & Schaufeli, 2011), based on recommendations by Walkey and Green (1992), or concluded that a solution without PE had a superior fit (Spontón et al., 2019). This is not surprising, as some of the included studies (Bocéréan et al., 2019; Chirkowska-Smolak & Kleka, 2011; Kitaoka-Higashiguchi et al., 2004; Wang et al., 2019; Xanthopoulou et al., 2012) indicated that PE might be unrelated to EX and/or CY. Our narrative analysis indicated that a change in the valence of professional inefficacy was not necessarily helpful in resolving the PE debate, as authors still needed to delete PE items (Shirom & Melamed, 2006). As the meta-analytic EFA illustrated that PE was a prominent third factor, with high factor loadings for the PE items onto their a priori factor, the omission of PE might be counterproductive. This would be even more so because a study using the MBI-GS for the

criterion or convergent validity of other burnout measures (e.g., the Burnout Assessment Tool) revealed interesting results regarding PE when modeling burnout (as measured by the MBI-GS) as a BESEM. Specifically, it showed that PE contributed less to the global burnout score compared to its specific factor but nonetheless contributed (De Beer, Schaufeli, & De Witte, 2022).

The prevailing body of literature provided moderate evidence for criterion validity. The MBI-GS factors correlated with other well-known burnout measures (Demerouti et al., 2003; Qiao & Schaufeli, 2011; Shirom & Melamed, 2006). However, caution should be exercised, as these criterion measures essentially measure exhaustion. Complicating conceptual matters is the overlap between the meta-analytic and some narrative results (Roelofs et al., 2005; Xanthopoulou et al., 2012) regarding the (non-significant) role of CY and PE in developing our constructed mental health disorders/problems construct comprising anxiety, depression, somatization, etc. This may cast doubt on the criterion validity of CY and PE (specifically with regard to health-related disorders/problems) as measured by the MBI-GS and adds to the current debate on whether burnout is more than just exhaustion (see Guseva-Canu et al., 2021; Schaufeli, 2021). However, given the uncertainty surrounding the structural validity of MBI-defined burnout (i.e., two versus three factors with[out] PE), the cross-sectional nature of the included studies, and the limited variety in "outcome" variables used (e.g., absence of well-being and performance indicators), this should be interpreted with caution.

All in all, our review indicated that the measurement properties of the MBI-GS did raise some questions. Firstly, the jury is still out on the structural validity of the MBI-GS, as different solutions are plausible. Matters are complicated when considering whether a total score is necessary to capture and identify a syndrome. Secondly, internal consistency is perhaps less concerning, challenging the common practice of excluding items. Thirdly, replicating factor structures within, but more importantly across, countries warrants attention. Lastly, the mediocre criterion validity of some of the MBI-GS scales in predicting mental health disorders/problems, specifically the CY and PE scales, raises interesting questions. These *psychometric* issues add to the *conceptual* debate regarding MBI-defined burnout (see Bianchi et al., 2019, 2022; Guseva-Canu et al., 2021; Schaufeli, 2021), which is seemingly the cornerstone of the World Health Organization's operationalization of burnout.

Recommendations

Primary validation studies have been comparatively sparse, particularly in a combination of WEIRD and non-WEIRD countries. This, combined with inconsistencies in reporting/results and frequent use of adapted versions of the MBI-GS, points to the necessity for more robust validation studies to ensure reliable and valid results, specifically in diverse linguistic and cultural settings, using more modern methods.

These studies, which could include a complete investigation into construct-relevant multidimensionality (CFA, ESEM, bifactor-CFA, BESEM) with measurement invariance (configural, metric, scalar/threshold, strict), are needed (Morin, 2023). This should also include the potential of PE as a reversed-scored scale. This will elucidate the factor structure and cross-cultural structure of the MBI-GS (16 items, 15 items, 10 items, or others) to be used in burnout research. It will also answer whether a total score with the MBI-GS is viable to identify burnout risk, in line with its conceptualization as a syndrome. Resolving the debate regarding the most optimal structure of the MBI-GS may also assist the developers in considering potential revisions that could improve its accuracy and utility for research and practice.

Based on the results of these suggested studies, if the MBI-GS remains a viable measure of burnout, we recommend incorporating the findings into a novel research version of the MBI. This iteration of the MBI should be available freely for academic purposes so that authors can reveal items in their manuscripts. The other commercial versions of the MBI should then be marked as not suitable for research purposes if they remain unchanged in light of new psychometric evidence. This would align with the modern movement toward open science practices and could help ensure that the MBI-GS remains a valid and reliable tool to assess burnout risk in the ongoing debate about the measurement and conceptualization of burnout. The potential challenges to the MBI-GS must be taken seriously, investigated, and reported to ensure its robustness and for burnout research to move forward.

Limitations

A strength of this study is that it followed the PRISMA guidelines to ensure quality and comprehensiveness, but some other limitations should be noted. Although we followed rigorous procedures to search for studies, we cannot guarantee that the reference list is complete. Furthermore, a call was issued for unpublished material, without a response, but it does not mean that no such material exists. This study only included research that primarily aimed to validate the MBI-GS and not research that may have inadvertently or secondarily included validation information, such as models investigating the relationship between MBI-GS-assessed burnout and musculoskeletal health outcomes.

The proprietary nature of the MBI-GS complicated the matching of item numbering with the original numbering in some studies. Authors were also reluctant to share any item list used with us privately due to their concern about violating copyright. The risk of bias assessments revealed some challenges with the quality of the primary studies. Although their findings were mostly aligned with those from good-quality studies, they may have an impact on the trustworthiness of our findings. Furthermore, the COSMIN initiative is strongly oriented toward clinical research. Although this is considered a strength, the application and adaptation of its tools in the current study may pose limitations (e.g., overly strict standards). Finally, the criterion validity results may be biased, given the small number and heterogeneity of studies used to establish criterion validity (i.e., mental health disorders/problems).

References

*References included in the systematic review.

**References included in the meta-analyses.

Aromataris, E., & Munn, Z. (2020). *JBI manual for evidence synthesis*,. https://doi.org/10.46658/JBIMES-20-01

**Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2002). Validation of the Maslach Burnout Inventory – General Survey: An Internet study. *Anxiety, Stress & Coping, 15*(3), 245–260. https://doi.org/10.1080/1061580021000020716

**Bedoya, L. M. G., & Agudelo, N. O. (2016). Propiedades psicométricas de confiabilidad y validez del Maslach Burnout Inventory – General Survey [Psychometric properties of reliability and validity of the Maslach Burnout Inventory – General Survey]. *Revisita Interamericana de Psicología Ocupacional, 33*(2), 128–142. http://revista.cincel.com.co/index.php/RPO/article/view/162

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement, 65*(5), 676–696. https://doi.org/10.1177/0013164404272507

Beyebach, M., Neipp, M. C., Solanes-Puchol, Á., & Martín-del-Río, B. (2021). Bibliometric differences between WEIRD and non-WEIRD countries in the outcome research on solution-focused brief therapy. *Frontiers in Psychology, 12*, Article 754885. https://doi.org/10.3389/fpsyg.2021.754885

Bianchi, R., Schonfeld, I. S., & Laurent, E. (2019). Burnout: Moving beyond the status quo. *International Journal of Stress Management, 26*(1), 36–45. https://doi.org/10.1037/str0000088

Bianchi, R., Wac, K., Sowden, J. F., & Schonfeld, I. S. (2022). Burned-out with burnout? Insights from historical analysis. *Frontiers in Psychology, 13*, Article 993208. https://doi.org/10.3389/fpsyg.2022.993208

**Bocéréan, C., Dupret, E., & Feltrin, M. (2019). Maslach Burnout Inventory–General Survey: French validation in a representative sample of employees. *SCIREA Journal of Health, 3*(2), 24–41. https://hal.science/hal-02480484/document

Bradley, H. B. (1969). Community-based treatment for young adult offenders. *Crime and Delinquency, 15*(2), 359–370. https://doi.org/10.1177/001112876901500307

**Bravo, D. M., Suárez-Falcón Juan, C., Bianchi, J. M., Segura-Vargas, M. A., Ruiz, F. J., & Tchounwou, P. B. (2021). Psychometric properties and measurement invariance of the Maslach Burnout Inventory – General Survey in Colombia. *International Journal of Environmental Research and Public Health, 18*(10), Article 5118. https://doi.org/10.3390/ijerph18105118

*Bria, M., Spânu, F., Băban, A., & Dumitraşcu, D. L. (2014). Maslach Burnout Inventory – General Survey: Factorial validity and invariance among Romanian healthcare professionals. *Burnout Research, 1*(3), 103–111. https://doi.org/10.1016/j.burn.2014.09.001

**Chen, W. S., Haniff, J., Siau, S. C., Seet, W., Loh, F. S., Jamil, A. H. M., Sa'at, N., & Baharum, N. (2014). Translation, cross-cultural adaptation and validation of the Malay version of the Maslach Burnout Inventory (MBI) in Malaysia. *International Journal of Social Science Studies, 2*(2), 66–74. https://doi.org/10.11114/ijsss.v2i2.309

Cheung, M. W.-L. (2015a). *Meta-analysis: A structural equation modeling approach*. Wiley.

Cheung, M. W. (2015b). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology, 5*, Article 1521. https://doi.org/10.3389/fpsyg.2014.01521

**Chirkowska-Smolak, T., & Kleka, P. (2011). The Maslach Burnout Inventory – General Survey: validation across different occupational groups in Poland. *Polish Psychological Bulletin, 42*(2), 86–94. https://doi.org/10.2478/v10059-011-0014-x

Clark, J., Glasziou, P., Del Mar, C., Bannach-Brown, A., Stehlik, P., & Scott, A. M. (2020). A full systematic review was completed in 2 weeks using automation tools: A case study. *Journal of Clinical Epidemiology, 28*(121), 81–90. https://doi.org/10.1016/j.jclinepi.2020.01.008

De Beer, L. (2023). *Systematic Review and Meta-analysis of the MBI-GS* [Data]. https://doi.org/10.17605/OSF.IO/N9V5T

De Beer, L. T., Schaufeli, W. B., & Bakker, A. B. (2022). Investigating the validity of the short form Burnout Assessment Tool: A job demands-resources approach. *African Journal of Psychological Assessment, 4*, Article a95. https://doi.org/10.4102/ajopa.v4i0.95

De Beer, L. T., Schaufeli, W. B., & De Witte, H. (2022). The psychometric properties and measurement invariance of the Burnout Assessment Tool (BAT-23) in South Africa. *BMC Public Health, 22*, Article 1555. https://doi.org/10.1186/s12889-022-13978-0

**Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology, 86*(3), 499–512. https://doi.org/10.1037/0021-9010.86.3.499

Demerouti, E., Bakker, A. B., Vardakou, I., & Kantas, A. (2003). The convergent validity of two burnout instruments: A multitrait-multimethod analysis. *European Journal of Psychological Assessment, 19*(1), 12–23. https://doi.org/10.1027/1015-5759.19.1.12

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

Gamer, M., Lemon, J., Singh, I. F., & Fellows, I. (2019). *irr: Various coefficients of interrater reliability and agreement* [Software]. https://CRAN.R-project.org/package=irr

*García, A. J., Soto, C. M., Arata, M. F., Jiménez, C. A. F., Caraballo, M., & Cristiá, C. C. (2020). Validación transcultural y funcionamiento diferencial del MBI-GS en docentes de tres países latinoamericanos [Cross-cultural validation and differential functioning of the MBI-GS in teachers from three Latin American countries]. *Avances en Psicología Latinoamericana, 38*(1), 135–156. https://doi.org/10.12804/revistas.urosario.edu.co/apl/a.6621

**Gil-Monte, P. R. (2002). Validez factorial de la adaptación al español del Maslach Burnout Inventory-General Survey [Factorial validity of the Spanish adaptation of the Maslach Burnout Inventory – General Survey]. *Salud Pública México, 44*, 33–40. http://www.insp.mx/salud/index.html

Greco, L. M., O'Boyle, E. H., Cockburn, B. S., & Yuan, Z. (2018). Meta-analysis of coefficient alpha: A reliability generalization study: Meta-analysis of coefficient alpha. *Journal of Management Studies, 55*(4), 583–618. https://doi.org/10.1111/joms.12328

Guseva-Canu, I., Marca, S. C., & Cernițanu, M. (2021). Harmonized definition of occupational burnout: A systematic review, semantic analysis, and Delphi consensus in 29 countries. *Scandinavian Journal of Work, Environment and Health, 47*(2), 95–107. https://doi.org/10.5271/sjweh.3935

Halbesleben, J. R., & Demerouti, E. (2005). The construct validity of an alternative measure of burnout: Investigating the English translation of the Oldenburg Burnout Inventory. *Work & Stress, 19*(3), 208–220. https://doi.org/10.1080/02678370500340728

Henrich, J. (2020). *The WEIRDest people in the world: How the West became psychologically peculiar and particularly prosperous*. MacMillan.

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

Iliescu, D., Rusu, A., Greiff, S., Fokkema, M., & Scherer, R. (2022). Why we need systematic reviews and meta-analyses in the testing and assessment literature [Editorial]. *European Journal of Psychological Assessment, 38*(2), 73–77. https://doi.org/10.1027/1015-5759/a000705

**Kitaoka-Higashiguchi, K., Ogino, K., & Masuda, S. (2004). 日本版MBI-GS (Maslach Burnout Inventory-General Survey) の妥当性の検討 [Validation of a Japanese research version of the Maslach Burnout Inventory – General Survey]. *The Japanese Journal of Psychology, 75*(5), 415–419. https://doi.org/10.4992/jjpsy.75.415

*Kleijweg, J. H. M., Verbraak, M. J. P. M., & Van Dijk, M. K. (2013). The clinical utility of the Maslach Burnout Inventory in a clinical population. *Psychological Assessment, 25*(2), 435–441. https://doi.org/10.1037/a0031334

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*(2), 202–220. https://doi.org/10.1177/1094428105284919

*Langballe, E. M., Falkum, E., Innstrand, S. T., & Aasland, O. G. (2006). The factorial validity of the Maslach Burnout Inventory – General Survey in representative samples of eight different occupational groups. *Journal of Career Assessment, 14*(3), 370–384. https://doi.org/10.1177/1069072706286497

*Le Roux, A. M. (2005). *The validation of two burnout measures in the South African earthmoving equipment industry* [Unpublished masters' thesis]. North-West University.

**Leiter, M. P., & Schaufeli, W. B. (1996). Consistency of the burnout construct across occupations. *Anxiety, Stress and Coping, 9*(3), 229–243. https://doi.org/10.1080/10615809608249404

Maslach, C., & Jackson, S. E. (1986). *Maslach Burnout Inventory manual* (2nd ed.). Consulting Psychologists Press.

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. (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

Meuleman, B., Żółtak, T., Pokropek, A., Davidov, E., Muthén, B., Oberski, D. L., Billiet, J., & Schmidt, P. (2023). Why measurement invariance is important in comparative research. A response to Welzel et al. (2021). *Sociological Methods & Research, 52*(3), 1401–1419. https://doi.org/10.1177/00491241221091755

**Millán de Lange, A. C., & D'Aubeterre López, M. E. (2012). Propiedades psicométricas del Maslach Burnout Inventory-GS en una muestra multiocupacional venezolana [Psychometric properties of the Maslach Burnout Inventory – GS in a multi-occupational Venezolean sample]. *Revista de Psicología (Pucp), 30*(1), 103–128. http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S0254-92472012000100005

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1171–1179. https://doi.org/10.1007/s11136-017-1765-4

Mokkink, L. B., Prisen, C. A. C., Partick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2018). *COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs): User guide*. https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf

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 Press. https://handbookofsem.com/

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* (4th ed., pp. 1044–1073). Wiley. https://doi.org/10.1002/9781119568124.ch51

**Mäkikangas, A., Hätinen, M., Kinnunen, U., & Pekkonen, M. (2011). Longitudinal factorial invariance of the Maslach Burnout Inventory – General Survey among employees with job-related psychological health problems. *Stress and Health, 27*(4), 347–352. https://doi.org/10.1002/smi.1381

**Mojsa, J., Dylag, A., & Palczynska, E. (2006). Psychometric properties of a Polish version of the Maslach Burnout Inventory General Survey (MBI-GS) in a group of information and communication technology (ICT) specialists. *Ergonomia IJE&HF, 28*(4), 351–361. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi= e1fd988d72a5f3fa9ecaeb0f3cf3aff647a448c5

**Moreno-Jiménez, B., Rodríguez-Carvajal, R., & Escobar Redonda, E. (2001). La evaluación del burnout profesional factorialización del MBI–GS. Un análisis preliminar [The assessment of burnout in work settings: MBI–GS factor preliminary analysis]. *Ansiedad y Estrés, 7*(1), 69–78. http://www.integraorg.com/wp-content/docs/La%20Evaluacion%20del%20Burnout%20Profesional.pdf

**Nguyen, H. T. T., Kitaoka, K., Sukigara, M., & Thai, A. L. (2018). Burnout study of clinical nurses in Vietnam: Development of job burnout model based on Leiter and Maslach's theory. *Asian Nursing Research, 12*(1), 42–49. https://doi.org/10.1016/j.anr.2018.01.003

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *British Medical Journal, 372*, Article n73. https://doi.org/10.1136/bmj.n71

**Pando Moreno, M., Aranda Beltrán, C., & López Palomar, M. (2015). Validez factorial del Maslach Burnout Inventory – General Survey en ocho países Latinoamericanos [Factorial validity of the Maslach Burnout Inventory – General Survey in eight Latin American countries]. *Ciencia & Trabajo, 17*, 28–31. https://doi.org/10.4067/S0718-24492015000100006

Prinsen, C. A., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., Williamson, P. R., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "core outcome set" – A practical guideline. *Trials, 17*(1), Article 449. https://doi.org/10.1186/s13063-016-1555-2

Pines, A. M., & Aronson, R. (1988). *Career burnout: Causes and cures*. Free Press.

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

*Qiao, H., & Schaufeli, W. B. (2011). The convergent validity of four burnout measures in a Chinese sample: A confirmatory factor-analytic approach. *Applied Psychology: An International Review, 60*(1), 87–111. https://doi.org/10.1111/j.1464-0597.2010.00428.x

R Core Team. (2021). *R: A language and environment for statistical* (4.2.2). R foundation for statistical computing.

Redelinghuys, K., & Morgan, B. (2023). Psychometric properties of the Burnout Assessment Tool across four countries. *BMC Public Health, 23*, Article 824. https://doi.org/10.1186/s12889-023-15604-z

Revelle, W. (2023). *psych: Procedures for psychological, psychometric, and personality research* (2.3.3). https://CRAN.R-project.org/package=psych

*Richardsen, A. M., & Martinussen, M. (2005). Factorial validity and consistency of the MBI-GS across occupational groups in Norway. *International Journal of Stress Management, 12*(3), 289–297. https://doi.org/10.1037/1072-5245.12.3.289

Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods, 11*(3), 306–322. https://doi.org/10.1037/1082-989X.11.3.306

**Roelofs, J., Verbraak, M., Keijsers, G. P., De Bruin, M. B., & Schmidt, A. J. (2005). Psychometric properties of a Dutch version of the Maslach Burnout Inventory General Survey (MBI-DV) in individuals with and without clinical burnout. *Stress and Health, 21*(1), 17–25. https://doi.org/10.1002/smi.1032

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

Schaufeli, W. B. (2017). Burnout: A short socio-cultural history. In S. Neckel, A. K. Schaffner, & G. Wagner (Eds.), *Burnout, fatigue, exhaustion: An interdisciplinary perspective on a modern affliction* (pp. 105–127). Palgrave Macmillan. https://10.1007/978-3-319-52887-8_5

Schaufeli, W. B. (2023). Burnout: A critical overview. In L. M. Lapierre & C. Cooper (Eds.), *The Cambridge companion to organizational stress and well-being* (pp. 214–259). Cambridge University Press. https://doi.org/10.1017/9781009268332

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

Schaufeli, W. B., Leiter, M. P., Maslach, C., & Jackson, S. E. (1996). Maslach Burnout Inventory – General Survey. In C. Maslach, S. E. Jackson, & M. P. Leiter (Eds.), *The Maslach Burnout Inventory – Test Manual* (3rd ed., pp. 19–26). Consulting Psychologists Press.

Schaufeli, W. B., & Salanova, M. (2007). Efficacy or inefficacy, that's the question: Burnout and work engagement, and their relationships with efficacy beliefs. *Anxiety, Stress, & Coping, 20*(2), 177–196. https://doi.org/10.1080/10615800701217878

Schaufeli, W. B., Salanova, M., Gonzalez-Roma, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A confirmative analytic approach. *Journal of Happiness Studies, 3*(1), 71–92. https://doi.org/10.1023/a:1015630930326

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

*Schuster, M. S., Dias, V. V., Battistella, L. F., & Grohmann, M. Z. (2015). Validação da escala MBI-GS: Uma investigação general survey sobre a percepção de saúde dos colaboradores [Validation of the MBI-GS Scale: A general survey investigation on employees' perception of health]. *Revista de Gestão, 22*(3), 403–416. https://doi.org/10.5700/rege569

*Schutte, N., Toppinen, S., Kalimo, R., & Schaufeli, W. (2000). The factorial validity of the Maslach Burnout Inventory – General Survey (MBI-GS) across occupational groups and nations. *Journal of Occupational and Organizational Psychology, 73*(1), 53–66. https://doi.org/10.1348/096317900166877

Shafer, A. (2005). Meta-analysis of the brief psychiatric rating scale factor structure. *Psychological Assessment, 17*(3), 324–335. https://doi.org/10.1037/1040-3590.17.3.324

*Shirom, A., & Melamed, S. (2006). A comparison of the construct validity of two burnout measures in two groups of professionals. *International Journal of Stress Management, 13*(2), 176–200. https://doi.org/10.1037/1072-5245.13.2.176

Siegel, M., Eder, J. S. N., Wicherts, J. M., & Pietschnig, J. (2022). Times are changing, bias isn't: A meta-meta-analysis on publication bias detection practices, prevalence rates, and predictors in industrial/organizational psychology. *Journal of Applied Psychology, 107*(11), 2013–2039. https://doi.org/10.1037/apl0000991

*Spontón, C., Trógolo, M., Castellano, E., & Medrano, L. A. (2019). Medición del burnout: estructura factorial, validez y confiabilidad en trabajadores argentinos [Measurement of burnout: Factor structure, validity and reliability in Argentinean workers]. *Interdisciplinaria, 36*(1), 87–103. https://doi.org/10.16888/interd.2019.36.1.7

*Storm, K., & Rothmann, S. (2003). A psychometric analysis of the Maslach Burnout Inventory – General Survey in the South African police service. *South African Journal of Psychology, 33*(4), 219–226. https://doi.org/10.1177/008124630303300404

**Taris, T. W., Schreurs, P. J. G., & Schaufeli, W. B. (1999). Construct validity of the Maslach Burnout Inventory – General Survey: A two-sample examination of its factor structure and correlates. *Work and Stress, 13*(3), 223–237. https://doi.org/10.1080/026783799296039

*Tomás, J. M., de los Santos, S., Alonso-Andres, A., & Fernández, I. (2016). Validation of the Maslach Burnout Inventory – General Survey on a representative sample of Dominican teachers: Normative data. *The Spanish Journal of Psychology, 19*(e83), 1–9. https://doi.org/10.1017/sjp.2016.91

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

*Vanheule, S., Rosseel, Y., Vlerick, P., Van de Ven, B., & Declercq, F. (2012). Onderzoeksnotitie: De factoriële validiteit en meet-invariantie van de Utrechtse Burnout Schaal. Algemene versie (UBOS-A) [Research note: The factorial validity and measurement invariance of the Dutch version (UBOS-A) of the Maslach Burnout Inventory]. *Gedrag en Organisatie, 25*(2), 192–201. https://doi.org/10.5117/2012.025.002.192

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

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

**Wang, Y., Zhang, H., Lei, J., & Yu, Y. (2019). Burnout in Chinese social work: Differential predictability of the components of the Maslach Burnout Inventory. *International Journal of Social Welfare, 28*(2), 217–228. https://doi.org/10.1111/ijsw.12339

Wheeler, D. L., Vassar, M., Worley, J. A., & Barnes, L. L. (2011). A reliability generalization meta-analysis of coefficient alpha for the Maslach Burnout Inventory. *Educational and Psychological Measurement, 71*(1), 231–244. https://doi.org/10.1177/0013164410391579

World Health Organization. (2019, May 28). *Burn-out an "occupational phenomenon": International classification of diseases*. https://www.who.int/news/item/28-05-2019-burn-out-an-occupational-phenomenon-international-classification-of-diseases

Worley, J. A., Vassar, M., Wheeler, D. L., & Barnes, L. L. B. (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

*Xanthopoulou, D., Bakker, A. B., Kantas, A., & Demerouti, E. (2012). Measuring burnout and work engagement: Factor structure, invariance, and latent mean differences across Greece and the Netherlands. *International Journal of Business Science & Applied Management, 7*(2), 40–52. http://hdl.handle.net/10419/190637

History
Received June 1, 2023
Revision received June 26, 2023
Accepted July 1, 2023
Published online February 5, 2024
EJPA Section / Category I/O Psychology

Conflict of Interest
Wilmar B. Schaufeli, is one of the original developers of the original Maslach Burnout Inventory – General Survey. This conflict was managed by the first three authors taking the lead in the design and execution of the methodology, as well as the writing of the manuscript. All the other authors declare that there was no undue influence or irregularities to report due to this conflict. Furthermore, the other authors declare no known conflicts of interest.

Authorship
Leon T. De Beer, Conceptualization, Data curation, Investigation, Writing – original draft; Leoni van der Vaart, Conceptualization, Methodology, Project administration, Data curation, Investigation, Writing – original draft; Maximiliano Escaffi-Schwarz, Methodology, Investigation, Formal analysis, Writing – original draft; Hans De Witte, Writing – review & editing; Wilmar B. Schaufeli, Writing – review & editing. Leon T. De Beer and Leoni van der Vaart are joint first authors, and their authorship order was determined alphabetically. All authors approved the final version of the manuscript.

Open Science
Open Data: We confirm that there is sufficient information for an independent researcher to reproduce our reported systematic review. All relevant supplementary material can be found online at https://doi.org/10.17605/OSF.IO/N9V5T (De Beer, 2023).
Open Materials: We confirm that there is sufficient information for an independent researcher to reproduce our reported systematic review. All applicable supplementary material can be found online at https://doi.org/10.17605/OSF.IO/N9V5T (De Beer, 2023). Note that the scripts for the analyses used in the meta-analysis can be obtained from the corresponding or third author upon reasonable request.
Preregistration of Study and Analysis Plan: The preregistration of this study was done online at PROSPERO: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=393159

Funding
This project received no internal or external funding from any private/public organization or an individual/group.

ORCID
Leon T. De Beer
https://orcid.org/0000-0001-6900-2192
Hans De Witte
https://orcid.org/0000-0002-6691-517X
Maximiliano Escaffi-Schwarz
https://orcid.org/0000-0002-4371-8739
Leoni Van der Vaart
https://orcid.org/0000-0002-8236-0637

Leon T. De Beer
Department of Psychology
Norwegian University of Science and Technology (NTNU)
Building 12, Edvard Bulls Veg 1
Dragvoll
7491 Trondheim
Norway
leondb@gmail.com

