WORK & STRESS
https://doi.org/10.1080/02678373.2025.2484758

Routledge
Taylor & Francis Group

REVIEW ARTICLE

Burnout Assessment Tool: a reliability generalisation meta-analysis

César Villacura-Herrera [a,b,c], Hedy Acosta-Antognoni [b], Javiera Maldonado [b], Fernanda Arriaza [b], Natalia Cancino-Letelier [d], Marcelo Nvo-Fernández [c] and Wilmar B. Schaufeli [e]

[a]Centro de Investigación en Ciencias Cognitivas (CICC), Faculty of Psychology, Universidad de Talca, Talca, Chile; [b]Research Team on Organizational Psychology (RTOP), Faculty of Psychology, Universidad de Talca, Talca, Chile; [c]Laboratory of Methodology, Behavioral Sciences and Neuroscience, Faculty of Psychology, Universidad de Talca, Talca, Chile; [d]Centro de Psicología Aplicada (CEPA), Faculty of Psychology, Universidad de Talca, Talca, Chile.; [e]Department of Psychology, Utrecht University, Utrecht, Netherlands

ABSTRACT

The aim of this study was to conduct a reliability generalisation meta-analysis (RGMA) for the Burnout Assessment Tool (BAT) on its original and shortened versions based on Cronbach's alpha and McDonald's omega. A systematic search was carried out on six databases, where 56 articles were included in the analyses. Results based on random-effects models show good pooled internal consistency indices for the BAT and its subscales on both its original ($\alpha = .798–.948$; $\omega = .754–.940$) and shortened versions ($\alpha = .763–.907$; $\omega = .750–.909$), while showing high heterogeneity overall ($p_Q \leq .004$; $I^2 \geq 77.59\%$). Sources of variability were examined through influence diagnostics and meta-regression analysis. Our findings confirm that the BAT is a highly reliable tool for the assessment of burnout across different populations and settings. These results demonstrate that this measure shows better reliability indices when compared to the Maslach Burnout Inventory (MBI). Methodological and practical implications for the assessment and diagnosis of burnout in the organisational context are discussed.

ARTICLE HISTORY
Received 24 July 2024
Accepted 21 March 2025

KEYWORDS
Burnout Assessment Tool; burnout; reliability generalization; meta-analysis

1. Introduction

Workplace conditions can trigger potentially negative effects on employees' physical and mental health. Beyond an increase in work pressure and physical exhaustion, the dynamics of this type of environment produce significant wear and tear on a psychological level for people (Salvagioni et al., 2017). This concept, understood as the burnout syndrome, is developed mainly in the work context (Maslach & Leiter, 2016), however, it cannot be limited to it (Bianchi et al., 2014). The burnout syndrome negatively impacts people emotionally, cognitively, and physically (Pines & Maslach, 1978), generating a mental state of disinterest, fatigue and frustration that can lead to depressive symptoms, aggression, and loss of physical energy (Bianchi et al., 2015; Schaufeli et al.,

CONTACT Hedy Acosta-Antognoni hacosta@utalca.cl
Supplemental data for this article can be accessed online at https://doi.org/10.1080/02678373.2025.2484758.

2009). The negative consequences evident in organisations are linked to the economic sphere, due to high rates of absenteeism, lack of work commitment, disability, and medical expenses (Salvagioni et al., 2017; Schaufeli, 2023).

The burnout syndrome has represented a significant challenge and currently constitutes one of the main psychosocial risks, both for people and organisations (Maslach & Leiter, 2008; Schaufeli, 2023). Jackson et al. (1986) pointed out that the burnout syndrome does not have a single definition; however, it highlighted its main dimensions, which correspond to emotional exhaustion, depersonalisation, and a critical sense of personal fulfilment. Likewise, the World Health Organization (WHO) officially recognised the burnout syndrome as a disease, after the latest revision of the International Classification of Diseases (ICD-11), which came into force in January 2022 (WHO, 2022). The syndrome was included as an occupational phenomenon, characterised by three dimensions which are feelings of exhaustion, increased mental distance from work and reduced professional effectiveness (Bianchi & Schonfeld, 2023; Parker & Tavella, 2022).

Freudenberger (1974) first described this phenomenon as a feeling of failure and exhaustion, along with its physical and psychological signs. Later, it was adopted as burnout syndrome, after Christina Maslach's (1982) contribution to the initial definition, adding negative attitudes and personal distancing as an inadequate response to chronic stress. While she initially pointed out that mainly healthcare professionals were the ones who suffered from burnout (Maslach & Jackson, 1982), later authors such as Pines and Aronson (1988) have stated that any professional can suffer from this syndrome.

While the literature on burnout syndrome has increased in recent years as it has evolved into a multidimensional phenomenon that has caused new generations of workers to be strongly affected (Márquez-Lugo et al., 2021), the medical diagnosis has only been established in a few countries, which means that it is not officially recognised in most health care systems (Heinemann & Heinemann, 2017). This, due to most instruments lacking a clinically valid threshold for a definite diagnosis (Edú-Valsania et al., 2022). For this reason, the proper measurement of burnout syndrome is highly relevant to develop primary prevention strategies, interventions, and personalised treatments (Edú-Valsania et al., 2022; Márquez-Lugo et al., 2021).

The measurement of burnout has been operationalised, primarily, through 13 self-report questionnaires: The Maslach Burnout Inventory (MBI), created by Maslach and Jackson (1981); the Burnout Measure (BM; Pines et al., 1981); the Oldenburg Burnout Inventory (OLBI; Demerouti et al., 2003); the Copenhagen Burnout Inventory (CBI; Kristensen, Borritz, et al., 2005); the Copenhagen Psychosocial Questionnaire (COPSOQ; Kristensen, Hannerz, et al., 2005); the BurnOut-Neurasthenia Complaints Scale (BO-NKS; Verbraak et al., 2008); the Shirom Melamed Burnout Measure (SMBM; Shirom & Melamed, 2006); the Boudreau Burnout Questionnaire (BBQ; Boudreau et al., 2006); the Bergen Burnout Inventory (BBI; Salmela-Aro et al., 2011); the Spanish Burnout Inventory (Gil-Monte & Faúndez, 2011); the Granada Burnout Questionnaire (GBQ; de la Fuente et al., 2015); the Hamburg Burnout Inventory (HBI; Burisch, 2017) and the Burnout Assessment Tool (BAT; Schaufeli et al., 2020). Of all those mentioned, the instrument most adapted and used internationally is the MBI, which has been used in about 90% of studies assessing for burnout syndrome (Erschens et al., 2019; Schaufeli et al., 2020).

However, despite its widespread use for the assessment of burnout, the MBI has been shown to have a number of limitations. First, it does not consider the inclusion of cognitive and emotional impairment resulting from a diminished functional capacity to regulate one's own internal processes when assessing burnout symptoms (Schaufeli & De Witte, 2023). It also lacks clinically validated cut-off criteria and statistical norms that are based on representative international samples (Kristensen, Hannerz, et al., 2005; Schaufeli et al., 2009; Schaufeli & Taris, 2005). It has also been suggested that this measure may not be robust for different cultural and organisational settings, as it was developed primarily for its use in western professional environments (Schaufeli & Enzmann, 1998), or that it directly fails to assess the underlying construct of burnout (Bianchi et al., 2024). In addition to the issues mentioned above, the MBI was designed as a research instrument rather than an evaluation or diagnosis tool (De Beer et al., 2024a; Moreno et al., 2006), so it does not guarantee efficient practical use (Sakakibara et al., 2020). Under this context, Schaufeli et al. (2020) developed the BAT through a deductive (theoretical) and inductive (empirical) approach, addressing the deficiencies and practical issues of previously developed instruments.

Among the advantages of the BAT is that its full version has proven to be useful in people who are currently working and in those who are on a short sick leave due to burnout to monitor their progress in treatment, counselling or return to work programmes (Schaufeli et al., 2020). In addition, its approach towards the concept of burnout allows it to encompass a broader spectrum of this syndrome considering different dimensions of its symptomatology, thus improving its sensitivity and specificity of use in different occupational settings (Schaufeli et al., 2019). Furthermore, research on the BAT has recently clinically validated cut-offs for assessing burnout using receiver operating characteristic (ROC) analysis. These results were based on patient samples from three European countries, where relatively similar cut-offs were obtained for the BAT-C and the BAT-12 (Schaufeli et al., 2023). For the BAT-C, a mean score of 2.59 would indicate to be "At risk," while a score of 3.02 would be indicative of "severe burnout." Psychometric studies have also found robust evidence for sustaining its validity across different cultural contexts and settings. Studies in countries such as Belgium (Schaufeli et al., 2020), Japan (Sakakibara et al., 2020), South Africa (De Beer, Schaufeli, & Bakker, 2022), Italy (Mazzetti et al., 2022) and Spain (Gómez et al., 2022) have examined the psychometric properties of the measure and its shortened versions, also demonstrating its strong convergent validity with well-established measures such as the MBI and the Oldenburg Burnout Inventory (OLBI; Demerouti et al., 2003), deeming the measure as a valid and versatile tool across multiple cultural groups and contexts such as healthcare workers, employees, and teachers.

Regarding the internal structure of the instrument, the BAT is composed of 33 items assessing for symptoms of burnout. The measure further classifies these symptoms into two main dimensions: Core and Secondary symptoms. The Core symptoms dimension, also known as the BAT-C, is comprised by 23 items assessing for exhaustion, mental distance, cognitive impairment, and emotional impairment. On the other hand, the Secondary symptoms dimension, also called BAT-S and comprised by 10 items, measures an individuals' psychological distress, and psychosomatic complaints. Participants respond based on a 5-point Likert-type frequency scale: 1 (never), 2

(rarely), 3 (sometimes), 4 (often) and 5 (always) (Schaufeli et al., 2019, 2020; Schaufeli & De Witte, 2023). Two brief versions of the measure have also been developed to date, focused exclusively on the four Core symptoms: The 12-item Short BAT (BAT-12; De Beer, Schaufeli, & Bakker, 2022), and the more recent 4-item Ultra Short version (BAT-4; Hadžibajramović et al., 2024). Since its formal release in 2019 (Schaufeli et al., 2019), the BAT has been used in a wide variety of countries and contexts, showing evidence sustaining its validity across multiple settings (Hadžibajramović, Hansson, et al., 2022). It has also been adapted to different languages such as Italian, Japanese, Portuguese, and Spanish (Consiglio et al., 2021; Sakakibara et al., 2020; Sinval et al., 2022). However, while the ability of the BAT to provide an adequate assessment of burnout has been demonstrated, an examination of its overall reliability has not been explored yet.

Reliability is not a static property of a test, but of the test scores obtained in a specific application (Sánchez-Meca & López-Pina, 2008). The reliability of a test's scores can vary depending on the composition, the characteristics of the samples and the context of assessment (Sánchez-Meca et al., 2009). Since reliability scores may experience changes in each evaluation, researchers should always calculate it based on the data of their respective samples. Despite this, it is common for researchers not to report the reliability indices of the instrument obtained with their data, but to induce them and assume that these indices are adequate based on previous research (Sánchez-Meca et al., 2021). Therefore, a statistical procedure that represents one of the most robust ways to obtain information on how the reliability coefficients obtained in different administrations of a test vary is through reliability generalisation meta-analysis (RGMA; Vacha-Haase, 1998). This psychometric meta-analytic approach allows obtaining an average estimate of the reliability of a test's scores, establishes whether the reliability coefficients obtained in the various applications of a test are heterogeneous among themselves and investigates what characteristics of the test, studies and the sample can account for said heterogeneity (Botella & Sánchez-Meca, 2015; Sánchez-Meca et al., 2021).

With this, the relevance of RGMA lies in estimating the expected reliability of the scores of a test, and communicating which test formats and application conditions could affect the reliability estimates (Sánchez-Meca et al., 2021). Likewise, this research approach contributes to applied research decisions about groups of individuals and specific people in which a test will be applied (Sánchez-Meca & López-Pina, 2008), also addressing questions regarding the average reliability estimates of a measure, the prediction intervals in the population, and to examine the between-study heterogeneity levels, among others (Badenes-Ribera et al., 2020). This is a characteristic of particular interest in psychological tests as the differences between subjects are legitimate and expected (Carvajal et al., 2011).

Therefore, the present research seeks to conduct a RGMA to evaluate the overall reliability estimates of the BAT on both its original and shortened versions, exploring reliability coefficients such as internal consistency and temporal stability vary across different cultural, occupational, and demographic contexts. With this, this study aims to contribute to the further development and enhancement of burnout measurement strategies, supporting more accurate and culturally sensitive approaches that can drive effective interventions and work as diagnosis tools.

2. Method

2.1. Search strategy

This meta-analysis was conducted under the Preferred Reporting Items Systematic Review and Meta Analysis (PRISMA) guidelines (Page et al., 2021), also following the recommendations by Sánchez-Meca et al. (2021) for reporting this kind of meta-analysis (See REGEMA Checklist in Supplementary Table S1). A systematic search was performed in the Web of Science, Scopus, ProQuest, PubMed, EBSCOhost and GoogleScholar databases. The final search was conducted on 13 May 2024. This study was registered in the PROSPERO database under registration number N°CRD42023442547.

Our search was based on the key terms "Burnout Assessment Tool," "BAT" and "Burnout," while searching for title, abstract and keywords when possible. The specific details of the advanced search syntax used for each database are provided in Supplementary Table S2 for replication purposes. Duplicates were detected based on their digital object identifier (DOI) and title, and finally through a manual check.

2.2. Eligibility criteria

Records obtained from the literature search were assessed using the following inclusion criteria: (a) the document addresses either the original or shortened versions of the BAT; (b) the document is a peer-reviewed scientific article; (c) the document is in either English or Spanish.

Subsequently, the remaining documents were thoroughly examined for the presence of at least one of the following exclusion criteria: (a) the document is a theoretical, review, or non-empirical study; (b) the document does not report reliability indices for inclusion in the meta-analysis; and (c) the study induces reliability indices. In this context, induction of reliability indices is defined as reporting these values based on data from prior research rather than calculating them in the relevant sample (Sánchez-Meca et al., 2021).

The initial systematic search was conducted by two independent reviewers (JM & FA) and a supervisor (NCL) acting as a third reviewer to resolve discrepancies or uncertainties regarding eligibility criteria.

2.3. Data extraction

After the study selection stage, all documents eligible for meta-analysis were examined for data extraction. Data extraction was performed by two independent reviewers (CVH & MNF) and a supervisor (HAA) acting as a third reviewer. Regarding reliability indices, we extracted data for the two most reported internal consistency indices: Cronbach's alpha (α) and McDonald's omega (ω). We also extracted data examining temporal stability of the measure, operationalised as test–retest correlations if available. Furthermore, we gathered descriptive data such as sample size, sample type, age, sex distribution (percentage of female participants), country, and which version of the BAT was used in each study.

2.4. Quality assessment

The assessment of the methodological rigour and the possible presence of bias in each of the included studies was examined using the Quality Assessment Tool for Observational

Cohort and Cross-Sectional Studies of the National Institute of Health (NIH; National Institutes of Health, 2021). This tool comprises 14 areas of analysis, covering crucial aspects such as formulation of the research question, composition of the study population, recruitment procedures, power of the study, methods for measuring exposures and outcomes, attrition rate and statistical approaches used, among others. Each assessment area is scored as "yes" "no," "cannot be determined," "not applicable," or "not reported."

We adopted a scoring system that assigns a score of 1 to "yes" responses, 0.5 to partial statements, and 0 to "no" or unclear responses. The description of each criterion is detailed in Supplementary Table S3. Studies that obtained 7 or more points were classified as high quality ("good"), while those that obtained between 5 and 6 points were considered acceptable ("fair") and those below 5 are considered low quality ("poor"), which would indicate potential bias. The quality assessment process was carried out by two independent researchers.

2.5. Data analysis

We performed multiple RGMAs in order to determine the pooled reliability indices of the BAT across studies. Though the construct measured by the BAT and its subscales remain constant, this measure has been validated with fairly different populations since its development, including employees from multiple areas, teachers, and students (e.g. Angelini et al., 2021; De Beer et al., 2020; Popescu et al., 2024; Romano et al., 2022). Thus, we assume that the true effect could differ based on more than just sampling error. Therefore, our analyses were based on a random effects model (REM), which allows for the possibility of genuine variation in effect sizes across studies (Borenstein et al., 2009; Dettori et al., 2022). We employed the restricted maximum likelihood (REML) estimator, which has been shown to be preferable when high levels of heterogeneity are to be expected (Tanriver-Ayder et al., 2021). In addition, we employed the Hartung-Knapp-Sidik-Jonkman method (Hartung, 1999; Knapp & Hartung, 2003; Sidik & Jonkman, 2002) for the estimation of more precise confidence intervals by adjusting the standard errors that better reflect the uncertainty around the true pooled effect (IntHout et al., 2014; Sánchez-Meca & Marín-Martínez, 2008).

Analyses were performed independently for each BAT version, its subscales, and reliability indices used. The three existing versions of the measure were examined: First, the original BAT, which comprises the 23-item Core symptoms (BAT-C; sometimes referred in the literature as BAT-23), and the 10-item Secondary symptoms (BAT-S), along with a total score based on its 33 items. Second, the BAT-12; and third, the BAT-4. Due to the nature of both Cronbach's alpha and McDonald's omega distribution, these indices were transformed based on the formula proposed by Bonett (2002, 2010) in order to stabilise their variances and normalise their distributions for analysis (Sánchez-Meca et al., 2013). After this, the indices were transformed back to alpha and omega coefficients respectively in order to make them easily interpretable. For temporal stability, classical meta-analysis based on correlation coefficients can be performed. As correlation coefficients also have a non-normal distribution, their values need to be transformed using Fisher's r-to-Z (Borenstein et al., 2009). Then, as

is the case for internal consistency indices, *Z*-values would be transformed back to correlation coefficients for interpretation.

To assess the level of variability between studies, we conducted heterogeneity tests using Cochran's *Q* test and the I^2 index. Cochran's *Q* test helps to determine the presence of heterogeneity testing the hypothesis that there are no statistically significant differences between the studies, while the I^2 index quantifies the percentage of variability in the estimates attributed to heterogeneity rather than sampling error (Higgins & Green, 2011). Established cutoff values for the I^2 index propose that values between 0% and 40% may not be significant, while values between 40% and 60% typically represent moderate levels of heterogeneity; values over 60% or 75% indicate substantial or considerable heterogeneity (Higgins et al., 2003; Huedo-Medina et al., 2006). These tests provide insights into the strength of evidence for heterogeneity and the degree to which it contributes to the overall variability in the results.

Heterogeneity sources were further examined first through influence diagnostics, which allows for the identification of potential outliers by detecting single studies that can potentially skew the overall results, either because of different methodologies, different population, or sampling variability. Influence diagnostics determine, through multiple tests, if the pooled results change significantly when a single study is removed, indicating that the results of the meta-analysis might be significantly influenced by that study (Viechtbauer & Cheung, 2010). In addition, meta-regression analyses were conducted to test the potential moderating role of age, sex distribution (operationalised as the percentage of female subjects in the sample), and employability rates, in explaining the observed between-study heterogeneity.

All analyses were conducted with RStudio version 2023.06.0 with R version 4.4.0 using the *metafor* package (Viechtbauer, 2010).

3. Results

3.1. Study selection

A total of 360 records were obtained from the six databases. After removing duplicates, 119 records remained for title and abstract screening in order to determine if they fulfilled the inclusion criteria. Thirty-eight studies were discarded at this stage. The remaining 81 studies went through a full-text examination in order to determine if any exclusion criteria were met. After the study selection process, 56 studies were deemed eligible for inclusion in our analysis. In the end, 42 studies provided data for the original BAT (BAT-C/BAT-S), 19 for the BAT-12, and 2 for the BAT-4. Figure 1 presents the corresponding flow diagram for the study selection process.

3.2. Quality assessment

The methodological quality of the included studies varied. Using the guidance provided by the National Institutes of Health (2021), a total of 56 studies were analysed. A total of 17 studies received a "good" quality rating, indicating a low risk of bias. In addition, 37 of the studies received a "fair" quality rating, indicating some risk of bias, and 2 studies received a "poor" quality rating, indicating potential risk of bias. However, these

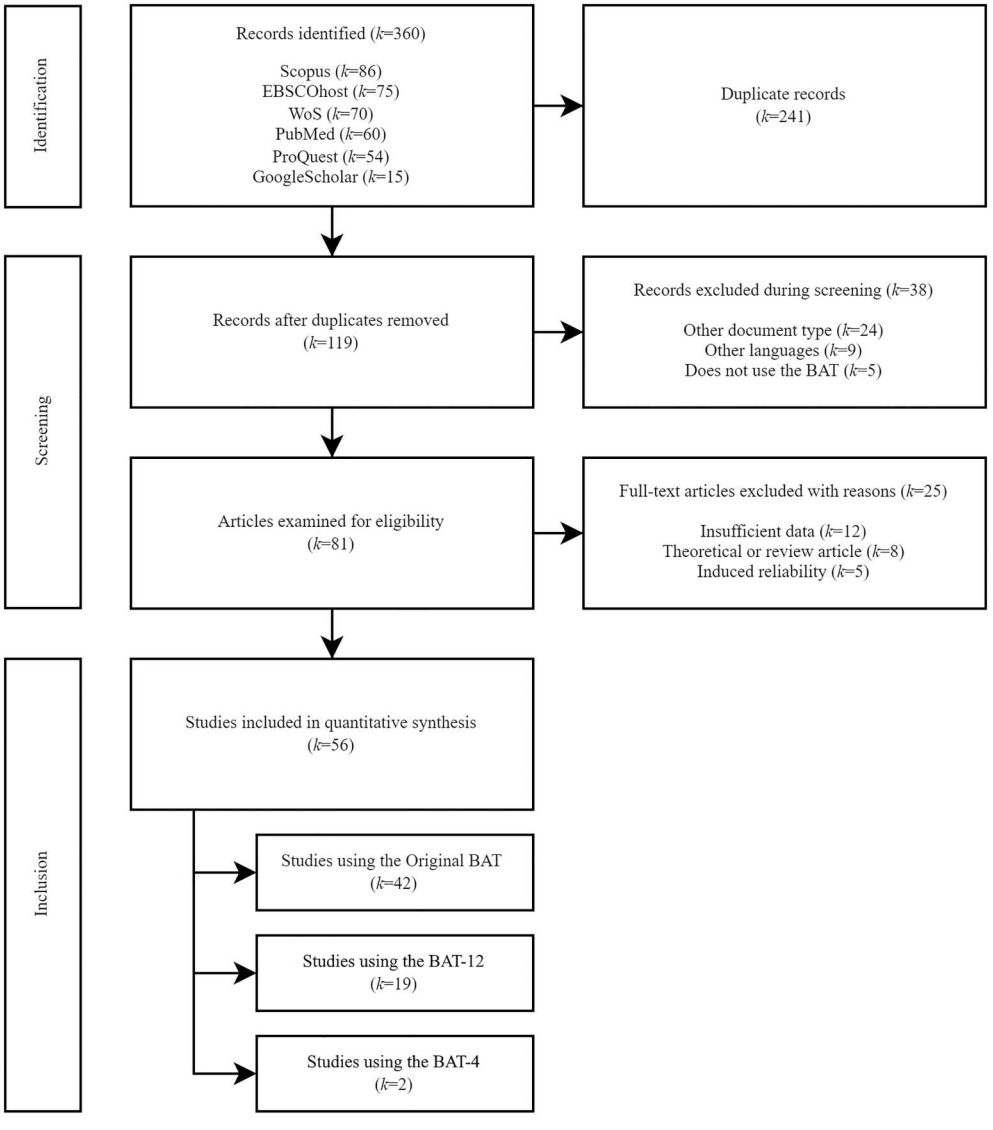

Figure 1. Flow diagram for the study selection process.

studies were not excluded aiming to provide a comprehensive perspective of the available evidence (Page et al., 2021).

It should be noted that a significant number of the included studies are cross-sectional. Cross-sectional designs have inherent limitations when evaluating changes over time, since they capture information at a single moment. This might explain the minimum scores on questions that refer to typical characteristics of longitudinal studies: sufficient time to observe associations (Q7), were exposures assessed more than once over time? (P10) and evaluation on loss of participants during follow-up (P13). Supplementary Table S4 provides the results for the 14 quality assessment areas and the overall quality rating for each specific study indicating potential risk of bias.

3.3. Sample characteristics

From the 56 included studies, a total of 61 independent samples were examined. The total sample size was 224,771 (M_{age} = 40.628; SD_{age} = 11.011; Female = 61.01%), comprising mainly individuals identified as employees and healthcare workers. For studies reporting the original BAT, sample size was 52,849 (M_{age} = 40.084; SD_{age} = 11.168; Female = 59.35%), while for studies reporting the BAT-12, the total sample size was 163,355 (M_{age} = 41.610; SD_{age} = 10.549; Female = 66.68%). Finally, for the BAT-4, the total sample size was 8,567 (M_{age} = 42.157 SD_{age} = 10.900; Female = 56.91%). The selected studies were carried out in a total of 30 countries, with the exception of Kannenberg (2022) who collected data from diving instructors around the globe, in an unspecified number of countries. A summary of all studies included for analyses based on the original and shortened versions is presented in Table 1.

3.4. Reliability generalisation meta-analysis (RGMA)

Based on the extracted data, a total of 28 separate RGMAs were conducted: 16 for the original BAT, 10 for the BAT-12, and 1 for the BAT-4. In terms of reliability indices, 15 meta-analyses were based on Cronbach's alpha, and 13 on McDonald's omega coefficients. While we initially expected to also perform meta-analysis for the BAT's temporal stability, in the end this had to be discarded as only one study, by Sakakibara et al. (2020), empirically examined for this reliability index ($r_{BAT\text{-}C}$ = .640; $r_{BAT\text{-}S}$ = .710; Interval = 60 days).

Results for all RGMAs and the corresponding heterogeneity tests are presented in Table 2. Forest, funnel, and influence diagnostics plots for the original BAT's exhaustion subscale are presented as example in Figures 2 and 3 for results based on Cronbach's alpha and McDonald's omega, respectively; this subscale had the largest number of included studies. Plots for the rest of the analyses are available in an OSF repository (see Data Availability section).

Results show that the pooled Cronbach's alpha and McDonald's omega values ranged from good to very good for both the original (α = .798–.948; ω = .754–.940) and both shortened versions (α = .763–.907; ω = .750–.909), demonstrating that all subscales of the BAT, along with its total scores, possess high internal consistency across different studies, samples, and versions. Notably, the 23-item BAT-C, in which both short versions are based, showed higher internal consistency all-round when compared to them.

3.5. Influence diagnostics

On the other hand, our results also revealed high levels of heterogeneity overall between the internal consistency indices reported across the literature (ps_Q = .004; $I^2 \geq 77.59\%$). Influence diagnostics were conducted for each meta-analysis. Results show that internal consistency coefficients were considered particularly influential in 10 cases. Additional meta-analyses were performed excluding these particular studies aiming to explore its impact on the results and overall heterogeneity levels. A summary of the changes in α, ω and I^2 coefficients is provided in Table 3.

Table 1. Summary of included studies.

Author	Sample (as described)	*N*	M_{age}	SD_{age}	Fem%	Country	Version used
Schaufeli et al. (2020)	Employees	1500	41.000	12.000	46.00%	Belgium	BAT
De Beer et al. (2020)	Sample 1: Employees	1500	41.260	13.360	45.90%	Netherlands	BAT
	Sample 2: Employees	1073	41.790	13.140	48.50%	Germany	BAT
	Sample 3: Employees	1059	42.980	13.310	49.90%	Austria	BAT
	Sample 4: Employees	431	42.100	12.300	53.60%	Ireland	BAT
	Sample 5: Employees	2299	43.500	11.340	50.30%	Finland	BAT
Kolachev and Novikov (2020)	Library employees	620	47.390	11.910	95.00%	Russia	BAT
Sakakibara et al. (2020)	Employed monitors	982	39.800	11.300	49.00%	Japan	BAT
Angelini et al. (2021)	Teachers	846	47.520	9.940	91.10%	Italy	BAT
Consiglio et al. (2021)	Employees	738	41.570	10.510	52.90%	Italy	BAT
Haar (2021a)	Managers	840	–	–	–	New Zealand	BAT
Haar (2021b)	Employees	1022	39.300	13.900	55.00%	New Zealand	BAT
Oprea et al. (2021)	Employees	648	33.560	9.960	78.00%	Romania	BAT
Pereira, Gonçalves, et al. (2021)	Teachers	302	46.750	11.020	55.00%	Brazil	BAT
Pereira, Fehér, et al. (2021)	Professionals	695	37.710	12.640	58.70%	Portugal	BAT
Vinueza-Solórzano et al. (2021)	Postgraduate programme workers	2237	34.000	6.800	65.60%	Ecuador	BAT; BAT-12
Aboutalebi et al. (2022)	Teachers and administrative staff	102	40.980	12.350	61.80%	Australia	BAT
Borrelli et al. (2022)	Healthcare workers	697	36.470	11.200	68.40%	Italy	BAT
Buono et al. (2022)	Employees and teachers	780	39.977	11.916	56.90%	Italy	BAT
De Beer, Schaufeli, and Bakker (2022)	Employees	660	38.000	10.600	58.00%	South Africa	BAT-12
De Beer, Schaufeli, and De Witte (2022)	Employees	1048	40.800	11.200	54.50%	South Africa	BAT
Gómez et al. (2022)	Teachers	265	41.610	9.680	74.70%	Spain	BAT-12
Haar and O'Kane (2022)	Employees	955	39.600	13.900	54.50%	New Zealand	BAT
Hadžibajramović, Hansson, et al. (2022)	Midwives	1664	48.000	–	99.60%	Sweden	BAT
Hadžibajramović, Schaufeli, et al. (2022)	Workers	1500	41.255	13.358	54.10%	Netherlands	BAT
Innstrand (2022)	Healthcare workers	268	–	–	76.50%	Norway	BAT-12
Kannenberg (2022)	Recreational diving instructors	1188	–	–	27.50%	Multiple	BAT
Lau et al. (2022)	Teachers	366	38.300	9.720	54.40%	China	BAT
Mazzetti et al. (2022)	Employees	2277	–	–	57.40%	Italy	BAT-12
Romano et al. (2022)	Middle school students	745	11.840	1.210	47.80%	Italy	BAT
Sinval et al. (2022)	Workers	2217	36.900	11.100	74.80%	Brazil	BAT
	Workers	886	38.900	11.400	72.50%	Portugal	BAT-12
Sørengaard and Langvik (2022)	Police employees	206	42.480	10.710	48.00%	Norway	BAT
van Den Boogert et al. (2022)	Mental health employees	116	44.700	12.200	71.60%	Netherlands	BAT
Androulakis et al. (2023)	Workers	356	–	–	69.90%	Greece	BAT; BAT-12
Barnard et al. (2023)	Emergency nurses	204	–	–	71.60%	South Africa	BAT-12
Basińska et al. (2023)	Nursing staff	252	44.700	10.100	–	Poland	BAT
Bianchi et al. (2023)	Civil servants	1612	44.000	9.000	60.00%	Brazil	BAT-12
Brulin et al. (2023)	Nurses and physicians	5615	–	–	79.10%	Sweden	BAT-12
Consiglio et al. (2023)	Remote workers	225	–	–	60.00%	Italy	BAT
De Amorim et al. (2023)	Professionals	3428	36.500	9.500	86.00%	Brazil	BAT-12
De Beer et al. (2023)	Workers	493	45.550	11.540	49.50%	Norway	BAT; BAT-12
De Freitas et al. (2023)	Health care workers	181	37.000	–	71.80%	Brazil	BAT
du Bois et al. (2023)	Employees	1006	46.018	12.305	51.00%	Belgium	BAT-12
Felicia et al. (2023)	Health workers	306	–	–	65.40%	Indonesia	BAT

(*Continued*)

Table 1. Continued.

Author	Sample (as described)	*N*	M_{age}	SD_{age}	Fem%	Country	Version used
Haar (2023)	Sample 1: Employees	709	39.800	14.300	59.00%	New Zealand	BAT
	Sample 2: Managers	313	38.200	12.800	43.00%	New Zealand	BAT
Izdebski et al. (2023)	Healthcare workers	2196	–	–	81.20%	Poland	BAT-12
Lazauskaitė-Zabielskė et al. (2023)	Working population	408	35.940	12.560	68.60%	Lithuania	BAT
Lee and Joo (2023)	Nurses	174	–	–	90.20%	Korea	BAT
Lyon and Galbraith (2023)	Mental health practitioners	144	–	–	90.00%	United States	BAT
Månsson et al. (2023)	General practicioners	6699	–	–	58.20%	Sweden	BAT
Marrinhas et al. (2023)	Teachers and students	333	50.000	10.000	56.50%	Portugal	BAT
Mazur et al. (2023)	Nurses	1187	50.400	10.100	100.00%	Poland	BAT-12
Palvimo et al. (2023)	Nurses	2115	45.600	10.900	92.50%	Finland	BAT-4
Popescu et al. (2024)	Bachelor students	399	20.760	4.620	60.70%	Romania	BAT
Ranabhat et al. (2023)	Pharmacy professionals	385	29.400	5.600	38.20%	Nepal	BAT
Redelinghuys and Morgan (2023)	Sample 1: Employees	794	–	–	–	Australia	BAT
	Sample 2: Employees	199	–	–	–	Netherlands	BAT
	Sample 3: Employees	197	–	–	–	South Africa	BAT
	Sample 4: Employees	198	–	–	–	United States	BAT
Schaufeli et al. (2023)	Sample 1: Healthy employees	1370	41.760	13.400	46.00%	Netherlands	BAT; BAT-12
	Sample 2: Healthy employees	1403	41.280	11.580	45.00%	Belgium	BAT; BAT-12
	Sample 3: Healthy employees	1372	46.090	11.020	57.60%	Finland	BAT; BAT-12
Tomas et al. (2023)	Sample 4: Employees	966	41.530	9.660	57.20%	Croatia	BAT-12
De Beer, Hakanen et al. (2024)	Sample 1: Employees	1059	42.980	13.310	49.90%	Austria	BAT
	Sample 2: Employees	1500	40.900	11.600	44.10%	Belgium	BAT
	Sample 3: Employees	1073	41.790	13.140	48.50%	Germany	BAT
	Sample 4: Employees	1567	45.800	10.980	59.50%	Finland	BAT
Hadžibajramović et al. (2024)	Sample 1: Workers	1600	41.500	–	50.00%	Multiple	BAT-4
	Sample 2: Workers	1054	42.000	–	51.20%	Austria	BAT-4
	Sample 3: Workers	1073	43.000	–	49.90%	Germany	BAT-4
	Sample 4: Workers	733	42.000	–	50.60%	Finland	BAT-4
	Sample 5: Workers	964	42.000	–	54.30%	Czech Republic	BAT-4
	Sample 6: Workers	1028	39.000	–	49.90%	Japan	BAT-4

Notes: *n* = Sample size; M_{age} = Mean age of the samples; SD_{age} = Standard deviation of the mean age of the sample; Fem % = Percentage of female participants in the sample; BAT = Burnout Assessment Tool; BAT-12 = Burnout Assessment Tool (Short version); BAT-4 = Burnout Assessment Tool (Ultra Short version); – = Not reported.

Results show that, in general, removing influential studies reduces heterogeneity levels ($\bigtriangleup I^2 \approx -0.32\%$ to -91.91%) and generates small variations in pooled variability. Reduction of heterogeneity is particularly strong for the total BAT score ($\bigtriangleup I^2 = -91.91\%$). However, this can be easily attributable to the reduced number of studies left included ($k_{adj} = 2$). Variations in pooled reliability indices also tended to be minimal ($\bigtriangleup \alpha/\omega \leq 0.019$).

3.6. Meta-regression

To further examine potential heterogeneity sources, we conducted multiple meta-regression analyses based on a mixed-effects model (MEM) for the main scales of the

Table 2. Reliability generalisation meta-analysis results.

			REM							Heterogeneity tests		
	Version	Subscale$_{(items)}$	*k*	IC	IC_{BT}	SE	*p*	CI (95%)		*Q*	*p*	I^2
Cronbach's alpha (*α*)	BAT	$EX_{(8)}$	40	.896	2.265	.048	.000	.886	– .906	1571.641	.000	97.47%
		$MD_{(5)}$	38	.856	1.936	.068	.000	.835	– .874	2594.119	.000	98.37%
		$CI_{(5)}$	37	.900	2.306	.052	.000	.889	– .910	1138.398	.000	97.16%
		$EI_{(5)}$	37	.880	2.121	.070	.000	.862	– .896	2265.012	.000	98.45%
		BAT-$C_{(23)}$	36	.930	2.657	.077	.000	.918	– .940	2445.919	.000	98.78%
		$PD_{(5)}$	10	.830	1.770	.065	.000	.803	– .853	109.270	.000	91.38%
		$PC_{(5)}$	12	.798	1.597	.069	.000	.764	– .826	197.493	.000	92.95%
		BAT-$S_{(10)}$	8	.872	2.053	.072	.000	.848	– .892	73.574	.000	90.46%
		$BAT_{(33)}$	3	.948	2.960	.138	.002	.906	– .971	31.613	.000	91.91%
	BAT-12	$EX_{(3)}$	15	.867	2.018	.075	.000	.844	– .887	419.903	.000	98.55%
		$MD_{(3)}$	14	.784	1.530	.111	.000	.725	– .830	1125.432	.000	99.35%
		$CI_{(3)}$	14	.846	1.872	.067	.000	.822	– .867	777.837	.000	98.15%
		$EI_{(3)}$	14	.848	1.883	.098	.000	.812	– .877	818.914	.000	99.15%
		BAT-$12_{(12)}$	15	.907	2.376	.096	.000	.886	– .924	3772.820	.000	99.32%
	BAT-4	BAT-$4_{(4)}$	7	.763	1.438	.063	.000	.723	– .797	82.299	.000	91.99%
McDonald's omega (*ω*)	BAT	$EX_{(8)}$	20	.906	2.360	.057	.000	.894	– .916	379.699	.000	96.15%
		$MD_{(5)}$	20	.843	1.848	.114	.000	.800	– .876	2093.546	.000	98.99%
		$CI_{(5)}$	20	.900	2.299	.090	.000	.879	– .917	1130.524	.000	98.34%
		$EI_{(5)}$	20	.871	2.045	.096	.000	.842	– .894	1134.398	.000	98.56%
		BAT-$C_{(23)}$	9	.940	2.820	.122	.000	.921	– .955	510.085	.000	98.94%
		$PD_{(5)}$	4	.809	1.655	.118	.001	.721	– .869	28.203	.000	91.60%
		$PC_{(5)}$	4	.754	1.402	.072	.000	.691	– .804	13.548	.004	77.59%
		BAT-$S_{(10)}$	1	–	–	–	–	–		–	–	–
		$BAT_{(33)}$	1	–	–	–	–	–		–	–	–
	BAT-12	$EX_{(3)}$	12	.843	1.851	.084	.000	.811	– .869	198.209	.000	97.07%
		$MD_{(3)}$	12	.750	1.386	.146	.000	.655	– .819	1174.302	.000	99.08%
		$CI_{(3)}$	12	.851	1.903	.076	.000	.824	– .874	218.223	.000	96.54%
		$EI_{(3)}$	12	.845	1.867	.114	.000	.802	– .880	470.931	.000	98.46%
		BAT-$12_{(12)}$	5	.909	2.395	.186	.000	.847	– .946	503.704	.000	98.91%
	BAT-4	BAT-$4_{(4)}$	6	.770	1.472	.069	.000	.726	– .808	61.042	.000	91.14%

Notes: REM = Random-effects model; *k* = number of independent samples included; IC = Internal consistency coefficient; *α* = Cronbach's alpha; *ω* = McDonald's omega; BT = Bonett transformed value; SE = Standard error; CI (95%) = Confidence interval; *Q* = Cochran's statistic; I^2 = Heterogeneity index; EX = Exhaustion; MD = Mental distance; CIm = Cognitive impairment; EI = Emotional impairment; BAT-C = Core symptoms; PD = Psychological distress; PC = Psychosomatic complaints; BAT-S = Secondary symptoms; BAT = Burnout Assessment Tool; BAT-12 = Burnout Assessment Tool (Short version); BAT-4 = Burnout Assessment Tool (Ultra Short version); – = Insufficient data for calculations, or the number of parameters to be estimated is larger than the number of observations. *p* values are statistically significant when < .05.

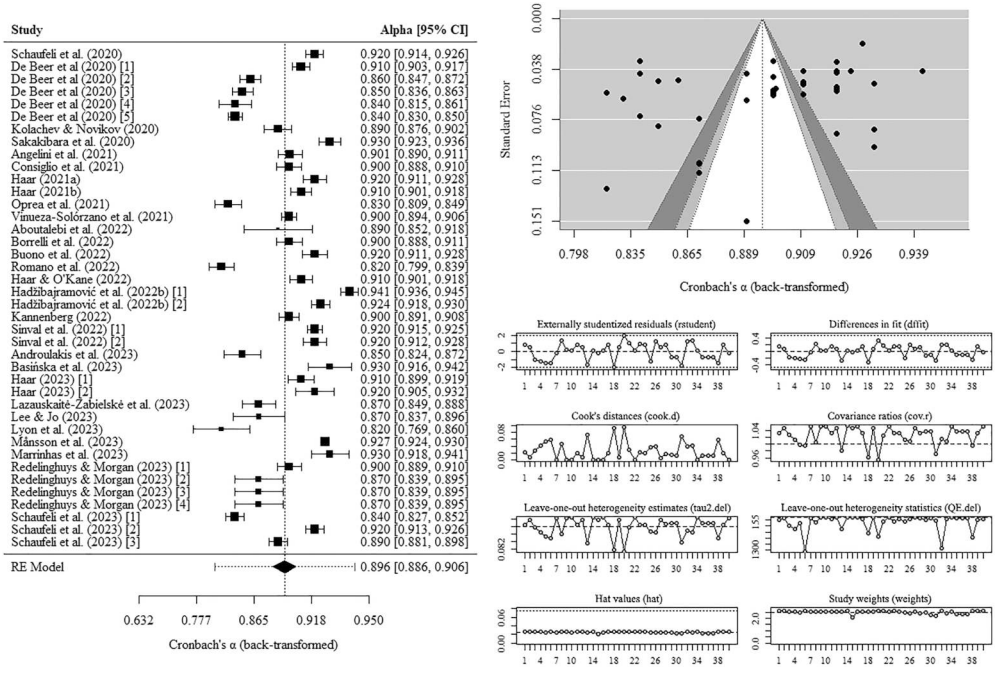

Figure 2. Forest, funnel, and influence diagnostics plots for the original BAT's exhaustion subscale meta-analysis based on Cronbach's alpha.

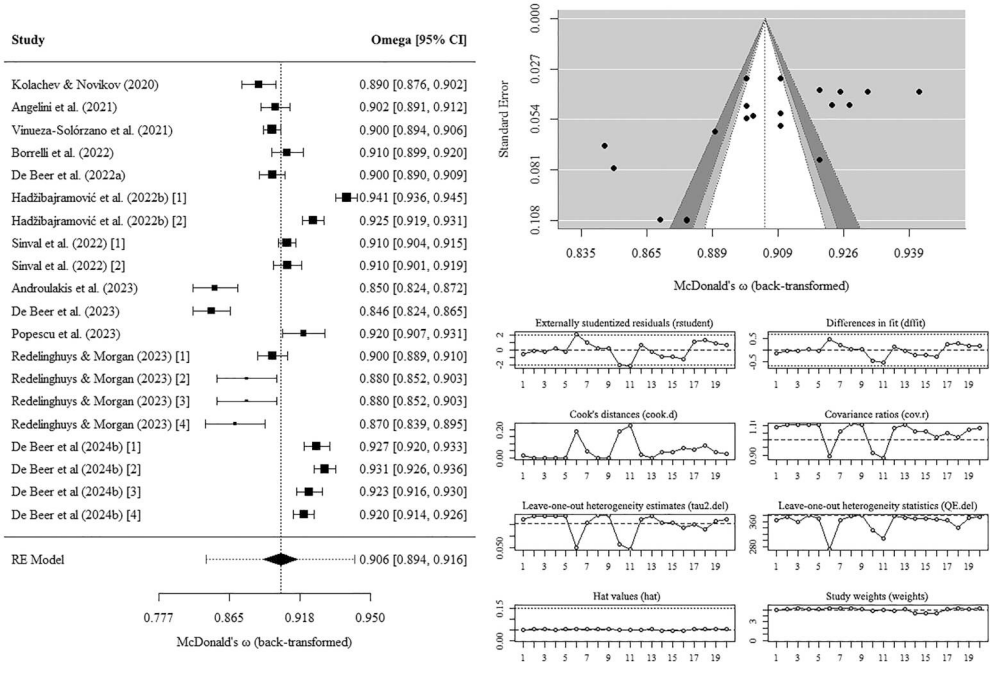

Figure 3. Forest, funnel, and influence diagnostics plots for the original BAT's exhaustion subscale meta-analysis based on McDonald's omega.

Table 3. Examination of influential studies.

IC	BAT version	Subscale	Influential study sample	Outlier type	k_{adj}	IC_{adj}	△IC	I^2_{adj}	△I^2
Cronbach's alpha (α)	Original BAT	EX	–	–	–	–	–	–	–
		MD	–	–	–	–	–	–	–
		CI	Oprea et al. (2021)	Low (.770)	36	.903	.002	96.54%	−0.62%
		EI	–	–	–	–	–	–	–
		BAT-C	Ranabhat et al. (2023)	Low (.740)	35	.933	.003	98.46%	−0.32%
		PD	Sakakibara et al. (2020)	High (.890)	9	.821	−.009	77.92%	−13.46%
		PC	–	–	–	–	–	–	–
		BAT-S	–	–	–	–	–	–	–
		BAT	Buono et al. (2022)	High (.960)	2	.940	−.008	0.00%	−91.91%
	BAT-12	EX	–	–	–	–	–	–	–
		MD	–	–	–	–	–	–	–
		CI	–	–	–	–	–	–	–
		EI	–	–	–	–	–	–	–
		BAT-12	–	–	–	–	–	–	–
	BAT-4	BAT-4	Hadžibajramović et al. (2024) (Sample 1)	High (.819)	6	.751	−.012	84.57%	−7.42%
McDonald's omega (ω)	Original	EX	–	–	–	–	–	–	–
		MD	–	–	–	–	–	–	–
		CI	–	–	–	–	–	–	–
		EI	–	–	–	–	–	–	–
		BAT-C	Hadžibajramović, Schaufeli, et al. (2022)	High (.973)	8	.934	−.006	94.09%	−4.85%
		PD	Androulakis et al. (2023)	Low (.730)	3	.828	.019	51.90%	−39.70%
		PC	Androulakis et al. (2023)	Low (.710)	3	.765	.011	74.57%	−3.02%
		BAT-S	–	–	–	–	–	–	–
		BAT	–	–	–	–	–	–	–
	BAT-12	EX	–	–	–	–	–	–	–
		MD	–	–	–	–	–	–	–
		CI	–	–	–	–	–	–	–
		EI	–	–	–	–	–	–	–
		BAT-12	Hadžibajramović, Schaufeli, et al. (2022)	High (.952)	4	.893	−.016	96.48%	−2.43%
	BAT-4	BAT-4	Hadžibajramović et al. (2024) (Sample 1)	High (.821)	5	.758	−.012	82.68%	−8.46%

Notes: k_{adj} = Adjusted number of independent samples included; IC_{adj} = Adjusted internal consistency coefficient; I^2_{adj} = Adjusted heterogeneity index; △ = Difference after excluding influential studies; EX = Exhaustion; MD = Mental distance; CI = Cognitive impairment; EI = Emotional impairment; BAT-C = Core symptoms; PD = Psychological distress; PC = Psychosomatic complaints; BAT-S = Secondary symptoms; BAT = Burnout Assessment Tool; BAT-12 = Burnout Assessment Tool (Short version); BAT-4 = Burnout Assessment Tool (Ultra Short version); – = No influential studies detected.

BAT. Once again, both its original and shortened versions were examined. The effect of three potential moderators, based on the descriptives provided in most studies, was explored: Mean age of the subjects, percentage of females, and employment rates (by country). Detailed results are provided in Supplementary Tables S5 and S6, respectively.

Our findings revealed that, for the original BAT, moderating effects were significant for the MD and CI subscales ($p_{QM} \leq .032$; $R^2 = .221$ to .464). For the BAT-12, moderating effects were observed for the MD and total score ($p_{QM} \leq .033$; $R^2 = .628$ to .676). No statistically significant moderating effects were observed for the BAT-4 ($p_{QM} = .714$; $R^2 = .000$).

The mean age of the sample showed a significant moderating effect for the original BAT's MD subscale, while the distribution of female participants in the sample showed was a moderator for both MD (Estimate = −1.467; SE = .540; p = .012) and CI (Estimate = −1.366; SE = .489; p = .010). On the other hand, for the BAT-12, only the employment rate demonstrated a moderating effect on both the MD (Estimate = 2.729; SE = .922; p = .025) and total score (Estimate = 2.316; SE = .704; p = .017).

4. Discussion

This study aimed to synthesise the current evidence on the reliability of the Burnout Assessment Tool (BAT; Schaufeli et al., 2020) and its shortened versions: The 12-item Short BAT (BAT-12; De Beer, Schaufeli, & Bakker, 2022), and the 4-item Ultra Short BAT (BAT-4; Hadžibajramović, Schaufeli, et al., 2022). This reliability generalisation meta-analysis was based on the two most commonly reported reliability indices: Cronbach's alpha and McDonald's omega. To date, this is the first study aimed at determining the overall internal consistency on this measure across different samples and populations, and to examine both internal consistency coefficients through a meta-analytical approach. We analysed a total of 56 independent articles examining the reliability of the BAT up until May 2024, in the WoS, Scopus, ProQuest, PubMed, EBSCOhost and GoogleScholar databases.

Our findings confirmed the BAT as a highly reliable tool for the assessment of the burnout syndrome, demonstrating strong internal consistency indices for both its original and shortened versions. The pooled Cronbach's alpha and McDonald's omega values for the BAT ranged from good to very good across all versions, reaffirming the measure's high reliability. Specifically, the original BAT showed slightly higher internal consistency compared to the shortened versions, which we consider can be attributed to its more comprehensive item set. That said, while the reliability estimates for the BAT and its subscales are predominantly above the common 0.70 cut-off, the lower bounds of some confidence intervals fall slightly below 0.80, a threshold that has been proposed as a more stringent criterion for acceptable reliability (Lance et al., 2006; Nunnally & Bernstein, 1994). This finding suggests that while the BAT demonstrates acceptable internal consistency overall, the results warrant cautious interpretation in contexts requiring higher precision.

The high between-study heterogeneity levels observed indicate that there is a significant variability in reliability estimates. This variability is likely influenced by differences in sample characteristics, study designs, and contextual or cultural factors. Influence diagnostics revealed that certain studies had a considerable impact on the overall results but removing these influential studies generally reduced heterogeneity levels and resulted in minimal changes to pooled reliability indices. When exploring this variability, potentially influential studies were detected. However, the exclusion of these studies did not reduce heterogeneity significantly. Further exploration of heterogeneity sources through meta-regression analyses based on age, sex distribution and employability rates were, in general, informative but not entirely conclusive. For the original BAT, mean age, and female percentage in the sample significantly moderated reliability indices for the MD and CI subscales. Specifically, the percentage of female participants negatively impacted reliability estimates, suggesting potential gender differences in the consistency

of burnout manifestation or reporting between predominantly female populations. For the BAT-12, employment rates by country emerged as a significant moderator for both the MD subscale and the total score, indicating that socio-economic factors might influence the consistency of burnout assessments. This goes in line with findings on the negative relationship between higher employability and lower levels of job insecurity in the population (De Witte, 1999). Thus, this moderating effect suggests that more stable organisational environments and work conditions lead to a more consistent way to experience work-related stressors (Hakanen & Bakker, 2017).

Regarding sex as a potential moderating factor, the literature has previously reported that males tend to exhibit higher levels of mental distance, while females report higher levels of exhaustion (Purvanova & Muros, 2010; Shankland et al., 2019). This suggests that the manifestation of the different dimensions of burnout could vary significantly between males and females. In addition, it has been also reported that workplace gender composition can have an impact in overall work environment, organisational culture and psychological distress levels (Elwér et al., 2014). On the other hand, the potential effect of age could be related to perceptions of job demands, which are closely related to stress and burnout. In this way, studies have revealed that young workers tend to report higher burnout levels due to career uncertainty, inexperience, and a growing need to demonstrate their abilities and job competences (Schaufeli et al., 2023). Moreover, while senior workers might have higher levels of stress accumulation across their work life, they could also benefit from experience and resilience, developing more effective coping strategies to deal with job stressors (Schaufeli et al., 2006; Siu, 2003). Previous meta-analytic reviews have provided evidence for age to be considered as a relevant factor and have also shown evidence that more experienced workers report less burnout attributable to better-coping strategies than younger workers (Brewer & Shapard, 2004; Xie et al., 2021). It should be noted that age has also shown to be a relevant moderator in the relationship between burnout and negative mental health outcomes such as stress or depression (Meier & Kim, 2021; Salvagioni et al., 2017), which reinforces its role in this dynamic.

Considering the practical implications of the study, one of high relevance is that the BAT can be considered a valid and consistent measure for its use in organisations to make visible the deterioration of workers' health, also determining which particular dimension of this syndrome is the most highly present in employees. Burnout is considered a condition that is caused by the work context, therefore, having instruments capable of identifying the presence of the syndrome is of crucial importance for organisations from all occupational sectors (ICD-11; WHO, 2019, 2022). In this way, human resource managers or practitioners can use the BAT as a tool for assessing burnout symptoms, allowing to establish intervention strategies at different levels (i.e. individual, team an organisations) aimed at early detection, prevention of addressing its consequences an impact in job performance and in the workplace (Bes et al., 2023). In addition, our results show that the original version of the instrument is the one with the highest internal consistency overall. Nevertheless, both the BAT-12 and BAT-4 versions also demonstrate good reliability indices. These findings can allow organisations, researchers and practitioners to make strategic decisions when seeking to integrate shorter questionnaires to consider different aspects of mental health in the workplace.

Studies focused on examining the internal consistency of other burnout measures have also been conducted in the past. In particular, three previous studies have examined the overall reliability (as measured by the Cronbach's alpha coefficient) of the MBI through meta-analytical approaches. The first being a study by Wheeler et al. (2011), which included 84 studies reporting sample-specific coefficients, pointing out that the pooled alpha coefficients of the MBI's subscales typically ranged between .70 and .80. Their results showed that the emotional exhaustion scale was the strongest in terms of internal consistency, while the depersonalisation and personal accomplishment subscales demonstrated high variability. Their findings also indicate the presence of an important language bias, with studies in English being more consistent overall. In addition, previous studies have shown a lack of factorial invariance of the MBI, as the structural validity of its subscales tends to vary significantly across different occupational settings (Vanheule et al., 2006).

The second study, by Aguayo et al. (2011), analysed 45 empirical articles and found that the overall internal consistency of the MBI ranged roughly between .71 and .88. Their results were in line with those of Wheeler et al. (2011), with emotional exhaustion being the strongest in terms of internal consistency and followed by personal accomplishment and depersonalisation. In Aguayo et al. (2011)'s study, moderators such as age, country and version of the measure were crucial in explaining the between-study heterogeneity, specifically, for emotional exhaustion and personal accomplishment. A third and more recent study meta-analyzing the MBI-General Survey version by De Beer et al. (2024a) conducted a comprehensive examination of its measurement properties including internal consistency along with a synthesis of the existing research on its structural and criterion validity. Their findings highlighted persisting challenges in dimensionality, non-stable internal consistency indices on some subscales, and limited applicability in cross-cultural contexts.

In contrast, the findings of the present study demonstrate that the BAT possesses higher internal consistency indices when compared to the MBI, ranging between .798 and .948 for the original BAT, and between .763 and .907 for its shorter versions. This represents a significant improvement in the reliability of burnout measurement, suggesting that the BAT provides a more consistent assessment even across diverse samples. In addition, our study not only examined Cronbach's alpha, but also included studies examining internal consistency through McDonald's omega, which is currently more advisable due to its suitability for assessing the reliability of multidimensional scales (Dunn et al., 2014). The omega coefficients for the BAT ranged from .754 to .940 for the original BAT and from .750 to .909 for the shortened versions, further supporting the high internal consistency indices of this measure when compared to the MBI. Overall, these findings provide robust evidence for the BAT to be considered as a reliable tool for assessing burnout across different samples.

4.1. Limitations and strengths

In terms of limitations, it can be noted that most of the studies that were included in the final analysis belonged to adaptations of the instrument in European countries. This raises the question regarding if same reliability indices are to be expected in countries with less favourable realities, in which adverse work practices are more frequent. In

this line, our analysis also included studies focused on Latin America, North America, Asia, Africa and Oceania, which have also reported good internal consistency indices for the BAT. This suggests that, while the overall evidence on this measure's validity in non-European contexts can still be considered relatively limited, results have been similar to those obtained in European population, thus demonstrating the versatility of the BAT and its shortened versions regardless of the cultural setting or language.

A second limitation comes from the existing literature on the measure, and it's related to the current inability to calculate pooled estimates for examining the overall temporal stability of the BAT due to it being reported by only one study (Sakakibara et al., 2020). This represents a significant gap that future studies should address, considering the need for a tool that can accurately and reliably assess the occurrence and development of burnout symptoms across time. This also represents a limitation for other measures such as the MBI, as its overall temporal stability also remains unexplored (Aguayo et al., 2011; Wheeler et al., 2011). Moreover, the role of potential moderators that significantly impact the reliability estimates of the measure is still unclear. While age, sex distribution and employment rates all showed an effect in specific subscales of the BAT and the BAT-12, these results alone are unable to explain most of the heterogeneity levels observed. Given this circumstance, researchers are invited to exhaustively report characterisations of the sample that may be of relevance to understand this underlying variability in subsequent studies. We also intended to consider position as a potential moderator (i.e. employee, supervisor), but noticed that the number of studies aimed at supervisors was so low that the analyses couldn't be conducted as the number of parameters that were to be estimated was higher than the number of included study samples.

In the same line, there are challenges on establishing direct comparisons on the findings for the BAT with those of the MBI, given differences in the methodological approaches and contexts of prior studies. While earlier meta-analyses of the MBI, such as those conducted by Aguayo et al. (2011) and Wheeler et al. (2011), also employed reliability generalisation meta-analysis (RGMA), variations in the conceptual focus or sample characteristics, and specific estimation techniques create inherent constraints for direct comparisons. Additionally, the recently updated meta-analysis of the MBI-General Survey by De Beer, Van Der Vaart, et al. (2024) provides a more contemporary approach towards the MBI's psychometric properties but focusing on its multidimensionality and structural validity. Future studies on the assessment and measurement of burnout should consider using both the BAT and MBI within the same sample and context to provide more robust evidence for their comparative strengths and applications in terms of validity and reliability.

On the other hand, among the strengths of the study was the use of indexed journal databases for their scientific quality stands out, which guarantees the scientific rigour of the studies incorporated in our work. Another strength, which is related to the exhaustiveness of the study, is that, although it was considered as a selectivity criterion that the articles were available in both English and Spanish. As previously mentioned, our aim was not only to examine internal consistency through the commonly reported Cronbach's alpha, but also based on McDonald's omega which is currently more recommended for this purpose. This allows us to provide a comprehensive examination of the BAT's internal consistency. In addition, while the role of the moderators tested was relatively limited, it still provides highly relevant information for future studies, researchers, and practitioners on the potential impact that these particular variables

could have on the internal consistency estimates. Finally, it is worth highlighting the innovation of the study, which corresponds to the first reliability generalisation meta-analysis carried out on the BAT instrument, whose favourable result drives future adaptations of the instrument.

4.2. Future directions

The findings of the present study allow us to identify that the dimensions of the instrument are highly reliable and internally consistent in both its original and shortened versions. As a result, we propose that future studies should encourage the adaptation of the instrument in different countries and occupational settings to strengthen the existing evidence on this measures' validity and reliability. While we have observed that this doesn't seem to be a common problem for the BAT (with only 5 cases at the time of this review), it is still recommended for future studies to avoiding the induction of reliability indices as this practice produces bias, ultimately failing to provide accurate information on a measures' reliability for that particular sample (Sánchez-Meca et al., 2021).

Likewise, future research is invited to consider the adaptation of the instrument, especially in both the BAT-12 and BAT-4 versions. This choice can bring positive benefits for institutions, considering the lower number of invested resources derived from the application and tabulation of the instrument when compared to its original version. Finally, through this study it was possible to determine that the different adaptations of the BAT present similar reliability statistics as those of the original version. Which, together with the theoretical improvements that the measure was built on based on the limitations of the MBI, positions the BAT with the potential to be the new preferred instrument for the evaluation of burnout syndrome in organisational contexts.

Based on our results, and considering the samples and settings in which the BAT either in its original or shortened versions have been used, we propose the original version of the BAT provides a better overall assessment of burnout syndrome in clinical and occupational settings as it allows to accurately differentiate between the three Core symptoms of burnout using relatively longer and more reliable subscales. This version can be supplemented with other scales that assess secondary burnout symptoms (i.e. psychological and psychosomatic complaints, and depression), as recommended by Schaufeli et al. (2020). That way, a detailed picture emerges of the various components of burnout that can be used as a basis for designing treatment. The 12-item Short version can be best used in employee surveys in organisations in order to identify those with elevated scores that are at risk for developing burnout. Relatively little information is lost when the BAT-C is shortened from 23 to 12 items, at least as far as the total-score is concerned (Hadžibajramović, Schaufeli, et al., 2022). Finally, the BAT-4 can be best used in large-scale epidemiological or international surveys to estimate the prevalence of burnout in particular populations (Hadžibajramović et al., 2024).

5. Conclusion

In summary, this study represents a significant contribution to the current knowledge on the measurement of burnout, the BAT and its properties. Through a psychometric meta-analytical approach, we conclude that this measure possesses good reliability indices in all its subscales and its total score, and in both its original and shortened versions. With this,

we provide strong evidence on the robustness of the measure as a sound and internally consistent tool that can be used across multiple cultural and organisational settings and languages. Our findings also support the suitability of the BAT for the screening of burnout symptoms in different occupational sectors, and in work and non-work contexts, hence contributing towards a better understanding of the burnout phenomena.

Disclosure statement

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

Funding

This work was supported by Programa de Investigación Asociativa (PIA), Universidad de Talca [RU-158-2019]; National Research and Development Agency of Chile (ANID)'s National Doctorate Grant [21221538].

Data availability statement

Documentation for all steps on this study, along with raw exports from the six databases and the dataset used for analysis are publicly available on an Open Science Framework repository on https://osf.io/dusmx/?view_only=71c97c49edca486b9bad72cfa0365b14.

ORCID

César Villacura-Herrera http://orcid.org/0000-0001-6665-3774
Hedy Acosta-Antognoni http://orcid.org/0000-0002-8412-4128
Javiera Maldonado http://orcid.org/0009-0004-9972-6748
Fernanda Arriaza http://orcid.org/0009-0006-3528-392X
Natalia Cancino-Letelier http://orcid.org/0000-0002-0860-1287
Marcelo Nvo-Fernández http://orcid.org/0000-0002-0183-8708
Wilmar B. Schaufeli http://orcid.org/0000-0002-6070-7150

References

Aboutalebi, M., Gallagher, H. C., Wang, P., Kyndt, E., Lusher, D., Block, K., & McKenzie, V. (2022). School staff wellbeing: A network-based assessment of burnout. *Frontiers in Psychology*, *13*, Article 920715. https://doi.org/10.3389/fpsyg.2022.920715

Aguayo, R., Vargas, C., de la Fuente, E. I., & Lozano, L. M. (2011). A meta-analytic reliability generalization study of the Maslach Burnout Inventory. *International Journal of Clinical and Health Psychology*, *11*(2), 343–361.

Androulakis, G. S., Georgiou, D. A., Lainidi, O., Montgomery, A., & Schaufeli, W. B. (2023). The Greek Burnout Assessment Tool: Examining its adaptation and validity. *International Journal of Environmental Research and Public Health*, *20*(10), 5827. https://doi.org/10.3390/ijerph20105827

Angelini, G., Buonomo, I., Benevene, P., Consiglio, P., Romano, L., & Fiorilli, C. (2021). The Burnout Assessment Tool (BAT): A contribution to Italian validation with teachers'. *Sustainability*, *13*(16), 9065. https://doi.org/10.3390/su13169065

Badenes-Ribera, L., Rubio-Aparicio, M., & Sánchez-Meca, J. (2020). Meta-análisis de generalización de la fiabilidad. *Información Psicológica*, *119*, 17–32. https://doi.org/10.14635/IPSIC.2020.119.6

Barnard, N. B., Rothmann, S., De Beer, L. T., & Lubbe, W. (2023). Burnout of emergency nurses in a South African context: The role of job demands and resources, and capabilities. *Frontiers in Psychology*, *14*, 1–16. https://doi.org/10.3389/fpsyg.2023.1119063.

Basińska, B., Gruszczyńska, E., & Schaufeli, W. (2023). The Polish adaptation of the Burnout Assessment Tool (BAT-PL) by Schaufeli et al. *Psychiatria Polska*, *57*(1), 223–235. https://doi.org/10.12740/PP/OnlineFirst/141563

Bes, I., Shoman, Y., Al-Gobari, M., Rousson, V., & Canu, I. G. (2023). Organizational interventions and occupational burnout: A meta-analysis with focus on exhaustion. *International Archives of Occupational and Environmental Health*, *96*(9), 1211–1223. https://doi.org/10.1007/s00420-023-02009-z

Bianchi, R., Cavalcante, D. C., Queirós, C., Santos, B. D. M., Verkuilen, J., & Schonfeld, I. S. (2023). Validation of the Occupational Depression Inventory in Brazil: A study of 1612 civil servants. *Journal of Psychosomatic Research*, *167*, Article 111194. https://doi.org/10.1016/j.jpsychores.2023.111194

Bianchi, R., & Schonfeld, I. S. (2023). Examining the evidence base for burnout. *Bulletin of the World Health Organization*, *101*(11), 743–745. https://doi.org/10.2471/BLT.23.289996

Bianchi, R., Schonfeld, I. S., & Laurent, E. (2014). Is burnout a depressive disorder? A reexamination with special focus on atypical depression. *International Journal of Stress Management*, *21*(4), 307–324. https://doi.org/10.1037/a0037906

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

Bianchi, R., Swingler, G., & Schonfeld, I. S. (2024). The Maslach Burnout Inventory is not a measure of burnout. *Work*, *79*(3), 1525–1527. https://doi.org/10.3233/wor-240095

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics*, *27*(4), 335–340. https://doi.org/10.3102/10769986027004335

Bonett, D. G. (2010). Varying coefficient meta-analytic methods for alpha reliability. *Psychological Methods*, *15*(4), 368–385. https://doi.org/10.1037/a0020142

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Borrelli, I., Santoro, P. E., Fiorilli, C., Angelini, G., Buonomo, I., Benevene, P., Romano, L., Gualano, M. R., Amantea, C., & Moscato, U. (2022). A new tool to evaluate burnout: The Italian version of the BAT for Italian healthcare workers. *BMC Public Health*, *22*(1), 474. https://doi.org/10.1186/s12889-022-12881-y

Botella, J., & Sánchez-Meca, J. (2015). *Meta-análisis en Ciencias Sociales y de la Salud*. Síntesis.

Boudreau, R. A., Grieco, R. L., Cahoon, S. L., Robertson, R. C., & Wedel, R. J. (2006). The pandemic from within: Two surveys of physician burnout in Canada. *Canadian Journal of Community Mental Health*, *25*(2), 71–88. https://doi.org/10.7870/cjcmh-2006-0014

Brewer, E. W., & Shapard, L. (2004). Employee burnout: A meta-analysis of the relationship between age or years of experience. *Human Resource Development Review*, *3*(2), 102–123. https://doi.org/10.1177/1534484304263335

Brulin, E., Lidwall, U., Seing, I., Nyberg, A., Landstad, B., Sjöström, M., Bååthe, F., & Nilsen, P. (2023). Healthcare in distress: A survey of mental health problems and the role of gender among nurses and physicians in Sweden. *Journal of Affective Disorders*, *339*, 104–110. https://doi.org/10.1016/j.jad.2023.07.042

Buono, C., Cortese, C. G., & Spagnoli, P. (2022). A further contribution to the Italian validation of the Burnout Assessment Tool: Measurement invariance in teachers and employees. *BPA Applied Psychology Bulletin*, *80*(295), 1–16. https://doi.org/10.26387/bpa.2022.00007

Burisch, M. (2017). *The Hamburg Burnout Inventory (HBI): Background and some early results*. Burnout Institut. https://www.burnout-institut.eu/fileadmin/user_upload/HBI-REP_V2_FINAL.pdf

Carvajal, A., Centeno, C., Watson, R., Martínez, M., & Sanz Rubiales, Á. (2011). ¿Cómo validar un instrumento de medida de la salud? *Anales del Sistema Sanitario de Navarra*, *34*(1), 63–72. https://doi.org/10.4321/S1137-66272011000100007

Consiglio, C., Massa, N., Sommovigo, V., & Fusco, L. (2023). Techno-stress creators, burnout and psychological health among remote workers during the pandemic: The moderating role of e-work self-efficacy. *International Journal of Environmental Research and Public Health*, *20*(22), 7051. https://doi.org/10.3390/ijerph20227051

Consiglio, C., Mazzetti, G., & Schaufeli, W. B. (2021). Psychometric properties of the Italian version of the Burnout Assessment Tool (BAT). *International Journal of Environmental Research and Public Health*, *18*(18), 9469. https://doi.org/10.3390/ijerph18189469

De Amorim, 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., Christensen, M., Sørengaard, T. A., Innstrand, S. T., & Schaufeli, W. B. (2023). The psychometric properties of the Burnout Assessment Tool in Norway: A thorough investigation into construct-relevant multidimensionality. *Scandinavian Journal of Psychology*, *65*(3), 479–489. https://doi.org/10.1111/sjop.12996

De Beer, L. T., Hakanen, J. J., Schaufeli, W. B., De Witte, H., Glaser, J., Kaltiainen, J., Seubert, C., & Morin, A. J. S. (2024). The burnout-depression conundrum: Investigating construct-relevant multidimensionality across four countries and four patient samples. *Psychology & Health*, 1–28. https://doi.org/10.1080/08870446.2024.2321358

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*, 95. 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*(1), 1555. https://doi.org/10.1186/s12889-022-13978-0

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

De Beer, L. T., Van Der Vaart, L., Escaffi-Schwarz, M., De Witte, H., & Schaufeli, W. B. (2024). Maslach Burnout Inventory – general survey. *European Journal of Psychological Assessment*, *40*(5), 360–375. https://doi.org/10.1027/1015-5759/a000797

De Freitas, M. F. M. B., Enumo, S. R. F., De Lara Machado, W., Messias, J. C. C., & Mendes, E. D. T. (2023). Burnout at the hospital: Healthcare workers coping with COVID-19 stress. *Estudos De Psicologia*, *40*, e220097. https://doi.org/10.1590/1982-0275202340e220097

de la Fuente, E. I., García, J., Cañadas, G. A., San Luis, C., Cañadas, G. R., Aguayo, R., de la Fuente, L., & Vargas, C. (2015). Psychometric properties and scales of the Granada Burnout Questionnaire applied to nurses. *International Journal of Clinical and Health Psychology*, *15*(2), 130–138. https://doi.org/10.1016/j.ijchp.2015.01.001

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

Dettori, J. R., Norvell, D. C., & Chapman, J. R. (2022). Fixed-effect vs random-effects models for meta-analysis: 3 points to consider. *Global Spine Journal*, *12*(7), 1624–1626. https://doi.org/10.1177/21925682221110527

De Witte, H. (1999). Job insecurity and psychological well-being: Review of the literature and exploration of some unresolved issues. *European Journal of Work and Organizational Psychology*, *8*(2), 155–177. https://doi.org/10.1080/135943299398302

du Bois, K., Sterkens, P., Lippens, L., Baert, S., & Derous, E. (2023). Beyond the hype: (How) are work regimes associated with job burnout? *International Journal of Environmental Research and Public Health*, *20*(4), 3331. https://doi.org/10.3390/ijerph20043331

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Edú-Valsania, S., Laguía, A., & Moriano, J. A. (2022). Burnout: A review of theory and measurement. *International Journal of Environmental Research and Public Health*, *19*(3), 1780. https://doi.org/10.3390/ijerph19031780

Elwér, S., Johansson, K., & Hammarström, A. (2014). Workplace gender composition and psychological distress: The importance of the psychosocial work environment. *BMC Public Health*, *14*(1), 241. https://doi.org/10.1186/1471-2458-14-241

Erschens, R., Keifenheim, K. E., Herrmann-Werner, A., Loda, T., Schwille-Kiuntke, J., Bugaj, T. J., Nikendei, C., Huhn, D., Zipfel, S., & Junne, F. (2019). Professional burnout among medical students: Systematic literature review and meta-analysis. *Medical Teacher*, *41*(2), 172–183. https://doi.org/10.1080/0142159X.2018.1457213

Felicia, F., Sudibjo, N., & Harsanti, H. R. (2023). Impact of psychosocial risk on intention to leave work during COVID-19 in Indonesia: The mediatory roles of burnout syndrome and job satisfaction. *Heliyon*, *9*(7), e17937. https://doi.org/10.1016/j.heliyon.2023.e17937

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

Gil-Monte, P. R., & Faúndez, V. E. O. (2011). Psychometric properties of the "Spanish Burnout Inventory" in Chilean professionals working to physical disabled people. *The Spanish Journal of Psychology*, *14*(1), 441–451. https://doi.org/10.5209/rev_SJOP.2011.v14.n1.40

Gómez, V., Gómez, M. T., Sánchez, L., & Navarro, D. (2022). Impacto del COVID-19: Los riesgos psicosociales en docentes no universitarios. *International Humanities Review*, *11*, 1–13. https://doi.org/10.37467/revhuman.v11.4233

Haar, J. (2021a). The state of job burnout amongst New Zealand managers: Implications for employment relations. *New Zealand Journal of Employment Relations*, *46*(1), 36–50. https://doi.org/10.24135/nzjer.v46i1.49

Haar, J. (2021b). Burnt to a crisp? Understanding drivers of burnout amongst New Zealand workers. *Evidence-Based HRM: A Global Forum for Empirical Scholarship*, *10*(2), 174–188. https://doi.org/10.1108/EBHRM-07-2021-0132

Haar, J. (2023). What are the odds of burnt-out risk and leaving the job? Turnover intent consequences of worker burnout using a two sample New Zealand study. *International Journal of Selection and Assessment*, *31*(1), 189–197. https://doi.org/10.1111/ijsa.12393

Haar, J., & O'Kane, C. (2022). A post-lockdown study of burnout risk amongst New Zealand essential workers. *Social Science & Medicine*, *306*, Article 115157. https://doi.org/10.1016/j.socscimed.2022.115157

Hadžibajramović, E., Hansson, M., Akerstrom, M., Dencker, A., & Hensing, G. (2022). Burnout among midwives – the factorial structure of the burnout assessment tool and an assessment of burnout levels in a Swedish national sample. *BMC Health Services Research*, *22*(1), 1167. https://doi.org/10.1186/s12913-022-08552-8

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

Hadžibajramović, E., Schaufeli, W., & De Witte, H. (2024). The ultra-short version of the Burnout Assessment Tool (BAT4) – development, validation, and measurement invariance across countries, age and gender. *PLoS One*, *19*(2), e0297843. https://doi.org/10.1371/journal.pone.0297843

Hakanen, J. J., & Bakker, A. B. (2017). Born and bred to burn out: A life-course view and reflections on job burnout. *Journal of Occupational Health Psychology*, *22*(3), 354–364. https://doi.org/10.1037/ocp0000053

Hartung, J. (1999). An alternative method for meta-analysis. *Biometrical Journal*, *41*(8), 901–916. https://doi.org/10.1002/(SICI)1521-4036(199912)41:8<901::AID-BIMJ901>3.0.CO;2-W

Heinemann, L. V., & Heinemann, T. (2017). Burnout research: Emergence and scientific investigation of a contested diagnosis. *Sage Open*, *7*(1), 1–12. https://doi.org/10.1177/2158244017697154.

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions. Version 5.1.0.* The Cochrane Collaboration.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*(2), 193–206. https://doi.org/10.1037/1082-989X.11.2.193

Innstrand, S. T. (2022). Burnout among health care professionals during COVID-19. *International Journal of Environmental Research and Public Health*, *19*(18), 11807. https://doi.org/10.3390/ijerph191811807

IntHout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology*, *14*(1), 25. https://doi.org/10.1186/1471-2288-14-25

Izdebski, Z., Kozakiewicz, A., Białorudzki, M., Dec-Pietrowska, J., & Mazur, J. (2023). Occupational burnout in healthcare workers, stress and other symptoms of work overload during the COVID-19 pandemic in Poland. *International Journal of Environmental Research and Public Health*, *20*(3), 2428. https://doi.org/10.3390/ijerph20032428

Jackson, S. E., Schwab, R. L., & Schuler, R. S. (1986). Toward an understanding of the burnout phenomenon. *Journal of Applied Psychology*, *71*(4), 630–640. https://doi.org/10.1037/0021-9010.71.4.630

Kannenberg, J. (2022). Occupational burnout among recreational diving instructors: Relationships with personality and sociodemographic variables. *Advances in Cognitive Psychology*, *18*(3), 190–202. https://doi.org/10.5709/acp-0361-6

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Kolachev, N., & Novikov, I. (2020). Age as a predictor of burnout in Russian public librarians. *Evidence Based Library and Information Practice*, *15*(4), 33–48. https://doi.org/10.18438/eblip29753

Kristensen, T. S., Borritz, M., Villadsen, E., & Christensen, K. B. (2005). The Copenhagen Burnout Inventory: A new tool for the assessment of burnout. *Work & Stress*, *19*(3), 192–207. https://doi.org/10.1080/02678370500297720

Kristensen, T. S., Hannerz, H., Høgh, A., & Borg, V. (2005). The Copenhagen Psychosocial Questionnaire – a tool for the assessment and improvement of the psychosocial work environment. *Scandinavian Journal of Work, Environment & Health*, *31*(6), 438–449. https://doi.org/10.5271/sjweh.948

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

Lau, S. S., Shum, E. N., Man, J. O., Cheung, E. T., Amoah, P. A., Leung, A. Y., … Okan, O. (2022). Assessing COVID-19-related health literacy and associated factors among school teachers in Hong Kong, China. *Frontiers in Public Health*, *10*, 1–14. https://doi.org/10.3389/fpubh.2022.1057782.

Lazauskaitė-Zabielskė, J., Žiedelis, A., Jakštienė, R., Urbanavičiūtė, I., & De Witte, H. (2023). The Lithuanian version of the Burnout Assessment Tool (BAT-LT): Psychometric characteristics of the primary and secondary symptoms scales. *Frontiers in Psychology*, *14*, 1–11. https://doi.org/10.3389/fpsyg.2023.1287368.

Lee, S., & Joo, M. (2023). The moderating effects of self-care on the relationships between perceived stress, job burnout and retention intention in clinical nurses. *Healthcare*, *11*(13), 1870. https://doi.org/10.3390/healthcare11131870

Lyon, T. R., & Galbraith, A. (2023). Mindful self-compassion as an antidote to burnout for mental health practitioners. *Healthcare*, *11*(20), 2715. https://doi.org/10.3390/healthcare11202715

Månsson, H., Landstad, B. J., Tjulin, Å, & Brulin, E. (2023). COVID-19-related work, managerial factors and exhaustion among general practitioners in Sweden: A cross-sectional study. *BMC Primary Care*, *24*(1), 1–12. https://doi.org/10.1186/s12875-023-02228-w.

Márquez-Lugo, I., Mosquera-Quiñónez, M., Ochoa-Granados, C., Pacavita-Sánchez, D., Palencia-Sánchez, F., & Riaño-Casallas, M. (2021). Revisión de los Instrumentos de Medición del Síndrome De Burnout-Documento de Trabajo. *SSRN*. http://doi.org/10.2139/ssrn.3841093

Marrinhas, D., Santos, V., Salvado, C., Pedrosa, D., & Pereira, A. (2023). Burnout and technostress during the COVID-19 pandemic: The perception of higher education teachers and researchers. *Frontiers in Education*, *8*, 1–10. https://doi.org/10.3389/feduc.2023.1144220.

Maslach, C. (1982). *Burnout: The cost of caring*. Prentice-Hall.

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., & Jackson, S. E. (1982). Burnout in health professions: A social psychological analysis. In G. S. Sanders & J. Suls (Eds.), *Social psychology of health and illness* (pp. 227–251). Erlbanm.

Maslach, C., & Leiter, M. P. (2008). Early predictors of job burnout and engagement. *Journal of Applied Psychology*, *93*(3), 498–512. https://doi.org/10.1037/0021-9010.93.3.498

Maslach, C., & Leiter, M. P. (2016). Understanding the burnout experience: Recent research and its implications for psychiatry. *World Psychiatry*, *15*(2), 103–111. https://doi.org/10.1002/wps.20311

Mazur, J., Kozakiewicz, A., Białorudzki, M., & Izdebski, Z. (2023). Selected aspects of patient communication and occupational burnout among nurses working during the COVID-19 pandemic. *Pielęgniarstwo XXI Wieku*, *22*(3), 139–146. https://doi.org/10.2478/pielxxiw-2023-0022

Mazzetti, G., Consiglio, C., Santarpia, F. P., Borgogni, L., Guglielmi, D., & Schaufeli, W. B. (2022). Italian validation of the 12-item version of the Burnout Assessment Tool (BAT-12). *International Journal of Environmental Research and Public Health*, *19*(14), 8562. https://doi.org/10.3390/ijerph19148562

Meier, S. T., & Kim, S. (2021). 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

Moreno, B., Herrer, M. G., Hernández, E. G., & Adán, J. C. M. (2006). Nuevos planteamientos en la evaluación del burnout. La evaluación específica del desgaste profesional médico. *Atención Primaria*, *38*(*10*), 544–549. http://doi.org/10.1157/13095925

National Institutes of Health. (2021). *Study quality assessment tools*. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Oprea, B., Iliescu, D., & De Witte, H. (2021). Romanian short version of the Burnout Assessment Tool: Psychometric properties. *Evaluation & the Health Professions*, *44*(4), 406–415. https://doi.org/10.1177/01632787211048924

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The Prisma 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*(71), 1–9. https://doi.org/10.1136/bmj.n71.

Palvimo, T., Vauhkonen, A., & Hult, M. (2023). The associations among destructive leadership, job demands and resources, and burnout among nurses: A cross-sectional survey study. *Journal of Nursing Management*, *2023*(1), Article 4289450. https://doi.org/10.1155/2023/4289450

Parker, G., & Tavella, G. (2022). Burnout: A case for its formal inclusion in classification systems. *World Psychiatry*, *21*(3), 467–468. https://doi.org/10.1002/wps.21025

Pereira, H., Fehér, G., Tibold, A., Monteiro, S., Costa, V., & Esgalhado, G. (2021). The impact of shift work on occupational health indicators among professionally active adults: A comparative study. *International Journal of Environmental Research and Public Health*, *18*(21), 11290. https://doi.org/10.3390/ijerph182111290

Pereira, H., Gonçalves, V. O., & Assis, R. M. D. (2021). Burnout, organizational self-efficacy and self-esteem among Brazilian teachers during the COVID-19 pandemic. *European Journal of Investigation in Health, Psychology and Education*, *11*(3), 795–803. https://doi.org/10.3390/ejihpe11030057

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

Pines, A., Aronson, E., & Kafry, D. (1981). *Burnout: From tedium to personal growth*. The Free Press.

Pines, A., & Maslach, C. (1978). Characteristics of staff burnout in mental health settings. *Hospital & Community Psychiatry*, *29*(4), 233–237.

Popescu, B., Maricuțoiu, L. P., & De Witte, H. (2024). The student version of the Burnout Assessment Tool (BAT): Psychometric properties and evidence regarding measurement validity on a Romanian sample. *Current Psychology*, *43*(3), 2037–2051. https://doi.org/10.1007/s12144-023-04232-w

Purvanova, R. K., & Muros, J. P. (2010). Gender differences in burnout: A meta-analysis. *Journal of Vocational Behavior*, *77*(2), 168–185. https://doi.org/10.1016/j.jvb.2010.04.006

Ranabhat, R., Giri, A., Sapkota, B., Shrestha, R., & Shrestha, S. (2023). A cross-sectional assessment of burnout presence among pharmacy professionals at Kathmandu Valley, Nepal. *SAGE Open Medicine*, *11*. http://doi.org/10.1177/20503121231215237

Redelinghuys, K., & Morgan, B. (2023). Psychometric properties of the Burnout Assessment Tool across four countries. *BMC Public Health*, *23*(1), 824. https://doi.org/10.1186/s12889-023-15604-z

Romano, L., Angelini, G., Consiglio, P., & Fiorilli, C. (2022). An Italian adaptation of the Burnout Assessment Tool-Core symptoms (BAT-C) for students. *Education Sciences*, *12*(2), 124. https://doi.org/10.3390/educsci12020124

Sakakibara, K., Shimazu, A., Toyama, H., & Schaufeli, W. B. (2020). Validation of the Japanese version of the Burnout Assessment Tool. *Frontiers in Psychology*, *11*, 1819. https://doi.org/10.3389/fpsyg.2020.01819

Salmela-Aro, K., Rantanen, J., Hyvönen, K., Tilleman, K., & Feldt, T. (2011). Bergen Burnout Inventory: Reliability and validity among Finnish and Estonian managers. *International Archives of Occupational and Environmental Health*, *84*(6), 635–645. https://doi.org/10.1007/s00420-010-0594-3

Salvagioni, D. A., Melanda, F. N., Mesas, A. E., González, A. D., Gabani, F. L., & Andrade, S. M. (2017). Physical, psychological and occupational consequences of job burnout: A systematic review of prospective studies. *PLoS One*, *12*(10), e0185781. https://doi.org/10.1371/journal.pone.0185781

Sánchez-Meca, J., López-López, J. A., & López-Pina, J. A. (2013). Some recommended statistical analytic practices when reliability generalization studies are conducted. *British Journal of Mathematical and Statistical Psychology*, *66*(3), 402–425. https://doi.org/10.1111/j.2044-8317.2012.02057.x

Sánchez-Meca, J., & López-Pina, J. A. (2008). El enfoque meta-analítico de generalización de la fiabilidad. *Acción Psicológica*, *5*(2), 37–64. https://doi.org/10.5944/ap.5.2.457

Sánchez-Meca, J., López-Pina, J. A., & López López, J. A. (2009). Generalización de la fiabilidad: Un enfoque metaanalítico aplicado a la fiabilidad. *Fisioterapia*, *31*(6), 262–270. https://doi.org/10.1016/j.ft.2009.05.005

Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods*, *13*(1), 31–48. https://doi.org/10.1037/1082-989X.13.1.31

Sánchez-Meca, J., Marín-Martínez, F., López-López, J. A., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-García, J. J., López-Pina, J. A., Blázquez-Rincón, D. M., López-Ibáñez, C., & López-Nicolás, R. (2021). Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist. *Research Synthesis Methods*, *12*(4), 516–536. https://doi.org/10.1002/jrsm.1487

Schaufeli, W. (2023). Burnout: A critical overview. In L. M. Lapierre & S. C. Cooper (Eds.), *Organizational stress and well-being* (pp. 214–259). Cambridge University Press. https://doi.org/10.1017/9781009268332.012

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire. *Educational and Psychological Measurement*, *66*(4), 701–716. https://doi.org/10.1177/0013164405282471

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., & De Witte, H. (2023). A fresh look at burnout: The Burnout Assessment Tool (BAT). In C. U. Krägeloh, M. Alyami, & O. N. Medvedev (Eds.), *International handbook of behavioral health assessment* (pp. 1–24). Springer. https://doi.org/10.1007/978-3-030-89738-3_54-1

Schaufeli, W. B., De Witte, H., & Desart, S. (2019). *User manual – Burnout Assessment Tool (BAT) – version 2.0* (Internal Report). KU Leuven.

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., & Enzmann, D. (1998). *The burnout companion to study and practice: A critical analysis*. Taylor and Francis.

Schaufeli, W. B., Leiter, M. P., & Maslach, C. (2009). Burnout: 35 years of research and practice. *Career Development International*, *14*(3), 204–220. https://doi.org/10.1108/13620430910966406

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

Shankland, R. R., Kotsou, I., Vallet, F. F., Bouteyre, E. E., Dantzer, C. C., & Leys, C. (2019). Burnout in university students: The mediating role of sense of coherence on the relationship between daily hassles and burnout. *Higher Education*, *78*(1), 91–113. https://doi.org/10.1007/s10734-018-0332-4

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

Sidik, K., & Jonkman, J. N. (2002). A simple confidence interval for meta-analysis. *Statistics in Medicine*, *21*(21), 3153–3159. https://doi.org/10.1002/sim.1262

Sinval, J., Vazquez, A. C. S., Hutz, C. S., Schaufeli, W. B., & Silva, S. (2022). Burnout Assessment Tool (BAT): Validity evidence from Brazil and Portugal. *International Journal of Environmental Research and Public Health*, *19*(3), 1344. https://doi.org/10.3390/ijerph19031344

Siu, O. (2003). Job stress and job performance among employees in Hong Kong: The role of Chinese work values and organizational commitment. *International Journal of Psychology*, *38*(6), 337–347. https://doi.org/10.1080/00207590344000024

Sørengaard, T. A., & Langvik, E. (2022). The protective effect of fair and supportive leadership against burnout in police employees. *Safety and Health at Work*, *13*(4), 475–481. https://doi.org/10.1016/j.shaw.2022.09.002

Tanriver-Ayder, E., Faes, C., van de Casteele, T., McCann, S. K., & Macleod, M. R. (2021). Comparison of commonly used methods in random effects meta-analysis: Application to preclinical data in drug discovery research. *BMJ Open Science*, *5*(1), e100074. https://doi.org/10.1136/bmjos-2020-100074

Tomas, J., Seršić, D. M., Mikac, U., Rebernjak, B., Buško, V., & De Witte, H. (2023). Validation of the Croatian version of the short form of the Burnout Assessment Tool: Findings from a nationally representative sample. *International Journal of Selection and Assessment*, *32*(1), 40–53. https://doi.org/10.1111/ijsa.12447

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*(1), 6–20. https://doi.org/10.1177/0013164498058001002

van Den Boogert, F., Spaan, P., Sizoo, B., Bouman, Y. H. A., Hoogendijk, W. J. G., & Roza, S. J. (2022). Sensory processing, perceived stress and burnout symptoms in a working population during the COVID-19 crisis. *International Journal of Environmental Research and Public Health*, *19*(4), 2043. https://doi.org/10.3390/ijerph19042043

Vanheule, S., Rosseel, Y., & Vlerick, P. (2006). The factorial validity and measurement invariance of the Maslach Burnout Inventory for human services. *Stress and Health*, *23*(2), 87–91. https://doi.org/10.1002/smi.1124

Verbraak, M., Kleyweg, J., van de Griendt, J., & Hoogduin, K. (2008). Nadere verkenning van de psychometrische kwaliteiten van de BurnOut-Neurasthenie Klachten Schaal (BO-NKS): Wat valt er te meten? *Directieve Therapie*, *28*(3), 187–211. https://doi.org/10.1007/BF03076438

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

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Vinueza-Solórzano, A. M., Portalanza-Chavarría, C. A., De Freitas, C. P. P., Schaufeli, W. B., De Witte, H., Hutz, C. S., & Souza Vazquez, A. C. (2021). The Ecuadorian version of the Burnout Assessment Tool (BAT): Adaptation and validation. *International Journal of Environmental Research and Public Health*, *18*(13), 7121. https://doi.org/10.3390/ijerph18137121

Wheeler, D. L., Vassar, M., Worley, J. A., & Barnes, L. L. B. (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). *11th revision of the International Classification of Diseases (ICD-11)*. https://icd.who.int/en

World Health Organization. (2022). *ICD-11: International Classification of Diseases (11th revision)*. https://icd.who.int/

Xie, W., Chen, L., Feng, F., Okoli, C. T., Tang, P., Zeng, L., Jin, M., Zhang, Y., & Wang, J. (2021). The prevalence of compassion satisfaction and compassion fatigue among nurses: A systematic review and meta-analysis. *International Journal of Nursing Studies*, *120*, Article 103973. https://doi.org/10.1016/j.ijnurstu.2021.103973