WILEY

Towards a new approach to job-related distress: A three-sample study of the Occupational Depression Inventory

Renzo Bianchi[1,2] 🟢 | Jay Verkuilen[3] | James F. Sowden[4] | Irvin Sam Schonfeld[3]

[1]University of Neuchâtel, Neuchâtel, Switzerland

[2]Norwegian University of Science and Technology (NTNU), Trondheim, Norway

[3]City University of New York, New York City, New York, USA

[4]Flinders University, Adelaide, Australia

Correspondence
Renzo Bianchi, Institute of Work and Organizational Psychology (IPTO), University of Neuchâtel, Émile-Argand 11, Neuchâtel 2000, Neuchâtel, Switzerland.
Email: renzo.bianchi@unine.ch

Abstract

The Occupational Depression Inventory (ODI) was recently developed to assess depressive symptoms that individuals specifically attribute to their work. One purpose of the ODI is to respond to limitations of current assessments of job-related distress, most notably, assessments relying on the burnout construct. In this study, we conducted a thorough examination of the psychometric and structural properties of the ODI using exploratory structural equation modelling bifactor analysis and Mokken scale analysis. The study involved three samples of employed individuals, recruited in France ($N = 3454$), Switzerland ($N = 1971$), and Australia ($N = 1485$). Results were consistent across the three samples. The ODI exhibited essential unidimensionality and excellent total-score reliability—as indexed by McDonald's omega, Cronbach's alpha, Guttman's lambda-2, and the Molenaar-Sijtsma statistic. We found evidence for measurement invariance across sexes, age groups, and samples. Mokken scale analysis revealed that the ODI's scalability was strong. No monotonicity violation was detected. Invariant item ordering showed sufficient accuracy. In all three samples, suicidal ideation was the least commonly endorsed item—thus acting as a sentinel item—and fatigue/loss of energy was the most commonly endorsed item. The ODI exhibits excellent psychometric and structural properties, suggesting that occupational health specialists can effectively employ the instrument.

KEYWORDS
factor analysis, item response theory, job-related distress, Mokken scale analysis, psychosocial hazards, psychosocial risks, work-related health

1 | INTRODUCTION

The depressogenic effects of insurmountable adversity in working life (also known as unresolvable job stress) have elicited growing concerns among occupational health specialists (Madsen et al., 2017; Matthews et al., 2021; McEwen, 2012; Melchior et al., 2007; Nieuwenhuijsen

et al., 2010). A large body of empirical findings and theory indicates that unresolvable stress is a basic depressogenic factor, even among individuals with no observable susceptibility to depression (Dohrenwend, 2000; Grahek et al., 2019; Pryce et al., 2011; Seligman, 1975; Wichers, 2014; Willner et al., 2013). The existence of unresolvable stress extends to the workplace, where uncontrollable and

unchangeable stressors can be encountered (Schonfeld & Chang, 2017). Workplace suicides, which can be viewed as an endpoint of job-related depressive processes, have been rising in recent years in countries such as the U.S.A (Howard et al., 2021; Tiesman et al., 2015; U.S. Bureau of Labor Statistics, 2019). In addition to its health-damaging and life-threatening character, job-related distress causes reduced performance, lost productivity, and job turnover, imposing a financial burden on organizations and society as a whole (European Agency for Safety and Health at Work, 2014; Gonzalez-Mulé & Cockburn, 2021; Hassard et al., 2018; Park & Shaw, 2013). Occupational health specialists' need for efficient means of assessing job-related distress, allowing informed decisions to be made at micro (e.g., individual), meso (e.g., organizational), and macro (e.g., public health) levels, is particularly salient in such a context.

The Occupational Depression Inventory (ODI) was recently developed to (a) quantify the severity of depressive symptoms that individuals specifically attribute to their jobs and (b) establish provisional diagnoses of job-ascribed depression (Bianchi & Schonfeld, 2020, 2021b).[1] The ODI was devised with the aim of overcoming limitations of current assessments of job-related distress, most notably, assessments relying on the burnout construct (Bianchi et al., 2021; Cox et al., 2005; Meier & Kim, 2021; Schonfeld & Bianchi, 2021; Schonfeld et al., 2019; Schwenk & Gold, 2018; Taris, 2006; Vinkers & Schaafsma, 2021).

2 | LIMITATIONS OF BURNOUT-RELATED ASSESSMENTS OF JOB-RELATED DISTRESS

Several limitations to an effective use of the burnout construct in assessments of job-related distress have been identified over the years. These limitations bear on both the definition and the measurement of burnout.

A first limitation is that there are no clear or consensual diagnostic criteria for burnout (Heinemann & Heinemann, 2017; Schwenk & Gold, 2018; Vinkers & Schaafsma, 2021). As an illustration, Rotenstein et al. (2018) found 142 different characterisations of burnout in a review of 182 studies of physician distress. The impossibility of diagnosing burnout is problematic for occupational health specialists seeking to identify cases of burnout and ascertain burnout's prevalence (e.g., Hewitt et al., 2020; Rotenstein et al., 2018; Tyssen, 2018), treat 'burned out' individuals (e.g., Ahola et al., 2017; Dreison et al., 2018), or isolate biomarkers for burnout (e.g., Danhof-Pont et al., 2011; Strikweda et al., 2021). The absence of a nosological characterisation of burnout has also been an obstacle to the legal recognition of clinically-relevant levels of job-related distress (with consequences for the management of sick pay and health insurance policies), and has obstructed public health decision-making (Bianchi et al., 2021; Schears, 2017; Schwenk & Gold, 2018). Making authoritative public health decisions (e.g., to establish workplace rules and regulations) about a phenomenon that cannot be diagnosed in a clear and consensual manner is challenging (Rotenstein et al., 2018; Schwenk & Gold, 2018; Vinkers & Schaafsma, 2021).

Second, although burnout is deemed to be a syndrome comprising exhaustion, cynicism, and inefficacy symptoms, it has been expressly recommended that exhaustion, cynicism, and inefficacy symptoms be approached separately (Maslach et al., 2016). Following this recommendation, investigators end up with three different constructs (exhaustion, cynicism, and inefficacy), none of which is by itself burnout. The prescribed operationalisation of burnout thus does not align with the entity's formal conceptualisation (Kristensen et al., 2005; Shirom, 2005). On the one hand, burnout is supposed to manifest itself in the combination of exhaustion, cynicism, and inefficacy; on the other hand, combining exhaustion, cynicism, and inefficacy is considered ill-advised and explicitly discouraged (Maslach et al., 2016). On a related note, if it is not advisable to examine exhaustion, cynicism, and inefficacy together, the reason for considering the three entities a *syndrome* is not clear (Kristensen et al., 2005; Shirom, 2005). By definition, a syndrome refers to a 'grouping of signs and symptoms, based on their frequent co-occurrence' (American Psychiatric Association, 2013, p. 830; for a similar definition applied to burnout research, see Shirom, 2005). Commenting on the idea that burnout is a syndrome, Semmer et al. (2015) concluded that 'neither a reduction to one dimension ... nor the use of each dimension ... separately would capture burnout as a construct in its own right' (p. 38).

A third concern pertains to the conditions under which the Maslach Burnout Inventory (MBI; Maslach et al., 2016), by far the most popular measure of burnout, was developed. The MBI has played a key role in bringing the burnout construct into the spotlight. Taking a historical perspective on the burnout construct, Schaufeli (2017) reported that prior to the release of the MBI, "the scientific community deemed burnout a 'pseudoscientific' or 'fad' concept and denounced it as 'pop psychology'" (p. 108). The MBI has crystallised the dominant definition of burnout and profoundly influenced the entire area of burnout research (Schaufeli & Enzmann, 1998; Shirom, 2005). However, Maslach and Jackson (1981) underlined that the studies on which the MBI is based were 'very exploratory' (p. 100), in the sense that their designs and analyses were rudimentary and that they incorporated few methodological safeguards (e.g., for the control of researcher bias). In a similar vein, Schaufeli (2003) indicated that the development of the MBI was 'neither grounded in firm clinical observation nor based on sound theorising' (p. 3). Looking into the process through which the MBI's items and components were generated, Schaufeli and Enzmann (1998) noted that the instrument had been created inductively 'by factor-analysing a rather arbitrary set of items' (p. 188). The authors observed that if a different 'arbitrary set of items' had been initially selected, most probably, the burnout syndrome would have been defined differently.[2] Taken together, these observations suggest that the dominant, three-component definition of burnout may be an artefact of the MBI's problematic development process.

In summary, the burnout construct is affected by important definitional and measurement problems that bear on its usability. The ODI responds to this state of affairs and offers occupational health specialists and work and organizational psychologists a different

approach to job-related distress (Schonfeld & Bianchi, 2021). Because depressive symptoms constitute a basic human response to insurmountable adversity (or unresolvable stress), even among individuals with no noticeable vulnerability to clinical depression (Dohrenwend, 2000; Grahek et al., 2019; Pryce et al., 2011; Seligman, 1975; Wichers, 2014; Willner et al., 2013), addressing the issue of job-related distress in the area of depression research appears to be particularly relevant (Bianchi et al., 2021; Schwenk & Gold, 2018).

3 | THE OCCUPATIONAL DEPRESSION INVENTORY (ODI)

The ODI reflects an approach to job-related distress that is anchored in stress and depression research (Bianchi & Schonfeld, 2020). The ODI was designed with reference to the diagnostic criteria for major depression of the *Diagnostic and statistical manual of mental disorders* (fifth ed.; DSM-5; American Psychiatric Association, 2013). The symptom items of the ODI thus target anhedonia (i.e., loss of pleasure and interest in things previously experienced as pleasurable and interesting), depressed mood (e.g., feelings of helplessness and hopelessness, loss of motivation), sleep alterations, fatigue/loss of energy, appetite alterations, feelings of worthlessness, cognitive impairment, psychomotor alterations, and suicidal ideation. In contrast to classical depression scales, which are 'cause-neutral', the ODI assesses depressive symptoms in etiological connection to work (Bianchi & Schonfeld, 2020). Thus, for instance, respondents are not merely asked if they felt worthless (a symptom of major depression); respondents are asked if their *experience at work* made them feel worthless. Reliance on causal attributions has been commonplace in (occupational) health research. The *Stress in America*™ survey commissioned by the American Psychological Association, for instance, has relied on causal attributions to identify leading sources of stress among the general public (American Psychological Association, 2015). Moreover, causal attributions are key to the diagnosis of a number of conditions, including posttraumatic stress disorder and acute stress disorder (American Psychiatric Association, 2013). It is noteworthy that, by assessing symptoms such as depressed mood, fatigue/loss of energy, and feelings of worthlessness, the ODI covers the state of exhaustion and demotivation and sense of failure purportedly at the core of the experience of burnout. As previously noted, the ODI also assesses a host of other symptoms (e.g., cognitive impairment, suicidal ideation).

The ODI combines dimensional (or continuum-based) and categorical (or diagnostic) foci on job-related distress (Bianchi & Schonfeld, 2020). This dual-lens approach reconciles process- and state-based views of job-related distress (see Schaufeli & Enzmann, 1998; see also Bianchi, 2020). In the ODI framework, dimensions and categories are not regarded as antinomic; dimensions and categories are treated as two complementary perspectives on the phenomenon of interest (Pickles & Angold, 2003). Depending on their research goals, investigators may prefer one perspective to the other or choose to coordinate the two approaches. In ODI research, a categorical approach appears to be particularly relevant when it comes, for instance, to estimating the prevalence of the phenomenon or identifying individuals in urgent need of help. A dimensional approach may be preferred when investigating, for example, the dynamics of symptom development because the entire continuum of occupational depression is of interest in such processes (e.g., Wichers, 2014).

Available evidence, based on data collected in the U.S.A., New Zealand, South Africa, Spain, and France, consistently suggests that the ODI is a reliable and valid instrument (Bianchi et al., 2022; Bianchi & Schonfeld, 2020; Hill et al., 2021). Cronbach's alphas and McDonald's omegas around 0.90 were observed. Exploratory structural equation modelling (ESEM) bifactor analysis suggested that the ODI meets the requirements for essential unidimensionality,[3] with the general factor accounting for nearly 90% of the common variance extracted (Bianchi & Schonfeld, 2020; Rodriguez et al., 2016). Furthermore, research on the ODI indicates that the scale has satisfactory convergent and discriminant validities vis-à-vis cause-neutral measures of depressive symptoms (Bianchi & Schonfeld, 2020; Hill et al., 2021). A balance of convergent and discriminant validities can be expected because the ODI is meant to assess depressive symptoms that individuals attribute to their work, whereas classical measures of depressive symptoms involve no etiological considerations. In terms of criterion validity, the ODI manifested moderate to large correlations, in the expected direction, with a number of job-related variables, including job satisfaction, dedication to work, willingness to stay in the job, social support in working life, and active search for another job/position. Regarding context-free variables, ODI-assessed symptoms showed substantial associations with higher trait anxiety, lower general health status, and lower life satisfaction. The ODI was also negatively linked to objective cognitive performance (Bianchi et al., 2022; Bianchi & Schonfeld, 2021a, 2022). Though highly encouraging, initial findings pertaining to the psychometric and structural properties of the ODI have unclear generalisability. The ODI has been in circulation for only a short period. Research that delves deeper into the instrument's characteristics is required.

4 | THE PRESENT STUDY

In this study, we inquired into the psychometric and structural properties of the ODI in three different countries—Australia, France, and Switzerland—relying on both ESEM bifactor analysis (Marsh et al., 2014; Reise et al., 2007) and Mokken scale analysis (MSA; Mokken, 1971; Molenaar, 1982; van der Ark, 2007, 2012). In so doing, we aimed to submit the ODI to a stringent examination and learn more about the instrument's characteristics. A key aspect of our examination was to ascertain whether the ODI can be confidently used as a unidimensional measure of occupational depression based on the scale's total score. The need for deep(er) examinations of psychological scales' psychometric and structural characteristics has often been underscored (Hussey & Hughes, 2020). Cortina

et al. (2020) recently noted that '[t]he distance between actual and recommended scale development and evaluation practices may have reached a magnitude that should lead us to question our conclusions regarding organizational phenomena…' (p. 1352). The importance of correcting such a trend cannot be overstated. The reliability and validity of the knowledge that we produce directly depend on the reliability and validity of our measures. Some psychological scales are used for years, even decades, before it is realized that their psychometric properties are, in fact, problematic (Cortina et al., 2020; Hussey & Hughes, 2020). It is thus essential that the characteristics of new instruments, such as the ODI, be subjected to a stringent examination.

Exploratory structural equation modelling bifactor analysis is built on assumptions that are less rigid, more complexity-compatible, and ultimately more realistic than those underlying classical confirmatory factor analysis (CFA) (Marsh et al., 2014; Morin et al., 2016). On a related note, bifactor analysis is useful for handling violations of local independence should such violations be detected (Reise et al., 2007). A bifactor model partitions the covariance among a set of items into a general factor and specific factors (or bifactors) that can be defined in a theory-driven manner. Exploratory structural equation modelling bifactor analysis is particularly useful in determining whether a scale is 'unidimensional enough' in order for its total score to be justifiably employed by investigators (Rodriguez et al., 2016).

Mokken scale analysis is a nonparametric Item Response Theory (IRT) method that is related to the older Guttman scaling method but with an error theory. Mokken scale analysis has recently gained popularity in applied psychological research because of its ability to evaluate the usability of a test's total score in a highly detailed manner (e.g., Dima, 2018; Stochl et al., 2012). Mokken scale analysis notably allows for an inquiry into scalability, monotonicity, and invariant item ordering (IIO) properties. These properties inform the investigator on the extent to which (a) a scale's items are ordered by degree of 'difficulty' (i.e., probability of endorsement) in measuring a latent variable, (b) items order respondents accurately on the latent continuum assumed to underlie the scale, and (c) respondents order items in a similar manner regardless of where respondents stand on the latent continuum. In the context of a depression scale such as the ODI, item 'difficulty' reflects the idea that all items do not have the same probability of being endorsed because all depressive symptoms do not have the same gravity (Meijer & Baneke, 2004). For instance, in the ODI, the fatigue/loss of energy item is likely to be 'easier' to endorse than the item referring to feelings of worthlessness. Such a hierarchical pattern can be expected because fatigue/loss of energy, though relevant to depression, constitutes a relatively common and nonspecific symptom whereas feelings of worthlessness are less common, more specific to depression, and denote more severe depressive symptomatology (Beck & Alford, 2009).

Learning more about the psychometric and structural properties of the ODI is crucial for ascertaining whether occupational health specialists can confidently employ the instrument (e.g., based on its total score). The ODI is the only measure of its kind and may help occupational health specialists overcome the limitations of current approaches to, and measures of, job-related distress (Bianchi et al., 2021; Cox et al., 2005; Meier & Kim, 2021; Rotenstein et al., 2018; Schwenk & Gold, 2018; Taris, 2006; Vinkers & Schaafsma, 2021). Because job-related distress can have devastating effects on worker health, it is essential that occupational health researchers and practitioners have instruments allowing them to assess job-related distress reliably and validly.

5 | METHODS

5.1 | Study samples

The present study involved three samples. Our first sample was recruited in France; our second sample, in Switzerland; and our third sample, in Australia. Our French and Australian samples consisted of education staff members employed in K-12 schools. Education staff members are substantially affected by chronic job stress (e.g., through workplace violence) and exhibit worryingly high rates of job-leaving (American Psychological Association, 2021; Galand et al., 2007; Ingersoll et al., 2018; Longobardi et al., 2019; Reddy et al., 2013; Schonfeld, 2001, 2006). The French sample comprised 3454 participants, of whom 584 were men and 2870 were women, consistent with the large overrepresentation of women in the French education system (https://www.education.gouv.fr/). Participants' mean age was 45 ($SD = 10$; range: 20–67). The Australian sample comprised 1485 participants, of whom 97 were men and 1347 were women, consistent with the large overrepresentation of women among Australian education staff members (https://www.abs.gov.au/); sex-related information was unreported for 41 participants. Participants' mean age was 40 ($SD = 10$; range: 21–69), noting that 255 participants did not provide information about their age. The Swiss sample comprised 1971 participants employed in a variety of occupational areas (e.g., marketing, human resources, financial counselling). Of the 1971 participants, 1411 were women and 560 were men. Participants' mean age was 36 ($SD = 12$; range: 18–75), noting that four participants did not provide information about their age. The size of each of our three samples meets the minimum sample size requirements established in factor analytic and Mokken scale analytic research (Kline, 2016; Straat et al., 2014).

All participants took part in online surveys administered through Qualtrics (https://www.qualtrics.com/). The only eligibility criterion for participating in the study was to be employed. Participants were reached through direct contacts with organizations and professional associations as well as social media. Participation in the study was voluntary. No compensation or incentives were offered. Confidentiality was guaranteed for each respondent. Participants were informed they could withdraw from the survey at any time and for any reason. The samples were recruited in the context of studies conducted in accordance with the ethical standards of the institutional review boards of Flinders University and the University of Neuchâtel.

5.2 | Measure of interest

Our measure of interest was the ODI (Bianchi & Schonfeld, 2020). The ODI comprises nine core polytomous items (e.g., 'My experience at work made me feel like a failure'; 'I felt exhausted because of my work'; 'I thought that I'd rather be dead than continue in this job'). The items of the ODI reference the nine diagnostic symptom criteria for major depression found in the DSM-5 (American Psychiatric Association, 2013). Symptoms are assessed within a 2-week time window. Items are graded on a 4-point frequency scale, from 0 for 'never or almost never' to 3 for 'nearly every day'. For all items, higher scores signal more severe symptoms. The ODI was initially developed in English and French. We used the French version of the instrument in France and Switzerland, and the English version of the instrument in Australia. Missing values were not allowed for any of the ODI's items given the conduct of analyses at the item level. Descriptive statistics pertaining to the nine core symptom items of the ODI are available in Table S1a–c.

The ODI incorporates a diagnostic algorithm intended to identify likely cases of job-ascribed depression (Bianchi & Schonfeld, 2020). The ODI's diagnostic algorithm reflects DSM-5 diagnostic criteria for major depression and is oriented towards a categorical approach to depression in the workplace. About 7% of French participants ($n = 237$), 3% of Swiss participants ($n = 64$), and 17% of Australian participants ($n = 258$) met the criteria for a provisional diagnosis of job-ascribed depression in this study.

The ODI's instructions to respondents were designed to discourage hasty attributions of depressive symptoms to work. Respondents are instructed to envisage multiple symptom sources before answering. The possibility that respondents may not know where their symptoms come from is also explicitly planned for. If a respondent believes a symptom derives from a source other than work (e.g., marital problems, family problems) or cannot attribute a symptom to any particular source, the respondent is instructed to select '0' when answering. Ultimately, respondents are asked to report symptoms (by selecting a score other than '0') only if they feel able to establish *a clear link* between their symptoms and their job.

Cronbach's alpha was 0.89 in the French sample, 0.88 in the Swiss sample, and 0.90 in the Australian sample. Kendall τ_B and Spearman ρ correlations among the items of the ODI, together with their 95% confidence intervals, are displayed in Table S1d–f. All correlations were positive and statistically significant at $p < 0.001$.

5.3 | Data analyses

We first examined the factorial structure and dimensionality of the ODI by relying on ESEM bifactor analysis (Marsh et al., 2014; Reise et al., 2007). We conducted our ESEM bifactor analysis in Mplus version 8.6 (Muthén & Muthén, 1998-2021). As per Bianchi and Schonfeld (2020), we considered two bifactors in addition to the general factor in our ESEM bifactor analysis. The two bifactors corresponded to the anhedonic-somatic symptom items and dysphoric symptom items found in the ODI. We treated all items as ordinal, employed the weighted least squares -- mean and variance adjusted -- estimator, and used *both* a bi-geomin and a partially specified target rotation (PSTR) to check on the consistency of the findings. A PSTR renders the analysis more confirmatory (Marsh et al., 2014; Verkuilen et al., 2021). The two target bifactors were defined based on the anhedonic-somatic symptom items (items 1, 3, 4, 5, 7, and 8) and dysphoric symptom items (items 2, 6, and 9) populating the ODI. Compared to common-practice CFA, an advantage of the PSTR is that no loadings are fixed to be equal to 0. Instead, loadings are 'encouraged' to get as close to 0 as possible by the loss function, allowing factor complexity to be taken into account. Bianchi and Schonfeld's (2020) bifactor modelling of the ODI reflects the view that, although the instrument is likely to exhibit a degree of multidimensionality on account of its anhedonic-somatic symptom items and dysphoric symptom items (a degree of multidimensionality that needs to be modelled), this degree of multidimensionality is not high enough to prevent the use of the test's total score or call into question the unitary nature of the occupational depression construct (essential unidimensionality). We relied on the following fit indices: The Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Standardized Root Mean Squared Residual (SRMR). We used the Explained Common Variance (ECV) statistic to estimate the proportion of common variance extracted accounted for by the general factor. An ECV index close to, or above, 0.80 is suggestive of essential unidimensionality (Rodriguez et al., 2016). We relied on the ω and ω hierarchical (ωH) coefficients derived from our bifactor model to further inspect the reliability of the ODI, compute the correlations between the general factor and the observed total scores, and estimate the proportion of the reliable variance in the observed total scores that was attributable to the general factor.

In addition, we examined, in each of our three samples, measurement invariance across sexes (males vs. females) and age groups ('younger' vs. 'older' respondents based on a median split of the age variable). We investigated the equivalence of (a) overall factorial structures—configural invariance, (b) factor loadings—metric invariance, and (c) item thresholds—scalar invariance (Putnick & Bornstein, 2016). As recommended in the context of ordinal factor analysis (Shi et al., 2020, 2021), we focussed on delta changes in CFI (ΔCFI) and SRMR (ΔSRMR). We adopted conservative threshold values for flagging violations of measurement invariance, that is, -0.005 for ΔCFI and 0.005 for ΔSRMR (Putnick & Bornstein, 2016; Rutkowski & Svetina, 2014). Following the same modus operandi, we examined measurement invariance at a between-sample level. In order to maximise the scope of our examination, and consistent with our aim of subjecting the ODI to a stringent analysis, our between-sample analysis mobilised not only our French, Swiss, and Australian samples but also the three samples used in the ODI's initial validation study conducted by Bianchi and Schonfeld (2020). These three samples came from the U.S.A. ($N = 312$), New Zealand ($N = 492$), and France ($N = 1450$) and are described in detail in Bianchi and Schonfeld's (2020) article.

We then turned to MSA. Like other scales built with IRT, especially those built in the Rasch tradition, a Mokken scale consists of items ordered by degree of 'difficulty' in measuring a latent variable. Unlike parametric IRT models, Mokken scaling does not assume a particular functional form for the relationship between item and latent variable. While this entails some potential loss of statistical power and is not as flexible in terms of instrument design (e.g., targeting a particular test information curve), MSA better aligns with the idea that a scale's total score should accurately reflect each individual's status on a latent variable, a common assumption when a total score is used as a barometer in applied settings.

Mokken scale analysis is founded on the Loevinger's homogeneity (*H*), or scalability, coefficient, which itself is rooted in Guttman scaling but considers items pairwise. Scalability implies that the endorsement of more 'difficult' items (or items referencing more severe symptoms) is related to a higher probability of endorsing 'easier' items (or items referencing less severe symptoms), while the opposite does not apply (Dima, 2018). A Guttman error occurs when an item that should be easier is not endorsed but a more difficult one is. *H* for a pair of items is based on the ratio of observed Guttman errors compared to those expected if the items were independent. For a pair of binary items, *H* is equivalent to the ϕ/ϕ_{max} coefficient (Sijtsma & Molenaar, 2002, p. 52). Item and scale *H* statistics are based on averages of *H* statistics for item pairs. Later research efforts generalised beyond binary items to polytomous ones.

In addition to scalability, we inspected monotonicity and IIO, two other aspects of MSA (Stochl et al., 2012). Monotonicity implies that, as one moves from the low end of a latent variable to its high end, the probability of endorsing an item should not decrease (Dima, 2018). Invariant item ordering implies that items keep the same order of 'difficulty' at all levels of the latent variable (Dima, 2018).[4]

We conducted our MSA using the Mokken package version 3.0.6 (van der Ark, 2007, 2012) in R version 4.0.3 (R Core Team, 2020). With respect to scalability, the following rules of thumb are generally followed for interpreting scale-level *H* coefficients: scalability is considered weak if $0.30 \leq H < 0.40$; moderate, if $0.40 \leq H < 0.50$; and strong, if $H \geq 0.50$. The predicates 'weak,' 'moderate,' and 'strong' refer to the degree to which the ordering of individuals by test score accurately reflects their ordering on the latent variable. Item-level *H* coefficients

(H_is) in a unidimensional scale should be >0.30. Pairwise *H* coefficients should be >0. The standard error (*SE*) of the scalability coefficients needs to be taken into account in the scalability coefficients' interpretation (van der Ark, 2012). For instance, an *H* of, say, 0.52 with a *SE* of, say, 0.05 should be interpreted to reflect moderate, rather than strong, scalability. We additionally explored the ODI's scalability relying on the Automated Item Selection Procedure (AISP), which partitions a set of items into one or more Mokken scales if the data allow (items can be left unscalable). The AISP employs user-defined thresholds of homogeneity based on the scale *H*. As recommended (e.g., Stochl et al., 2012), we explored thresholds ranging from 0.30 to 0.55 (a high threshold), in increments of 0.05.

Monotonicity violations were analysed in terms of their presence, statistical significance, and seriousness by means of the *crit* statistic. It has been suggested that items for which the *crit* statistic falls below 80 can be viewed as *not* seriously violating monotonicity requirements and, as a result, can be safely included in a Mokken scale (e.g., van Schuur, 2003). Invariant item ordering was investigated based on the Manifest IIO (MIIO) method and the H^T coefficient. The MIIO method allows investigators to identify IIO violations and also employs the *crit* statistic to estimate the magnitude of the violations. The coefficient H^T represents the degree to which respondents order items invariantly. Assuming that MIIO holds, H^T should reach at least 0.30 for IIO to be considered sufficiently accurate (Ligtvoet et al., 2010).

6 | RESULTS

6.1 | ESEM bifactor analysis

Our bifactor model showed a satisfactory fit in the three samples (Table 1). All ODI items loaded more strongly on the general factor than on any of the two bifactors (Figures 1–3). Although the bifactors were weak, they did not entirely collapse. The ECV statistics revealed that the general factor accounted for 80%–90% of the common variance extracted depending on the sample and rotation considered (Table 1). The ECV statistics and mean loadings on the general factor were thus consistently indicative of essential unidimensionality.

TABLE 1 Exploratory structural equation modelling bifactor analysis of the ODI: fit statistics and explained common variance

Sample	Sample size	χ^2 (df)	RMSEA	RMSEA 90% CI	CFI	TLI	SRMR	ML-GF$_{PSTR}$	ML-GF$_{BI-GEO}$	ECV$_{PSTR}$	ECV$_{BI-GEO}$	ωH$_{PSTR}$	ωH$_{BI-GEO}$
French	3454	49.142 (12)	0.030	0.022, 0.039	0.999	0.997	0.010	0.751	0.765	0.875	0.903	0.859	0.880
Swiss	1971	57.636 (12)	0.044	0.033, 0.056	0.997	0.991	0.012	0.720	0.751	0.799	0.863	0.814	0.853
Australian	1485	21.839 (12)	0.023	0.005, 0.039	0.999	0.998	0.007	0.774	0.768	0.878	0.868	0.877	0.878

Note: The rotation chosen does not affect model fit.

Abbreviations: CFI, Comparative Fit Index; d*f*, degrees of freedom; ECV, Explained Common Variance; ECV$_{BI-GEO}$, ECV when using a bi-geomin rotation; ECV$_{PSTR}$, ECV when using a partially specified target rotation; ML-GF, mean loading on the general factor; ML-GF$_{BI-GEO}$, ML-GF with a bi-geomin rotation; ML-GF$_{PSTR}$, ML-GF with a partially specified target rotation; RMSEA, Root Mean Square Error of Approximation; SRMR, Standardized Root Mean Squared Residual; TLI, Tucker-Lewis Index; ωH$_{BI-GEO}$, omega hierarchical using a bi-geomin rotation; ωH$_{PSTR}$, omega hierarchical using a partially specified target.
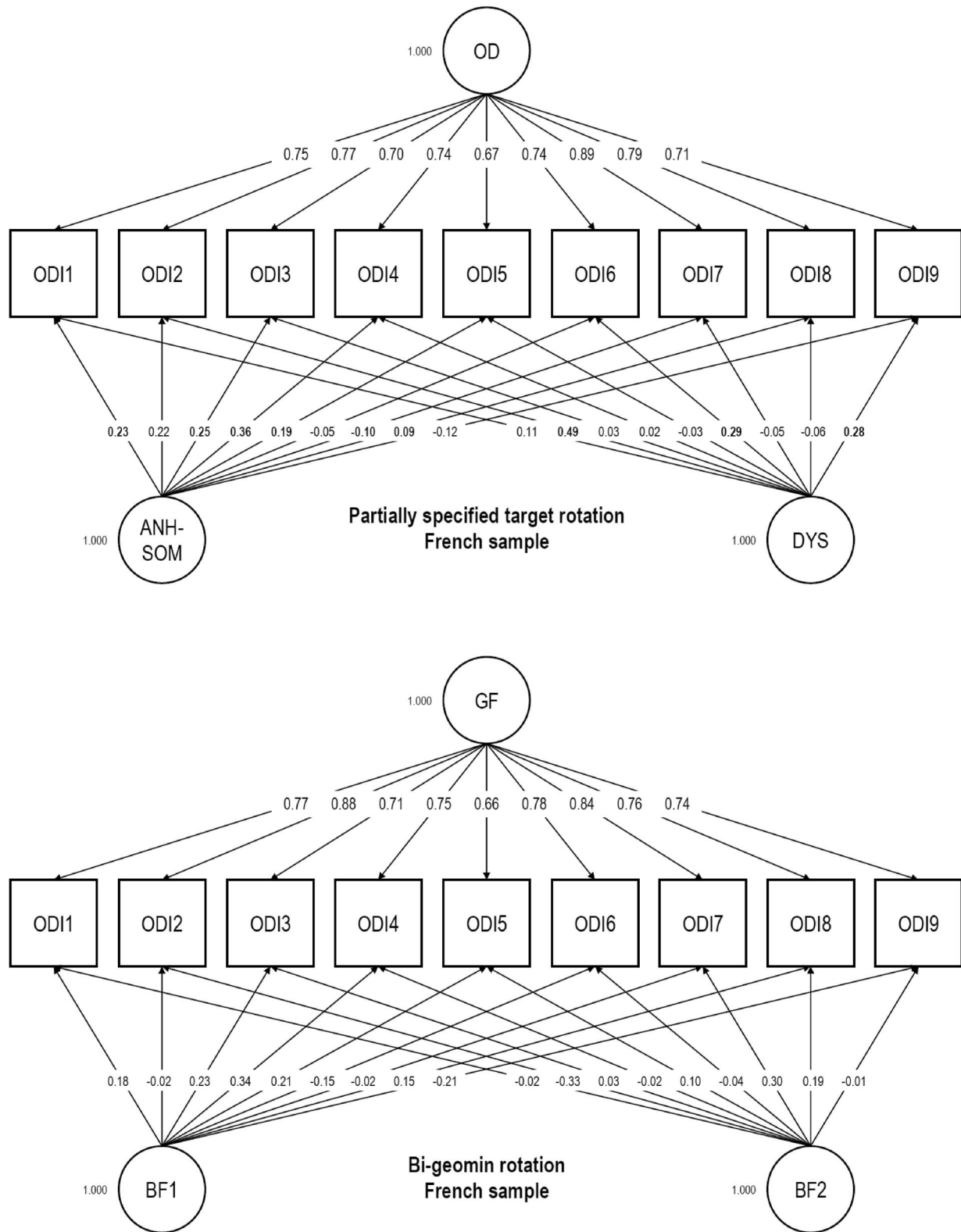
FIGURE 1 Exploratory structural equation modelling bifactor analyses of the Occupational Depression Inventory in the French sample (*N* = 3454): Factor loadings (GF: general factor; BF1: first bifactor; BF2: second bifactor; OD: Occupational Depression factor; ANH-SOM: Anhedonic-Somatic factor; DYS: Dysphoric factor). Target loadings are bolded
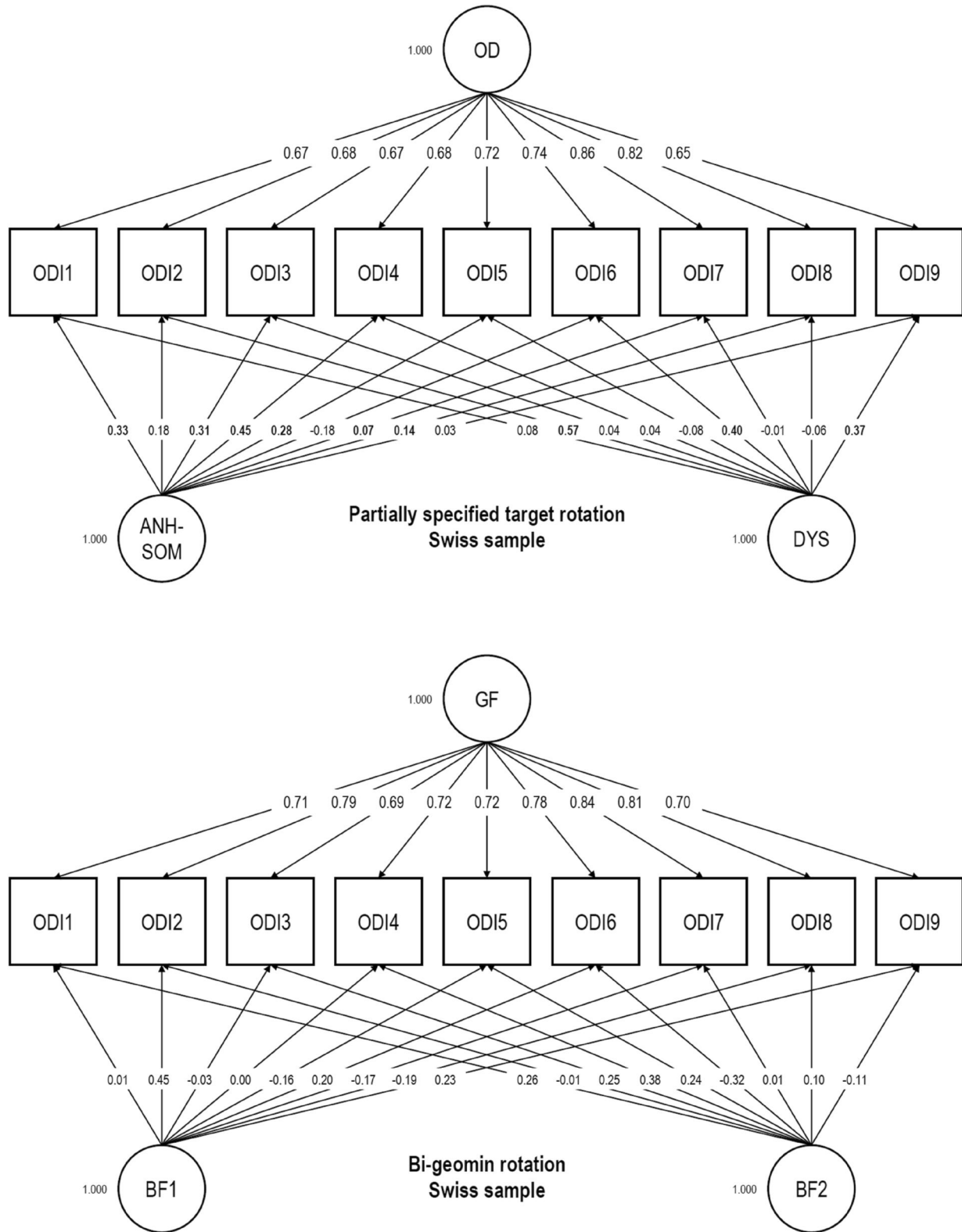
FIGURE 2 Exploratory structural equation modelling bifactor analyses of the Occupational Depression Inventory in the Swiss sample (*N* = 1971): Factor loadings (GF: general factor; BF1: first bifactor; BF2: second bifactor; OD: Occupational Depression factor; ANH-SOM: Anhedonic-Somatic factor; DYS: Dysphoric factor). Target loadings are bolded

FIGURE 3 Exploratory structural equation modelling bifactor analyses of the Occupational Depression Inventory in the Australian sample (N = 1485): Factor loadings (GF: general factor; BF1: first bifactor; BF2: second bifactor; OD: Occupational Depression factor; ANH-SOM: Anhedonic-Somatic factor; DYS: Dysphoric factor). Target loadings are bolded
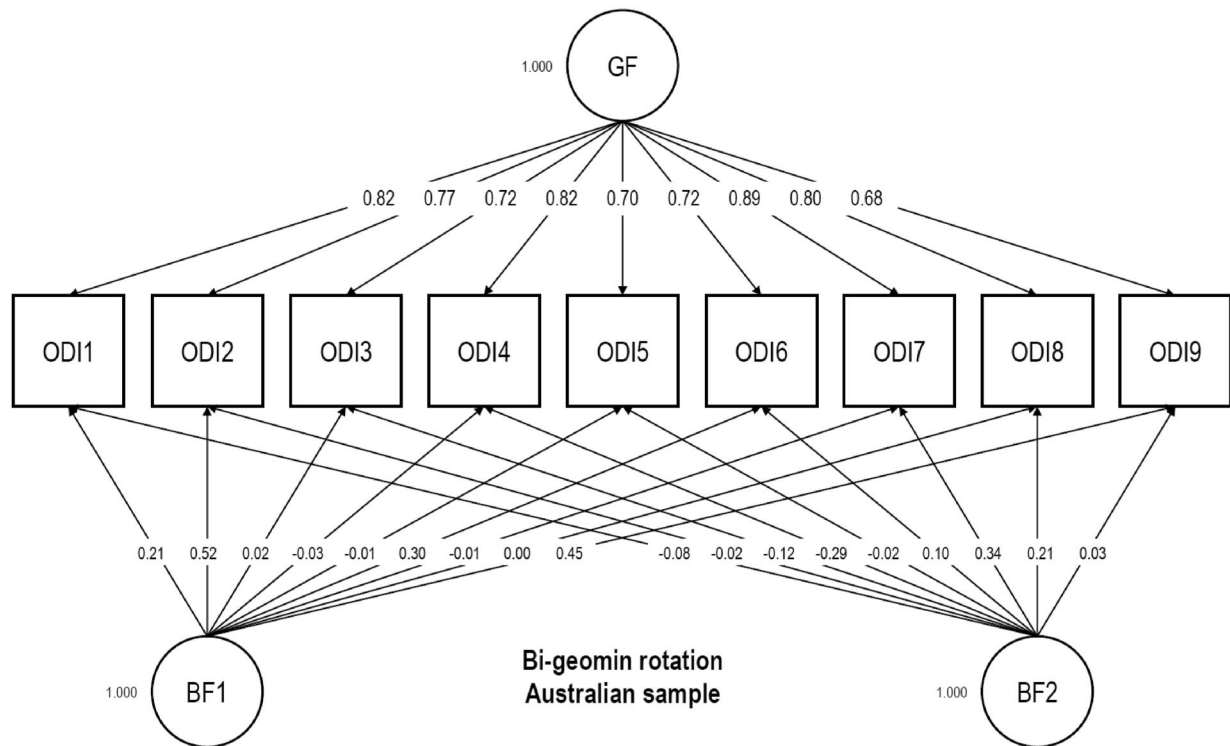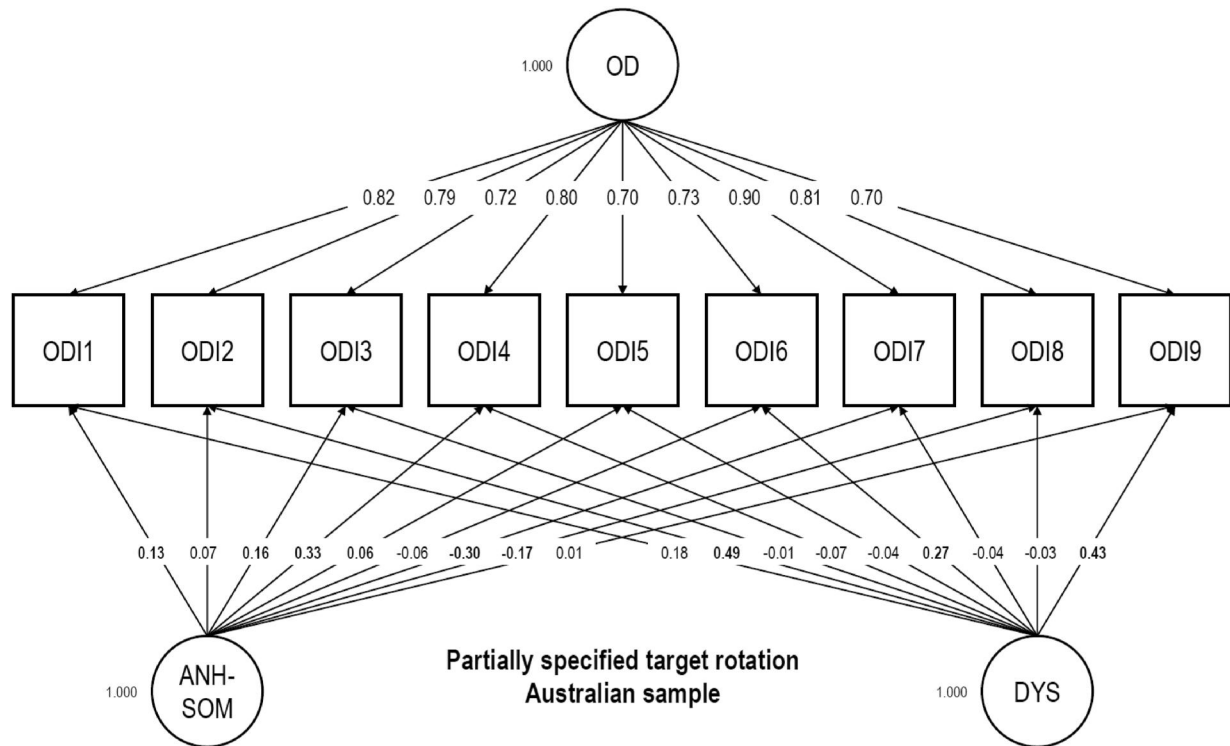
All ωH coefficients were high (Table 1). In the French sample, ωH was 0.880 when a bi-geomin rotation was used, and 0.859 with the PSTR. In the Swiss sample, ωH was 0.853 when a bi-geomin rotation was used, and 0.814 with the PSTR. In the Australian sample, ωH was 0.878 when a bi-geomin rotation was used, and 0.877 with the PSTR. We computed the correlations between the general factor and the observed total scores based on the square roots of the ωH coefficients (Rodriguez et al., 2016). The correlations were large, ranging from 0.902 to 0.938. As recommended by Rodriguez et al. (2016), we also compared the ωH and ω coefficients derived from the bifactor model. ω coefficients were 0.942 (bi-geomin) and 0.941 (target) in the French sample, 0.943 (bi-geomin) and 0.940 (target) in the Swiss sample, and 0.949 (bi-geomin) and 0.950 (target) in the Australian sample. Dividing ωH by ω, we found that most of the reliable variance in the observed total scores could be attributed to the general factor in the French sample (91%–93%), the Swiss sample (87%–91%), and the Australian sample (92%).

6.2 | Measurement invariance

Examining the measurement invariance of a unidimensional model, we found that measurement invariance held across sexes and age groups in the French, Swiss, and Australian samples (see Table S2a). ΔCFI and ΔSRMR were all very small. Changes in absolute values did not exceed 0.003. Measurement invariance was also likely across the six samples submitted to examination (our French, Swiss,

and Australian samples and the three original samples from the ODI's initial validation study by Bianchi and Schonfeld (2020)), with (a) ΔCFI of −0.004 and −0.003 for metric invariance and scalar invariance, respectively, and (b) ΔSRMR of 0.003 and 0.002 for metric invariance and scalar invariance, respectively (see Table S2b).[5]

6.3 | MSA

Scalability results are summarised in Table 2. The scale-level H coefficients were indicative of a strong scale in the French, the Swiss, and the Australian sample, with values of 0.57 ($SE = 0.01$), 0.54 ($SE = 0.01$), and 0.59 ($SE = 0.01$), respectively. H_is were similarly high, ranging from 0.51 (Item 5 [appetite alterations]) to 0.61 (Item 2 [depressed mood]) in the French sample, 0.49 (Items 6 [feelings of worthlessness] and 9 [suicidal ideation]) to 0.58 (Item 7 [cognitive impairment]) in the Swiss sample, and 0.53 (Item 5 [appetite alterations]) to 0.64 (Items 1 [anhedonia] and 7 [cognitive impairment]) in the Australian sample. All SEs were small ($0.01 \leq SE$s ≤ 0.03) and did not lead us to reinterpret the strength of the scale in any sample. No pairwise H was low (Table S3a–c). In all three samples, the AISP signalled a single scale up to a threshold of 0.50. At the highest threshold tested, that is, 0.55, item 5 appeared to be unscalable in the French sample as well as in the Australian sample. In the Swiss sample, two scales emerged at the threshold of 0.55. One scale corresponded to the ODI's anhedonic-somatic symptom items; the

TABLE 2 Mokken scale properties of the occupational depression inventory: scalability, reliability, and monotonicity

| | French sample (N = 3454) | | | | | | Swiss sample (N = 1971) | | | | | | Australian sample (N = 1485) | | | | | |
| | Scalability and reliability | | | Monotonicity | | | Scalability and reliability | | | Monotonicity | | | Scalability and reliability | | | Monotonicity | | |
	H_i	SE	95% CI	#vi	#zsig	crit	H_i	SE	95% CI	#vi	#zsig	crit	H_i	SE	95% CI	#vi	#zsig	crit
ODI1	0.59	0.01	[0.57, 0.61]	0	0	0	0.54	0.02	[0.51, 0.57]	0	0	0	0.64	0.01	[0.61, 0.66]	0	0	0
ODI2	0.61	0.01	[0.60, 0.63]	0	0	0	0.55	0.02	[0.52, 0.58]	0	0	0	0.63	0.01	[0.61, 0.65]	0	0	0
ODI3	0.56	0.01	[0.54, 0.58]	0	0	0	0.54	0.02	[0.51, 0.56]	0	0	0	0.55	0.02	[0.52, 0.58]	0	0	0
ODI4	0.60	0.01	[0.58, 0.61]	0	0	0	0.57	0.01	[0.54, 0.59]	0	0	0	0.58	0.02	[0.55, 0.61]	0	0	0
ODI5	0.51	0.01	[0.49, 0.53]	0	0	0	0.51	0.02	[0.48, 0.55]	0	0	0	0.53	0.02	[0.49, 0.56]	1	0	2
ODI6	0.55	0.01	[0.53, 0.57]	0	0	0	0.49	0.02	[0.46, 0.53]	0	0	0	0.57	0.01	[0.54, 0.60]	0	0	0
ODI7	0.60	0.01	[0.59, 0.62]	0	0	0	0.58	0.01	[0.55, 0.60]	0	0	0	0.64	0.01	[0.61, 0.66]	0	0	0
ODI8	0.57	0.01	[0.55, 0.59]	0	0	0	0.56	0.02	[0.53, 0.59]	0	0	0	0.59	0.01	[0.56, 0.61]	0	0	0
ODI9	0.59	0.02	[0.55, 0.63]	0	0	0	0.49	0.03	[0.44, 0.55]	0	0	0	0.62	0.02	[0.59, 0.65]	0	0	0
H	0.57	0.01	[0.56, 0.59]				0.54	0.01	[0.52, 0.56]				0.59	0.01	[0.57, 0.62]			
λ₂	0.90						0.89						0.90					
MS	0.90						0.89						0.91					
H^T	0.56						0.38						0.52					

Note: H: scale-level H; H_i: item-level H; H^T: invariant item ordering; λ₂: Guttman's lambda-2; MS: Molenaar-Sijtsma statistic; *SE*: standard error; 95% CI: 95% confidence interval; #vi: violations; #zsig: significant violations; *crit*: serious violations—items for which the *crit* statistic reaches 80 seriously violate requirements. ODI1: anhedonia; ODI2: depressed mood; ODI3: sleep alterations; ODI4: fatigue/loss of energy; ODI5: appetite alterations; ODI6: feelings of worthlessness; ODI7: cognitive impairment; ODI8: psychomotor alterations; ODI9: suicidal ideation.

other, to the ODI's dysphoric symptom items. A threshold of 0.55 is, however, a very high threshold.

We did not observe any statistically significant or serious violation of monotonicity for any of the ODI items (see Table 2; see also Figures S1–S3). In other words, all ODI items discriminated well between respondents at higher and lower levels of the latent variable in all three samples. Results pertaining to IIO are summarised in Table S3d (see also Figures S4–S6 for graphical illustrations). We found a couple of statistically significant IIO violations in the French and Swiss samples; however, none of them was serious, as reflected in crit indices far below 80. No IIO violation was detected in the Australian sample. Consistent with these observations, overall IIO was found to be sufficiently accurate in all three samples, with H^T values ranging 0.38–0.56.

Items were ordered in the same manner in the French and Swiss samples. In all three samples, the easiest items were Item 4—fatigue/loss of energy—and Item 3—sleep alterations—and the most difficult item was Item 9—suicidal ideation. Variations in item ordering were found between the French and Swiss samples on the one hand, and the Australian sample on the other hand, for the six remaining items of the ODI. As an illustration, the second most difficult item was Item 6 (feelings of worthlessness) in the French and Swiss samples and Item 2 (depressed mood) in the Australian sample (see Table S3d for a more comprehensive view).

Total score reliability, as indexed by Guttman's lambda-2 (λ_2) and the Molenaar-Sijtsma statistic (MS), was excellent (Table 2). In the French sample, Guttman's λ_2 and the MS each had a value of 0.90. In the Swiss sample, Guttman's λ_2 and the MS were both equal to 0.89. In the Australian sample, Guttman's λ_2 index was 0.90 and the MS, 0.91.

7 | DISCUSSION

The ODI was recently developed to improve the way job-related distress is assessed and managed. In this study, we conducted stringent analyses of the psychometric and structural properties of the ODI by employing ESEM bifactor analysis and MSA—a probabilistic, nonparametric IRT method. We relied on three samples of employed individuals (combined N = 6910) living in three different countries—Australia, France, and Switzerland—and speaking two different languages—English and French.

7.1 | Main findings

The ODI met the requirements for essential unidimensionality, scalability, monotonicity, and IIO and exhibited optimal total-score reliability. The ODI demonstrated Mokken scale analytic properties that are at least as strong as those of depression scales such as the PHQ-9 and stronger than those of (a) depression scales such as the Centre for Epidemiological Studies-Depression scale or the Montgomery-Åsberg Depression Rating Scale (e.g., Adler et al., 2012; Boothroyd

et al., 2019; Kliem et al., 2020) and (b) burnout scales such as the MBI (Jordan et al., 2018; see also; González-Romá et al., 2006). A unidimensional model showed measurement invariance across sex and age groups in each of our three samples (Table S2a), suggesting that the ODI had a largely similar structure, or meaning, across the groups under examination.

Our results are consistent with those of Bianchi and Schonfeld's (2020) initial validation study of the ODI, in which the ODI displayed strong alpha and omega reliabilities, high factorial validity, and essential unidimensionality. We found a similar unidimensional model to hold across our French, Swiss, and Australian samples and the three samples used in Bianchi and Schonfeld's (2020) initial ODI validation study (Table S2b). These findings provide support for the ODI's between-sample comparability across countries and languages. Measurement invariance across samples will need to be re-assessed as new ODI studies are conducted.

We note that between-item associations in self-report measures (and indicators linked to them, such as alpha reliability) are sometimes inflated because of the use of items having highly similar, virtually interchangeable content (Boyle, 1991; Cho & Kim, 2015). As an illustration, the MBI items 'Working with people all day is really a strain for me' and 'Working with people directly puts too much stress on me' are largely synonymous; the same applies to items of other burnout scales, such as the Shirom-Melamed Burnout Measure (SMBM; e.g., 'I have difficulty concentrating'; 'I feel I am not thinking clearly'; 'I feel I am not focussed on my thinking'). The amount of synonymy in burnout items gives burnout scales the patina of reliability, making them more attractive to researchers and practitioners than is warranted. Such an inflation phenomenon is improbable in the case of the ODI. Each of the nine core items of the ODI focuses on a specific symptom of major depression as characterised by the DSM-5 (American Psychiatric Associations, 2013), and the wording of the items shows no explicit redundancy. The associations between the items are likely to reflect genuine, clinically meaningful links among the symptoms of interest and not item-level content synonymy.

In all three samples, we found the easiest (i.e., most commonly endorsed) item to be Item 4, referring to fatigue/loss of energy, and the most difficult (i.e., least commonly endorsed) item to be Item 9, referring to suicidal ideation. Item 9 thus acted as a sentinel item. If someone were to endorse suicidal ideation, there is a high probability that the individual would have endorsed a host of other depressive symptoms. It is reasonable to expect that (a) the item referring to fatigue/loss of energy will consistently exhibit the lowest degree of difficulty, and (b) the item referring to suicidal ideation will consistently exhibit the highest degree of difficulty across ODI studies. Such expectations are reasonable because fatigue/loss of energy, though elevated in individuals standing at the higher end of the depression continuum, is fairly common in individuals at the lower end of the depression continuum. By contrast, suicidal ideation is extremely rare in individuals standing at the lower end of the depression continuum and more common in individuals at the higher end of the depression continuum. Whether a stable hierarchy of item

difficulties across samples/cultures can realistically be expected for *all* ODI items is left to be clarified (Meijer & Egberink, 2012).

7.2 | Contrasting the ODI and burnout measures

Given the popularity of the burnout construct in research on job-related distress, the potential added value of the ODI with reference to measures of burnout is worth examining. As previously mentioned, relying on the burnout construct is complicated because (a) the burnout syndrome remains loosely defined and undiagnosable, (b) the conceptualisation and operationalisation of burnout reflected in the MBI—the main measure of burnout—do not align, and (c) the MBI was developed inductively, without firm theoretical or clinical foundations (Bianchi et al., 2021; Heinemann & Heinemann, 2017; Schwenk & Gold, 2018; Vinkers & Schaafsma, 2021). Additional complications can be considered.

As underscored by Schaufeli and Enzmann (1998), the MBI has a narrow symptom scope. The MBI does not assess key aspects of job-related distress, including crucially informative symptoms such as suicidal thoughts (Bianchi et al., 2021).[6] That an employee contemplates suicide is a fact occupational physicians and occupational health psychologists (among other agents) want to be aware of. Furthermore, the psychometric and structural properties of the MBI have been repeatedly called into question, for instance, at the levels of factorial validity and reliability (de Beer & Bianchi, 2019; Shoman et al., 2021; Wheeler et al., 2011).[7] Numerous findings indicate that the three-factor model assumed to underlie the MBI does not fit well; to reach an acceptable fit, investigators generally have to engage in bricolage (e.g., item removal) and (questionable) model re-specification. Although alternative measures of burnout, such as the SMBM (Shirom & Melamed, 2006) and the Copenhagen Burnout Inventory (Kristensen et al., 2005), may be stronger from a psychometric and structural standpoint, their symptom scope is, unfortunately, even narrower than that of the MBI, with an almost exclusive focus on fatigue (Schaufeli & Taris, 2005). Such instruments add to burnout's definitional heterogeneity while leaving the issue of burnout's diagnosis unresolved.

The absence of diagnostic criteria for burnout has significant implications.[8] A sound diagnosis is a prerequisite for proper treatments and interventions—regardless of the level(s) at which treatments and interventions are envisaged (e.g., individual, organizational). Consistent with this observation, attempted treatments and interventions for burnout have produced disappointing outcomes to date (Ahola et al., 2017; Dreison et al., 2018). A sound diagnosis is also a prerequisite for proper prevalence estimations—a critical issue for public health decision-makers (Tyssen, 2018; Vinkers & Schaafsma, 2021). Interestingly, studies aiming to estimate the prevalence of burnout have flourished in recent years, despite reservations about the very idea of estimating the prevalence of an undiagnosable entity (Bianchi et al., 2017b; Mirkovic & Bianchi, 2019; Rotenstein et al., 2018; Schears, 2017). Such studies have typically relied on clinically and theoretically arbitrary identification criteria, leading to essentially uninterpretable estimates (Bianchi, 2017; Hewitt et al., 2020; Rotenstein et al., 2018; Schonfeld et al., 2018). As previously noted, Rotenstein et al. (2018) identified no fewer than 142 different characterisations of physician burnout in the literature, with prevalence estimates varying from 0% to 90%. In a national study of nearly 7000 health professionals, Hewitt et al. (2020) found burnout's prevalence estimates to vary from 3% to 91% depending on how burnout was defined (see also Prins et al., 2007). How public health policymakers can use these estimates to guide their decisions is unclear.

Another area of investigation affected by the impossibility of diagnosing burnout is that of the biology of burnout. In a systematic review carried out by Danhof-Pont et al. (2011), '[n]o potential biomarkers for burnout were found, largely due to the incomparability of studies' (p. 505). Other investigators have arrived at similar conclusions (e.g., Strikweda et al., 2021). Bianchi et al. (2017a) underscored the paradoxical situation in which research on the biology of burnout appears to stand: '…while biologically-driven research is intended to better characterise burnout, burnout would need to be better characterised in order for biologically-driven research to progress' (p. 239).

The problems affecting the burnout construct are profound and progress in their resolution has been underwhelming. Anchored in the area of stress and depression research, the ODI offers occupational health specialists interested in job-related distress a promising way forward (Bianchi & Schonfeld, 2021b; Bianchi & Sowden, 2022). While the caveats related to self-report measures inevitably apply to the ODI, the ODI responds encouragingly to the limitations attached to the burnout construct and its measures. The evidence provided in this study suggests that (a) the psychometric and structural properties of the ODI are particularly robust and (b) researchers and practitioners are justified in employing the ODI as an integrated measure of work-attributed depressive symptoms based on the scale's total score. That the ODI demonstrates essential unidimensionality despite covering nine different symptoms (including suicidal ideation) speaks to the unity of the phenomenon of (occupational) depression.

From a practical standpoint, the ODI is brief, straightforward in its use (e.g., its conceptualisation and operationalisation align), and available to occupational health specialists at no cost. In addition, the dual (dimensional-categorical) lens of the ODI may enable investigators to identify individuals with clinically-relevant levels of job-related distress more effectively and estimate the prevalence of the phenomenon more accurately. Approaching job-related distress within the sphere of depression may also pave the way for better protection of workers in terms of sick pay and provide health insurance companies with a clearer basis for defining their coverage (e.g., regarding treatment payment). The use of the ODI does not exempt occupational health specialists from conducting further etiological inquiries (e.g., based on peer reports or 'objective' indicators of working conditions and health status). Patient-provided information plays a key role in diagnostic, etiological, and treatment-related (e.g., treatment efficacy and side effects) investigations. In many cases, no one has access to more or better

information than patients themselves (Baldwin, 2000; Spector, 2019). However, corroborating patient-provided information with evidence from other sources is advisable. Furthermore, the ODI makes no presupposition regarding the extent to which internal dispositions (e.g., personal incompetence) or external conditions (e.g., management styles) account for the emergence of the symptoms assessed (Bianchi & Schonfeld, 2020). Through the causal attributions that it incorporates, the ODI points to *work* as a driving depressogenic factor. It is then incumbent on occupational health specialists to identify the *specific* job-related factors that may give rise to the symptoms.

Interestingly, addressing the issue of job-related distress within the sphere of depression may allow researchers and practitioners to close a historical loop. In a diachronic examination of 'exhaustion-related conditions', and of the narratives surrounding them, Schaffner (2016) concluded that entities such as nervous weakness (described in the 18th century), neurasthenia (described in the 19th century), and burnout (described in the 20th century) could be regarded as successive avatars of *melancholic weariness*, a depressive condition already described by physicians of the Hellenic and Roman eras. This view dovetails with contemporary research findings demonstrating that burnout's discriminant validity vis-à-vis depression is problematic (Ahola et al., 2014; Bianchi, Rolland, et al., 2018, 2021; Hodson, 2021; Sowden et al., 2022; Verkuilen et al., 2021).

7.3 | Limitations and strengths

The present study has at least two limitations. First, the study's external validity is limited by the use of convenience sampling. Studies involving samples representative of specific occupational groups and national working populations are needed to further clarify the generalisability of our findings. Second, we could not investigate some properties of the ODI, such as its test-retest reliability. Multiple-wave, longitudinal studies should be conducted to examine such characteristics. We note that depression scales having comparable architectures have shown high test-retest reliability (e.g., Kroenke et al., 2001).

As for its strengths, the conducted study involved three large samples of participants (*N*s ranging from 1485 to 3454) from three different countries (Australia, France, and Switzerland) and relied on advanced statistical techniques rooted in both classical test theory and IRT. The features of this study resonate with Cortina et al.'s (2020) call for thorough examinations of the psychometric and structural properties of the measures employed in applied psychology. The present study tested the ODI's properties against demanding standards. To our knowledge, few measures of distress have undergone such a stringent examination, and even fewer have passed it. Finally, this study widens the panel of national and cultural contexts in which the ODI demonstrates its viability.

Our study closely scrutinised the psychometric and structural properties of the ODI but did not address the instrument's clinical and diagnostic properties. Establishing the ODI's sensitivity (i.e., the probability that workers whom the ODI identifies as depressed are truly suffering from the condition), specificity (i.e., the probability that workers whom the ODI identifies as *not* depressed are truly *not* suffering from the condition), and predictive value (both positive and negative) will be important to further evaluate the instrument's potential as a screen and decision-guiding tool.

8 | CONCLUSION

This study suggests, with an unprecedented level of confidence, that the ODI has excellent psychometric and structural properties and can be used as a unidimensional measure of occupational depression based on the scale's total score. Depressive symptoms constitute basic responses to situations of unresolvable stress and are well-identified predictors of work impairment, shortened worklife expectancy, and suicide (Amagasa et al., 2005; Brown et al., 2000; Chesney et al., 2014; Howard et al., 2021; Pedersen et al., 2019; Probert-Lindström et al., 2020; Simon et al., 2013; Willner et al., 2013). Because the work domain represents a major source of life stress (American Psychological Association, 2015), assessing work-related depressive symptoms reliably and validly is vital. The balance of evidence from the present study indicates that occupational health specialists can confidently employ the ODI. In view of its characteristics and anchorage in the area of stress and depression research, the ODI offers researchers and practitioners a robust alternative to existing measures of job-related distress.

ORCID
Renzo Bianchi https://orcid.org/0000-0003-2336-0407

ENDNOTES
[1] Diagnoses are deemed *provisional* because a formal diagnosis of depression requires the use of clinician-driven standardized interviews.

[2] Many researchers consider, for instance, that cynicism and inefficacy are not components of burnout (Kristensen et al., 2005; Shirom, 2005).

[3] Unidimensionality means that only one latent factor is necessary to account for the inter-item associations in the empirical data.

[4] Although we examined IIO in this study, the issue of whether the IIO assumption is realistic for health and clinical scales is debated (Meijer & Egberink, 2012).

[5] Although the use of RMSEA was not recommended in our analytic context (Shi et al., 2020, 2021), we note for the reader's information that RMSEA never increased across our various analyses of

measurement invariance, consistent with the conclusion that measurement invariance applied (Putnick & Bornstein, 2016; Rutkowski & Svetina, 2014).

[6] As a reminder, there is no iatrogenic risk in assessing suicidality (DeCou & Schumann, 2018).

[7] In an ancillary examination of 734 U.S. workers (for a description of the sample, see Schonfeld et al.'s (2019) article), we reinspected the factorial structure of the MBI. We relied on confirmatory factor analysis in Mplus 8.6 using the weighted least squares -- mean and variance adjusted -- estimator. The three-factor model assumed to characterize the MBI (Maslach et al., 2016) did not fit the data well: $RMSEA = 0.122$; $CFI = 0.918$; $TLI = 0.908$; $SRMR = 0.091$; χ^2 (206) = 2444.263. Unsurprisingly, a second-order model involving an overarching Burnout factor had a similarly poor fit. We also considered separate one-factor models of each of the three item sets that the MBI comprises. None showed an acceptable fit. Because these analyses concern the original, English version of the MBI, potential translation-related issues can be ruled out.

[8] Being able to identify individuals with burnout is important whether one prefers to designate burnout as a *syndrome*, an *ailment*, an *illness*, a *disease*, or a *disorder*—a point that has generated misunderstandings (Bianchi, Schonfeld, et al., 2018; Maslach & Leiter, 1997; Melnick et al., 2017).

REFERENCES

Adler, M., Hetta, J., Isacsson, G., & Brodin, U. (2012). An item response theory evaluation of three depression assessment instruments in a clinical sample. *BMC Medical Research Methodology*, 12, 84. https://doi.org/10.1186/1471-2288-12-84

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

Ahola, K., Toppinen-Tanner, S., & Seppänen, J. (2017). Interventions to alleviate burnout symptoms and to support return to work among employees with burnout: Systematic review and meta-analysis. *Burnout Research*, 4, 1–11. https://doi.org/10.1016/j.burn.2017.02.001

Amagasa, T., Nakayama, T., & Takahashi, Y. (2005). *Karojisatsu* in Japan: Characteristics of 22 cases of work-related suicide. *Journal of Occupational Health*, 47(2), 157–164. https://doi.org/10.1539/joh.47.157

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

American Psychological Association. (2015). *Stress in America™: Paying with our health*. Author.

American Psychological Association. (2021). A silent national crisis: Violence against teachers. https://www.apa.org/education/k12/teacher-victimization#

Baldwin, W. (2000). Information no one else knows: The value of self-report. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 3–7). Lawrence Erlbaum Associates Publishers.

Beck, A. T., & Alford, B. A. (2009). *Depression: Causes and treatment* (2nd ed.). University of Pennsylvania Press.

Bianchi, R. (2017). Is the "burnout epidemic" an academic fiction? *British Medical Journal*, 358, j4389. https://doi.org/10.1136/bmj.j4389

Bianchi, R. (2020). Do burnout and depressive symptoms form a single syndrome? Confirmatory factor analysis and exploratory structural equation modeling bifactor analysis. *Journal of Psychosomatic Research*, 131, 109954. https://doi.org/10.1016/j.jpsychores.2020.109954

Bianchi, R., Manzano-García, G., Montañés-Muro, P., Schonfeld, E. A., & Schonfeld, I. S. (2022). Occupational depression in a Spanish-speaking sample: Associations with cognitive performance and work-life characteristics. *Journal of Work and Organizational Psychology*, 38(1), 59–74. https://doi.org/10.5093/jwop2022a5

Bianchi, R., Rolland, J.-P., & Salgado, J. F. (2018a). Burnout, depression, and borderline personality: A 1, 163-participant study. *Frontiers in Psychology*, 8, 2336. https://doi.org/10.3389/fpsyg.2017.02336

Bianchi, R., Schonfeld, I., & Laurent, E. (2018b). Burnout syndrome and depression. In Y.-K. Kim (Ed.), *Understanding depression: Volume 2. Clinical manifestations, diagnosis and treatment* (pp. 187–202). Springer. https://doi.org/10.1007/978-981-10-6577-4_14

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

Bianchi, R., & Schonfeld, I. S. (2021a). Occupational depression, cognitive performance, and task appreciation: A study based on Raven's advanced progressive Matrices. *Frontiers in Psychology*, 12, 695539. https://doi.org/10.3389/fpsyg.2021.695539

Bianchi, R., & Schonfeld, I. S. (2021b). Who needs to be "burned-out"? Time for a new approach to job-related distress. *European Psychiatry*, 64(1), e65. https://doi.org/10.1192/j.eurpsy.2021.2240

Bianchi, R., & Schonfeld, I. S. (2022). Is the occupational depression inventory predictive of cognitive performance? A focus on inhibitory control and effortful reasoning. *Personality and Individual Differences*, 184, 111213. https://doi.org/10.1016/j.paid.2021.111213

Bianchi, R., Schonfeld, I. S., & Laurent, E. (2017a). Biological research on burnout-depression overlap: Long-standing limitations and on-going reflections. *Neuroscience & Biobehavioral Reviews*, 83, 238–239. https://doi.org/10.1016/j.neubiorev.2017.10.019

Bianchi, R., Schonfeld, I. S., & Laurent, E. (2017b). Can we trust burnout research? *Annals of Oncology*, 28(9), 2320–2321. https://doi.org/10.1093/annonc/mdx267

Bianchi, R., & Sowden, J. F. (2022). A paradigm shift from burnout to occupational depression. *Journal of Affective Disorders, 303*, 230–232. https://doi.org/10.1016/j.jad.2022.02.049

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

Boothroyd, L., Dagnan, D., & Muncer, S. (2019). PHQ-9: One factor or two? *Psychiatry Research*, 271, 532–534. https://doi.org/10.1016/j.psychres.2018.12.048

Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*, 12(3), 291–294. https://doi.org/10.1016/0191-8869(91)90115-R

Brown, G. K., Beck, A. T., Steer, R. A., & Grisham, J. R. (2000). Risk factors for suicide in psychiatric outpatients: A 20-year prospective study. *Journal of Consulting and Clinical Psychology*, 68(3), 371–377. https://doi.org/10.1037/0022-006x.68.3.371

Chesney, E., Goodwin, G. M., & Fazel, S. (2014). Risks of all-cause and suicide mortality in mental disorders: A meta-review. *World Psychiatry*, 13(2), 153–160. https://doi.org/10.1002/wps.20128

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods*, 18(2), 207–230. https://doi.org/10.1177/1094428114555994

Cortina, J. M., Sheng, Z., Keener, S. K., Keeler, K. R., Grubb, L. K., Schmitt, N., Tonidandel, S., Summerville, K. M., Heggestad, E. D., & Banks, G. C. (2020). From alpha to omega and beyond! A look at the past, present, and (possible) future of psychometric soundness in the Journal of Applied Psychology. *Journal of Applied Psychology*, 105(12), 1351–1381. https://doi.org/10.1037/apl0000815

Cox, T., Tisserand, M., & Taris, T. (2005). The conceptualization and measurement of burnout: Questions and directions. *Work & Stress*, *19*(3), 187–191. https://doi.org/10.1080/02678370500387109

Danhof-Pont, M. B., van Veen, T., & Zitman, F. G. (2011). Biomarkers in burnout: A systematic review. *Journal of Psychosomatic Research*, *70*(6), 505–524. https://doi.org/10.1016/j.jpsychores.2010.10.012

de Beer, L. T., & Bianchi, R. (2019). Confirmatory factor analysis of the Maslach burnout inventory: A Bayesian structural equation modeling approach. *European Journal of Psychological Assessment*, *35*(2), 217–224. https://doi.org/10.1027/1015-5759/a000392

DeCou, C. R., & Schumann, M. E. (2018). On the iatrogenic risk of assessing suicidality: A meta-analysis. *Suicide and Life-Threatening Behavior*, *48*(5), 531–543. https://doi.org/10.1111/sltb.12368

Dima, A. L. (2018). Scale validation in applied health research: Tutorial for a 6-step R-based psychometrics protocol. *Health Psychology and Behavioral Medicine*, *6*(1), 136–161. https://doi.org/10.1080/21642850.2018.1472602

Dohrenwend, B. P. (2000). The role of adversity and stress in psychopathology: Some evidence and its implications for theory and research. *Journal of Health and Social Behavior*, *41*(1), 1–19. https://doi.org/10.2307/2676357

Dreison, K. C., Luther, L., Bonfils, K. A., Sliter, M. T., McGrew, J. H., & Salyers, M. P. (2018). Job burnout in mental health providers: A meta-analysis of 35 years of intervention research. *Journal of Occupational Health Psychology*, *23*(1), 18–30. https://doi.org/10.1037/ocp0000047

European Agency for Safety and Health at Work. (2014). *Calculating the cost of work-related stress and psychosocial risks*. Author.

Galand, B., Lecocq, C., & Philippot, P. (2007). School violence and teacher professional disengagement. *British Journal of Educational Psychology*, *77*(2), 465–477. https://doi.org/10.1348/000709906X114571

Gonzalez-Mulé, E., & Cockburn, B. S. (2021). This job is (literally) killing me: A moderated-mediated model linking work characteristics to mortality. *Journal of Applied Psychology*, *106*(1), 140–151. https://doi.org/10.1037/apl0000501

González-Romá, V., Schaufeli, W. B., Bakker, A. B., & Lloret, S. (2006). Burnout and work engagement: Independent factors or opposite poles? *Journal of Vocational Behavior*, *68*(1), 165–174. https://doi.org/10.1016/j.jvb.2005.01.003

Grahek, I., Shenhav, A., Musslick, S., Krebs, R. M., & Koster, E. H. W. (2019). Motivation and cognitive control in depression. *Neuroscience & Biobehavioral Reviews*, *102*, 371–381. https://doi.org/10.1016/j.neubiorev.2019.04.011

Hassard, J., Teoh, K. R. H., Visockaite, G., Dewe, P., & Cox, T. (2018). The cost of work-related stress to society: A systematic review. *Journal of Occupational Health Psychology*, *23*(1), 1–17. https://doi.org/10.1037/ocp0000069

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

Hewitt, D. B., Ellis, R. J., Hu, Y.-Y., Cheung, E. O., Moskowitz, J. T., Agarwal, G., & Bilimoria, K. Y. (2020). Evaluating the association of multiple burnout definitions and thresholds with prevalence and outcomes. *JAMA Surgery*, *155*(11), 1043–1049. https://doi.org/10.1001/jamasurg.2020.3351

Hill, C., de Beer, L. T., & Bianchi, R. (2021). Validation and measurement invariance of the occupational depression inventory in South Africa. *PLoS One*, *16*(12), e0261271. https://doi.org/10.1371/journal.pone.0261271

Hodson, G. (2021). Construct jangle or construct mangle? Thinking straight about (nonredundant) psychological constructs. *Journal of Theoretical Social Psychology*(4), 576–590. Advance online publication. https://doi.org/10.1002/jts5.120

Howard, M. C., Follmer, K. B., Smith, M. B., Tucker, R. P., & Van Zandt, E. C. (2021). Work and suicide: An interdisciplinary systematic literature review. *Journal of Organizational Behavior*(2), 260–285. Advance online publication. https://doi.org/10.1002/job.2519

Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 166–184. https://doi.org/10.1177/2515245919882903

Ingersoll, R. M., Merrill, E., Stuckey, D., & Collins, G. (2018). Seven trends: The transformation of the teaching force – updated October 2018. CPRE Research Reports. In *Consortium for policy research in education*. University of Pennsylvania. https://repository.upenn.edu/cpre_researchreports/108

Jordan, P., Steingen, U., Terschüren, C., & Harth, V. (2018). The Maslach burnout inventory: A test dimensionality assessment via item response theory. *Testing, Psychometrics, Methodology in Applied Psychology*, *25*, 101–120. https://doi.org/10.4473/TPM25.1.6

Kliem, S., Beller, J., Tibubos, A. N., Zenger, M., Schmalbach, B., & Brähler, E. (2020). A reanalysis of the center for epidemiological studies-depression scale (CES-D) using non-parametric item response theory. *Psychiatry Research*, *290*, 113132. https://doi.org/10.1016/j.psychres.2020.113132

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.

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

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Ligtvoet, R., van der Ark, L. A., te Marvelde, J. M., & Sijtsma, K. (2010). Investigating an invariant item ordering for polytomously scored items. *Educational and Psychological Measurement*, *70*(4), 578–595. https://doi.org/10.1177/0013164409355697

Longobardi, C., Badenes-Ribera, L., Fabris, M. A., Martinez, A., & McMahon, S. D. (2019). Prevalence of student violence against teachers: A meta-analysis. *Psychology of Violence*, *9*(6), 596–610. https://doi.org/10.1037/vio0000202

Madsen, I. E. H., Nyberg, S. T., Magnusson Hanson, L. L., Ferrie, J. E., Ahola, K., Alfredsson, L., Batty, G. D., Bjorner, J. B., Borritz, M., Burr, H., Chastang, J.-F., de Graaf, R., Dragano, N., Hamer, M., Jokela, M., Knutsson, A., Koskenvuo, M., Koskinen, A., Leineweber, C., …, & Kivimäki, M. (2017). Job strain as a risk factor for clinical depression: Systematic review and meta-analysis with additional individual participant data. *Psychological Medicine*, *47*, 1342–1356. https://doi.org/10.1017/S003329171600355X

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

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., & Leiter, M. P. (2016). *Maslach burnout inventory manual* (4th ed.). Mind Garden.

Maslach, C., & Leiter, M. P. (1997). *The truth about burnout: How organizations cause personal stress and what to do about it*. Jossey-Bass.

Matthews, T. A., Robbins, W., Preisig, M., von Känel, R., & Li, J. (2021). Associations of job strain and family strain with risk of major depressive episode: A prospective cohort study in U.S. working men and women. *Journal of Psychosomatic Research*, *147*, 110541. https://doi.org/10.1016/j.jpsychores.2021.110541

McEwen, B. S. (2012). Brain on stress: How the social environment gets under the skin. *Proceedings of the National Academy of Sciences of the*

United States of America, *109*(Suppl. 2), 17180–17185. https://doi.org/10.1073/pnas.1121254109

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*. Advance online publication. https://doi.org/10.1037/ocp0000273

Meijer, R. R., & Baneke, J. J. (2004). Analyzing psychopathology items: A case for nonparametric item response theory modeling. *Psychological Methods*, *9*(3), 354–368. https://doi.org/10.1037/1082-989X.9.3.354

Meijer, R. R., & Egberink, I. J. L. (2012). Investigating invariant item ordering in personality and clinical scales: Some empirical findings and a discussion. *Educational and Psychological Measurement*, *72*(4), 589–607. https://doi.org/10.1177/0013164411429344

Melchior, M., Caspi, A., Milne, B. J., Danese, A., Poulton, R., & Moffitt, T. E. (2007). Work stress precipitates depression and anxiety in young, working women and men. *Psychological Medicine*, *37*(8), 1119–1129. https://doi.org/10.1017/S0033291707000414

Melnick, E. R., Powsner, S. M., & Shanafelt, T. D. (2017). In reply—Defining physician burnout, and differentiating between burnout and depression. *Mayo Clinic Proceedings*, *92*(9), 1456–1458. https://doi.org/10.1016/j.mayocp.2017.07.005

Mirkovic, D., & Bianchi, R. (2019). Physician burnout: Let's avoid unsubstantiated claims. *Nature Reviews Clinical Oncology*, *16*(2), 136. https://doi.org/10.1038/s41571-018-0150-8

Mokken, R. J. (1971). *A theory and procedure of scale analysis*. De Gruyter Mouton.

Molenaar, I. W. (1982). Mokken scaling revisited. *Kwantitatieve Methoden*, *3*(8), 145–164.

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

Muthén, L. K., & Muthén, B. O. (1998-2021). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Nieuwenhuijsen, K., Bruinvels, D., & Frings-Dresen, M. (2010). Psychosocial work environment and stress-related disorders, a systematic review. *Occupational Medicine*, *60*(4), 277–286. https://doi.org/10.1093/occmed/kqq081

Park, T.-Y., & Shaw, J. D. (2013). Turnover rates and organizational performance: A meta-analysis. *Journal of Applied Psychology*, *98*(2), 268–309. https://doi.org/10.1037/a0030723

Pedersen, J., Thorsen, S. V., Andersen, M. F., Hanvold, T. N., Schlünssen, V., & Bültmann, U. (2019). Impact of depressive symptoms on worklife expectancy: A longitudinal study on Danish employees. *Occupational and Environmental Medicine*, *76*(11), 838. https://doi.org/10.1136/oemed-2019-105961

Pickles, A., & Angold, A. (2003). Natural categories or fundamental dimensions: On carving nature at the joints and the rearticulation of psychopathology. *Development and Psychopathology*, *15*(3), 529–551. https://doi.org/10.1017/S0954579403000282

Prins, J. T., Gazendam-Donofrio, S. M., Tubben, B. J., van der Heijden, F. M., van de Wiel, H. B., & Hoekstra-Weebers, J. E. (2007). Burnout in medical residents: A review. *Medical Education*, *41*(8), 788–800. https://doi.org/10.1111/j.1365-2923.2007.02797.x

Probert-Lindström, S., Berge, J., Westrin, Å., Öjehagen, A., & Skogman Pavulans, K. (2020). Long-term risk factors for suicide in suicide attempters examined at a medical emergency in patient unit: Results from a 32-year follow-up study. *BMJ Open*, *10*(10), e038794. https://doi.org/10.1136/bmjopen-2020-038794

Pryce, C. R., Azzinnari, D., Spinelli, S., Seifritz, E., Tegethoff, M., & Meinlschmidt, G. (2011). Helplessness: A systematic translational review of theory and evidence for its relevance to understanding and treating depression. *Pharmacology & Therapeutics*, *132*(3), 242–267. https://doi.org/10.1016/j.pharmthera.2011.06.006

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

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reddy, L. A., Espelage, D., McMahon, S. D., Anderman, E. M., Lane, K. L., Brown, V. E., Reynolds, C. R., Jones, A., & Kanrich, J. (2013). Violence against teachers: Case studies from the APA task force. *International Journal of School & Educational Psychology*, *1*(4), 231–245. https://doi.org/10.1080/21683603.2013.837019

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*(1), 19–31. https://doi.org/10.1007/s11136-007-9183-7

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Rotenstein, L. S., Torre, M., Ramos, M. A., Rosales, R. C., Guille, C., Sen, S., & Mata, D. A. (2018). Prevalence of burnout among physicians: A systematic review. *JAMA*, *320*(11), 1131–1150. https://doi.org/10.1001/jama.2018.12777

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, *74*(1), 31–57. https://doi.org/10.1177/0013164413498257

Schaffner, A. K. (2016). Exhaustion and the pathologization of modernity. *Journal of Medical Humanities*, *37*(3), 327–341. https://doi.org/10.1007/s10912-014-9299-z

Schaufeli, W. B. (2003). Past performance and future perspectives of burnout research. *South African Journal of Industrial Psychology*, *29*(4), 1–15. https://doi.org/10.10520/EJC88982

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

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

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

Schears, R. M. (2017). Defining physician burnout, and differentiating between burnout and depression—II. *Mayo Clinic Proceedings*, *92*(9), 1455–1456. https://doi.org/10.1016/j.mayocp.2017.07.006

Schonfeld, I. S. (2001). Stress in 1st-year women teachers: The context of social support and coping. *Genetic, Social, and General Psychology Monographs*, *127*(2), 133–168.

Schonfeld, I. S. (2006). School violence. In E. K. Kelloway, J. Barling, & J. J. HurrellJr. (Eds.), *Handbook of workplace violence* (pp. 169–229). Sage Publications. https://doi.org/10.4135/9781412976947.n9

Schonfeld, I. S., & Bianchi, R. (2021). From burnout to occupational depression: Recent developments in research on job-related distress and occupational health. *Frontiers in Public Health*, *9*(1978), 796401. https://doi.org/10.3389/fpubh.2021.796401

Schonfeld, I. S., & Chang, C.-H. (2017). *Occupational health psychology: Work, stress, and health*. Springer Publishing Company.

Schonfeld, I. S., Laurent, E., & Bianchi, R. (2018). Burnout research: Eyes wide shut. *Critical Care Medicine*, *46*(2), e179–e180. https://doi.org/10.1097/ccm.0000000000002795

Schonfeld, I. S., Verkuilen, J., & Bianchi, R. (2019). Inquiry into the correlation between burnout and depression. *Journal of Occupational Health Psychology*, *24*(6), 603–616. https://doi.org/10.1037/ocp0000151

Schwenk, T. L., & Gold, K. J. (2018). Physician burnout—A serious symptom, but of what? *JAMA*, *320*(11), 1109–1110. https://doi.org/10.1001/jama.2018.11703

Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death*. W. H. Freeman.

Semmer, N. K., Jacobshagen, N., Meier, L. L., Elfering, A., Beehr, T. A., Kalin, W., & Tschan, F. (2015). Illegitimate tasks as a source of work stress. *Work & Stress*, *29*(1), 32–56. https://doi.org/10.1080/02678373.2014.1003996

Shi, D., DiStefano, C., Maydeu-Olivares, A., & Lee, T. (2021). *Evaluating SEM model fit with small degrees of freedom*. Multivariate Behavioral Research. Advance online publication. https://doi.org/10.1080/00273171.2020.1868965

Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2020). Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 1–15. https://doi.org/10.1080/10705511.2019.1611434

Shirom, A. (2005). Reflections on the study of burnout. *Work & Stress*, *19*(3), 263–270. https://doi.org/10.1080/02678370500376649

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

Shoman, Y., Marca, S. C., Bianchi, R., Godderis, L., van der Molen, H. F., & Guseva Canu, I. (2021). Psychometric properties of burnout measures: A systematic review. *Epidemiology and Psychiatric Sciences*, *30*, e8. https://doi.org/10.1017/s2045796020001134

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Sage. https://doi.org/10.4135/9781412984676

Simon, G. E., Rutter, C. M., Peterson, D., Oliver, M., Whiteside, U., Operskalski, B., & Ludman, E. J. (2013). Does response on the PHQ-9 Depression Questionnaire predict subsequent suicide attempt or suicide death? *Psychiatric Services*, *64*(12), 1195–1202. https://doi.org/10.1176/appi.ps.201200587

Sowden, J. F., Schonfeld, I. S., & Bianchi, R. (2022). Are Australian teachers burned-out or depressed? A confirmatory factor analytic study involving the occupational depression inventory. *Journal of Psychosomatic Research*, *157*, 110783. https://doi.org/10.1016/j.jpsychores.2022.110783

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

Stochl, J., Jones, P. B., & Croudace, T. J. (2012). Mokken scale analysis of mental health and well-being questionnaire item responses: A nonparametric IRT method in empirical research for applied health researchers. *BMC Medical Research Methodology*, *12*(1), 74. https://doi.org/10.1186/1471-2288-12-74

Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2014). Minimum sample size requirements for Mokken scale analysis. *Educational and Psychological Measurement*, *74*(5), 809–822. https://doi.org/10.1177/0013164414529793

Strikweda, M., Beulens, J. W., Remmelzwaal, S., Schoonmade, L. J., van Straten, A., Schram, M. T., Elders, P. J., & Rutters, F. (2021). The association of burnout and vital exhaustion with type 2 diabetes: A systematic review and meta-analysis. *Psychosomatic Medicine*, (9), 1013–1030. Advance online publication. https://doi.org/10.1097/psy.0000000000000995

Taris, T. W. (2006). Bricks without clay: On urban myths in occupational health psychology. *Work & Stress*, *20*(2), 99–104. https://doi.org/10.1080/02678370600893410

Tiesman, H. M., Konda, S., Hartley, D., Chaumont Menéndez, C., Ridenour, M., & Hendricks, S. (2015). Suicide in U.S. workplaces, 2003-2010: A comparison with non-workplace suicides. *American Journal of Preventive Medicine*, *48*(6), 674–682. https://doi.org/10.1016/j.amepre.2014.12.011

Tyssen, R. (2018). What is the level of burnout that impairs functioning? *Journal of Internal Medicine*, *283*(6), 594–596. https://doi.org/10.1111/joim.12769

U.S. Bureau of Labor Statistics. (2019). *National census of fatal occupational injuries in 2019*. Author.

van der Ark, L. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, *20*(11), 1–19. https://doi.org/10.18637/jss.v020.i11

van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software*, *48*(5), 1–27. https://doi.org/10.18637/jss.v048.i05

van Schuur, W. H. (2003). Mokken scale analysis: Between the Guttman scale and parametric item response theory. *Political Analysis*, *11*(2), 139–163. https://doi.org/10.1093/pan/mpg002

Verkuilen, J., Bianchi, R., Schonfeld, I. S., & Laurent, E. (2021). Burnout-depression overlap: Exploratory structural equation modeling bifactor analysis and network analysis. *Assessment*, *28*(6), 1583–1600. https://doi.org/10.1177/1073191120911095

Vinkers, C. H., & Schaafsma, F. G. (2021). Burnout urgently needs robust research. *Nature*, *592*(7853), 188. https://doi.org/10.1038/d41586-021-00896-1

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

Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, *44*(7), 1349–1360. https://doi.org/10.1017/S0033291713001979

Willner, P., Scheel-Krüger, J., & Belzung, C. (2013). The neurobiology of depression and antidepressant action. *Neuroscience & Biobehavioral Reviews*, *37*(10), 2331–2371. https://doi.org/10.1016/j.neubiorev.2012.12.007

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.