BRIEF REPORT

An Exploratory Structural Equation Modeling Bi-Factor Analytic Approach to Uncovering What Burnout, Depression, and Anxiety Scales Measure

Irvin Sam Schonfeld
The City College of the City University of New York, and the
Graduate Center of the City University of New York

Jay Verkuilen
The Graduate Center of the City University of New York

Renzo Bianchi
The University of Neuchâtel, Switzerland

In this study, we addressed the ongoing debate about what burnout and depression scales measure by conducting an exploratory structural equation modeling (ESEM) bifactor analysis. A sample of 734 U.S. teachers completed a survey that included the Center for Epidemiologic Studies Depression scale (CES-D-10), the depression module of the Patient Health Questionnaire (PHQ-9), the Generalized Anxiety Disorder scale (GAD-7), and the Maslach Burnout Inventory (MBI), which contains emotional exhaustion (EE), depersonalization (DP), and (diminished) personal accomplishment (PA) subscales. Job adversity and workplace support were additionally measured for the purpose of a nomological network analysis. EE, burnout's core, was more highly correlated with the depression and anxiety scales than it was with DP and PA, even with controls for item content overlap. The CES-D-10, PHQ-9, GAD-7, and EE subscale of the MBI were similarly related to job adversity and workplace support. ESEM bifactor analysis revealed that the CES-D-10, PHQ-9, GAD-7, and EE items loaded highly on a general factor, which we labeled nonspecific psychological distress (NSPD). We conclude that depression, anxiety, and EE scales reflect NSPD. DP items largely reflect two factors, NSPD and depersonalization, about equally. PA items were found to be less related to NSPD. With respect to the debate surrounding burnout-depression overlap, our findings do not support the view that the burnout construct represents a syndrome that consists of EE, DP, and diminished PA and excludes (or does not primarily include) depressive symptoms.

Public Significance Statement
This study indicates that the principal scale that assesses the core of burnout is most likely assessing a condition reflecting an anxio-depressive continuum. With this knowledge, we can better identify individuals suffering from "burnout" and direct them to the appropriate treatment.

Burnout has been defined as a work-induced syndrome combining emotional exhaustion (EE), depersonalization (DP), and a reduced sense of personal accomplishment (PA; Maslach, Jackson, & Leiter, 2016). EE refers to the feeling of being stressed out and drained of one's energy. EE is considered the core of burnout, with DP viewed as a way of coping with EE and diminished PA a

---

long-term repercussion of EE (Maslach, Schaufeli, & Leiter, 2001). DP consists of withdrawal from one's job and detachment from the people connected with one's work. Diminished PA involves a negative self-evaluation in the job, entailing feelings of incompetence and professional inefficacy. Thus defined, burnout is assessed with the Maslach Burnout Inventory (MBI), a self-administered questionnaire that has played a referential role in burnout research (Maslach et al., 2001). A problem attached to the MBI pertains to discriminant validity, most notably vis-à-vis measures of depression (Ahola, Hakanen, Perhoniemia, & Mutanen, 2014; Wurm et al., 2016). Scores on burnout and depression scales correlate highly enough to suggest that burnout and depression scales largely measure the same construct (Bianchi & Schonfeld, 2018; Schonfeld & Bianchi, 2016). The two sets of scales also showed similar nomological networks (Bianchi, Schonfeld, & Laurent, 2018). At an etiological level, unresolvable job stress, thought to be the prime mover in burnout (Maslach et al., 2001), is a key depressogenic factor (Bianchi et al., 2018; Schonfeld & Chang, 2017).

Other evidence indicates that burnout is associated with anxiety (e.g., Rössler, Hengartner, Ajdacic-Gross, & Angst, 2015). Depression and anxiety have been found to reflect a common continuum (Caspi et al., 2014). The view of a general anxio-depressive dimension of ill-being is well instantiated in the notions of demoralization (Tellegen et al., 2003) and nonspecific psychological distress (NSPD; Dohrenwend, Shrout, Egri, & Mendelsohn, 1980).

In this study, we examined relations among burnout, depression, and anxiety scales in two ways. First, we examined the scales' intercorrelations and looked for parallels in the scales' nomological networks. Second, to better understand how the scale items align dimensionally, we conducted an exploratory structural equation modeling (ESEM) bifactor analysis, focusing on the items of (a) the MBI's subscales; (b) two depression measures, the 9-item depression module of the Patient Health Questionnaire (PHQ-9; Kroenke & Spitzer, 2002) and the 10-item version of the Center for Epidemiologic Studies Depression scale (CES-D-10; Cole, Rabin, Smith, & Kaufman, 2004), and (c) the 7-item Generalized Anxiety Disorder scale (GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006).

Method

During the 2015–2016 school year, teachers completed an Internet survey that included the MBI (Maslach et al., 2016), the PHQ-9 (Kroenke & Spitzer, 2002), the CES-D-10 (Cole et al., 2004), the GAD-7 (Spitzer et al., 2006), and measures of job adversity and workplace support. Research has suggested that the reliability of scores on self-administered Internet surveys is as high, and the interpretation of such scores is as valid, as those of paper-and-pencil surveys (Ritter, Lorig, Laurent, & Matthews, 2004).

We conducted an analytic, rather than a descriptive, study (see Kristensen, 1995, p. 21). Our aim was *not* to estimate the prevalence of the conditions of interest, but to ascertain the extent to which burnout, depression, and anxiety scale items measure the same construct. While sample representativeness was not a primary concern, the inclusion of teachers who attained high, medium, and low scores on the scales of interest was. All recruitment and other procedures were approved by the Institutional Review

Boards of the City University of New York and the New York City Department of Education.

Study Sample

The only eligibility criterion for study participation was that the participant be a public school teacher. Teachers are an apt group to study because the stressfulness of working conditions and the extent of depressive and burnout symptoms in teachers vary considerably (Schonfeld, Bianchi, & Luehring-Jones, 2017). The teachers ($N = 734$; 573 women and 161 men) were recruited by contacting school administrators in 22 states in the United States. The teachers' mean age was 43 ($SD = 11.7$); 580 (79%) were married or in a close relationship. The mean number of years in the teacher's current school was 7.8 ($SD = 7.1$); the mean number of years as a teacher was 13.6 ($SD = 8.9$). A total of 286 (39%) taught in high schools; 135 (18%), in middle schools; 279 (38%), in elementary schools; and 34 (5%), in kindergarten or PreK. Regarding ethnicity/race, 47 (6%) teachers were African American; 33 (4.5%), Latino; 579 (79%), White; the other teachers identified as Native American, Asian American, Arab American, East Indian, mixed, and no response.

Measures

MBI. Burnout symptoms were assessed with the MBI-Educators Survey (Maslach et al., 2016), a 22-item questionnaire that comprises three subscales, EE, DP, and PA. EE is considered the core of burnout (Maslach et al., 2001). The response alternatives are on a 7-point frequency scale that covers the previous year (from 0 for *never* to 6 for *every day*). The means, standard deviations, and alphas for all scales used in this study can be found in Table 1. See Table 2 for the items.

PHQ-9. The PHQ-9 references the nine diagnostic criteria for major depressive disorder found in the *DSM–5* (Kroenke & Spitzer, 2002). The instrument grades symptoms by a frequency scale that ranges from 0 (*not at all*) to 3 (*nearly every day*). The items cover the previous two weeks. Although the PHQ-9 was created in the *DSM–IV* era, no changes were made between *DSM–IV* and *DSM–5* in the symptom profiles for diagnosing a major depressive episode (American Psychiatric Association, 2013).

CES-D-10. The CES-D-10 also assesses depressive symptoms. Compared to the full 20-item CES-D, the CES-D-10 reduces the response burden on study participants. Cole et al. (2004) demonstrated the structural validity of the instrument. To reduce the influence of format differences on CES-D scores' correlation with scores on the MBI subscales, we altered the response alternatives of the CES-D-10, which ordinarily cover symptoms occurring over the last week, to match the response alternatives of the MBI items.

In view of Maslach and Leiter's (2016) concern that the correlation between depression and burnout scales is high because depression scales include fatigue-related items, for the correlational analyses in Table 1 we created alternative versions of the PHQ (PHQ-7) and CES-D (CES-D-9) by omitting items that reference fatigue. Omitting items was not necessary for the factor analysis described later.

GAD-7. The GAD-7 assesses anxiety symptoms on a 0–3 rating scale. The GAD-7 is, like the PHQ-9, a module of the

Table 1

The Means (M), Standard Deviations (SD), Alpha Coefficients (α), and Pearson Correlations Are Presented for the Depressive and Anxiety Symptom Scales and the MBI Subscales

Scales	M	SD	α	CES-D-10	CES-D-9	PHQ-9	PHQ-7	GAD-7	EE	DP	PA	Job adversity
CES-D-10	2.35	1.16	.86									
CES-D-9	2.58	.80	.84	.97								
PHQ-9	9.83	6.51	.89	.80	.77							
PHQ-7	6.42	5.08	.87	.81	.79	.98						
GAD-7	8.36	6.00	.93	.72	.70	.78	.75					
EE	3.56	1.42	.93	.76	.74	.74	.71	.69				
DP	1.88	1.30	.75	.56	.55	.54	.54	.50	.60			
PA	4.29	.92	.80	−.44	−.38	−.46	−.46	−.33	−.44	−.48		
Job adversity	1.34	.92	.80	.29	.30	.31	.32	.28	.36	.42	−.23	
Workplace support	1.92	.74	.79	−.39	−.40	−.40	−.40	−.38	−.42	−.34	.27	−.22

Note. CES-D-10 = 10-item version of the Center for Epidemiologic Studies Depression scale; CES-D-9 = CES-D-10 with one fatigue item deleted; PHQ-9 = 9-item depression module of the Patient Health Questionnaire; PHQ-7 = PHQ-9 with two items that pertain to fatigue deleted; GAD-7 = 7-item Generalized Anxiety Disorder scale; EE = emotional exhaustion subscale of the Maslach Burnout Inventory (MBI); DP = depersonalization subscale of the MBI; PA = personal accomplishment subscale of the MBI. All correlations are significant at $p < .001$.

Patient Health Questionnaire and covers the previous two weeks (Spitzer et al., 2006).

Other measures. The teachers additionally completed a 5-item job adversity scale (e.g., "Students disrupted your class"; Never/very rarely . . . Daily) and a 4-item workplace support scale (e.g., "How much are the following people helpful to you in getting your job done?" Supervisors/administrators; Not at all . . . Very much). The job adversity and workplace support scales were derived from longer, more detailed scales (Schonfeld, 2001) but shortened to reduce response burden (Schonfeld & Bianchi, 2016).

Data Analysis

We searched for aberrant response patterns but found no evidence that would lead us to remove respondents (see online supplemental materials, S1). We calculated the correlations among the scales.

The question we pose is strongly focused on the extent and nature of multidimensionality among the items of interest. The bifactor model (e.g., DeMars, 2013; Reise, 2012; Rodriguez, Reise, & Haviland, 2016) is aimed precisely at this question. It works by decomposing an item's systematic variance in terms of, ideally, two sources, a general factor and one source of additional systematic variance. Our analysis proceeded as follows.

We were interested in the composition and nature of the items that load strongly on a general, nonspecific psychological distress (NSPD) factor as well as the nature of secondary factors. We therefore chose to add a bifactor for each anticipated potential source of systematic secondary variance. Given three MBI subscales (EE, DP, and PA) and three additional scales (GAD-7; CES-D-10, PHQ-9), we added six bifactors in total. The number of items per bifactor ranged from 5 to 10, indicating that the factors were substantially overidentified, a highly useful property to ensure a good representation of the secondary factors (MacCallum, Widaman, Zhang, & Hong, 1999).

Following recommendations in Heiserman and Maydeu-Olivares (2018), we made use of target rotation in a seven-factor solution. Using the method described in Browne (1972), the target rotation is specified by setting zero targets for loadings on the bifactors for items that do not belong to the scale associated with the bifactor, while allowing all other loadings to be free. For instance, on the MBI EE bifactor, all items that do not belong to the EE subscale were set to have zero targets while the loadings associated with the EE subscale were freely estimated. Rotation is accomplished by minimizing a least squares criterion with respect to the specified target values. This target encourages the general factor and bifactors to be orthogonal but allows them potentially to be oblique. The correlations of the bifactors with the general factor were all zero; the correlations among the bifactors (mean $r = .12$, corrected for direction) were small (see online supplemental materials, S2). In addition, Heiserman and Maydeu-Olivares noted that unmanaged doublets exercise a strong and potentially biasing influence on factor loading estimates, recommending the use of correlated residuals in this case. We needed to do so with one doublet involving items on the MBI's EE subscale, items 6 (working with people is a strain for me) and 16 (working with people puts too much stress on me).

In addition to the configuration of the loadings, we focused on the Explained Common Variance (ECV) and Item Explained Common Variance (IECV) measures. These measures, extensively discussed in Rodriguez et al. (2016), are simple to understand in the context of the bifactor model, where the IECV quantifies the proportion of an item's communality, C^2, accounted for by the general factor, which is

$$IEVC = \frac{\lambda^2_{NSPD}}{C^2}$$

Because the bifactor factors are overidentified, the IECV is expected to provide a useful degree of the decomposition of common variance. The ECV is simply an aggregate of the IECV over all items. Given the structure of the bifactor model, these measures are directly focused on the nature of multidimensionality in the solution. However, they depend on the potential secondary sources of variance being represented adequately and not erroneously lumped into the general factor, which is another reason why we fully factored. The literature suggests that an ECV of .8 or higher is indicative of essential unidimensionality for the set of items, and an IECV of .8 or higher indicates that an item essentially loads on the general factor. Finally, we computed Omega Hierarchical (Omega-H), which represents the sum score reliability with respect to

Table 2

Results of the Target Rotation Bifactor Analysis

Scale	Sub-scale	Item		General NSPD	EE	PA	DP	GAD	PHQ	CESD	Commun.	IECV	IECV Scale Ave.
							Bifactors						
MBI	EE	1	Drained by work	.74	.55	−.01	−.02	.04	.00	.09	.87	.63	.73
		2	Used up at workday's end	.72	.56	−.01	−.01	.02	.05	.10	.88	.60	
		3	Fatigue in morning facing job	.76	.37	−.02	−.03	−.03	.12	.06	.75	.76	
		6	Working with people is a strain	.61	.12	.00	.18	−.04	−.01	.17	.46	.81	
		8	Burned out from work	.77	.48	−.05	.02	.03	.04	−.05	.83	.71	
		13	Job is frustrating	.71	.45	.03	.07	.02	−.10	−.01	.71	.71	
		14	Works too hard	.56	.45	.09	.02	.07	−.02	.02	.55	.58	
		16	Work with people stresses me	.58	.04	−.07	.16	.05	−.02	.04	.37	.89	
		20	End of my rope	.79	.34	−.04	.05	.01	−.03	−.01	.74	.85	
	PA	4	Understand students' feelings	.01	.18	.46	−.14	−.04	.08	.12	.34	.00	.30
		7	Effective with students' problems	−.30	.06	.60	−.01	.00	.03	−.05	.46	.20	
		9	Pos. influence on lives at work	−.46	−.03	.55	−.01	.14	.01	.08	.59	.36	
		12	Energetic	−.64	−.17	.28	.11	.04	−.16	.15	.58	.70	
		17	Create relaxed atmosphere	−.41	−.10	.60	−.02	−.12	.01	−.07	.52	.33	
		18	Work with students exhilarating	−.38	−.17	.58	−.07	.07	.04	.04	.55	.26	
		19	Accomplish good things on job	−.46	−.07	.54	−.01	.19	.05	.04	.61	.35	
		21	Deal with problems calmly	−.23	.08	.40	−.03	−.13	−.06	−.04	.23	.23	
	DP	5	Treat students as objects	.41	−.05	−.15	.46	−.04	−.01	.17	.48	.35	.39
		10	Callous	.63	.08	.05	.60	.10	.04	−.18	.77	.52	
		11	Job hardening me	.66	.17	−.01	.51	.10	.06	−.17	.75	.58	
		15	Don't care about students	.36	−.02	−.16	.50	−.13	−.06	.07	.48	.27	
		22	Students blame me	.32	.00	−.05	.31	−.06	−.06	.11	.48	.22	
GAD		1	Nervous/anxious/on edge	.76	.08	.00	−.04	.50	−.07	.07	.83	.69	.64
		2	Cannot stop worrying	.76	.07	−.05	−.14	.61	−.09	.02	.95	.61	
		3	Worry too much	.73	.10	−.03	−.15	.62	−.10	.00	.92	.58	
		4	Trouble relaxing	.72	.05	.04	.01	.45	.21	−.08	.82	.63	
		5	Restless	.56	−.09	.07	.12	.42	.29	.01	.65	.49	
		6	Irritable	.72	.08	−.01	.13	.31	.16	−.08	.69	.75	
		7	Something awful might happen	.69	−.11	−.03	−.05	.42	−.09	.05	.66	.72	
PHQ		1	Little interest/little pleasure	.83	.02	.04	−.03	−.07	.19	−.10	.73	.94	.80
		2	Down/depressed/hopeless	.91	−.02	.12	−.06	−.07	.04	−.16	.88	.95	
		3	Sleep problems	.58	.02	.01	−.01	.16	.41	−.11	.57	.58	
		4	Tired or little energy	.69	.13	−.01	−.02	.05	.40	−.01	.69	.69	
		5	Appetite problems	.65	−.02	.03	−.04	−.03	.33	.07	.53	.79	
		6	Feel bad about self	.85	−.16	−.08	−.06	.06	.08	−.02	.77	.95	
		7	Trouble concentrating	.72	−.08	.07	.01	.03	.34	.15	.66	.79	
		8	Motor slowness or fidgety	.68	−.12	.05	.11	.11	.29	.14	.60	.77	
		9	Thoughts of self-harm	.71	−.32	.02	−.06	−.09	.01	−.17	.67	.76	
CESD		1	Bothered by things	.68	.16	.09	.06	.01	.02	.05	.51	.91	.78
		2	Can't shake the blues	.84	.05	.11	−.04	−.04	−.03	−.04	.72	.97	
		3R	As good as other people	.46	−.14	−.20	−.15	.00	.07	−.04	.29	.71	
		4	Can't kind mind on things	.53	.12	.04	−.11	−.07	.19	.45	.56	.51	
		5	Everything is an effort	.63	.18	−.02	−.06	−.01	.10	.35	.59	.69	
		6R	Hopeful about future	.59	.08	−.18	−.09	−.09	−.04	−.23	.47	.74	
		7	Life is a failure	.74	−.16	−.03	−.02	−.11	−.10	.04	.61	.90	
		8	Fearful	.69	−.07	.02	−.09	.18	−.23	.24	.61	.77	
		9	Lonely	.70	−.15	.17	.01	−.03	−.18	.06	.58	.84	
		10	People unfriendly	.50	−.03	.15	.18	.02	−.14	.10	.34	.74	

the general factor. However, this measure is not directly aimed at the question of multidimensionality insofar as we are not advocating for the formation of a sum score of all these items.

Results and Discussion

Correlations

Table 1 indicates that the CES-D-10, CES-D-9, PHQ-9, PHQ-7, and GAD-7 scores were similarly related to EE, DP, and PA scores. Removing fatigue-related items to create the CES-D-9 and PHQ-7 barely changed the correlations. EE scores were almost as highly correlated with scores on the anxiety scale as they were with scores on the depression scales. The relationship of CES-D scores to scores on the MBI subscales paralleled the relationship of scores on the PHQ to EE, DP, and PA scores. EE scores were more closely related to depression and anxiety scores than to DP and PA scores. Using Steiger's (1980) *z* test, we found that depression, anxiety, EE, and PA scores were similarly ($p > .05$) related to scores on our measures of job adversity and workplace support. Compared to depression and

anxiety scores, DP scores were more highly related to job adversity; no significant differences were found for workplace support.

In our bifactor modeling, we fit the model using the WLSMV method in Mplus 8.1 (Muthén & Muthén, 2018), which is recommended for ordinal data (Li, 2016). Ordinal data are often skewed in a manner that is inconsistent with multivariate Gaussian assumptions, which is true for many of our variables, as expected for data from a community sample. The fitted model has an RMSEA = .041, CFI = .984, TLI = .977, and SRMR = .023. There is no evidence of an improper solution such as an extraordinarily large standard error or communality near 1. While the chi-square test versus the saturated model unsurprisingly rejects fit to the saturated model, the modification indices and bivariate residuals did not show a notable pattern of misfit. Table 2 shows the target-rotated standardized loadings, communalities, and IECVs. For the reader's convenience, we also computed mean IECVs within each set of items making up a bifactor. All values are included in the Mplus output, which can be found online (see online supplemental materials, S3).

Omega-H for the general factor was .95 and the ECV was .63. The latter number suggests some multidimensionality despite a high total score reliability. The IECVs and the scale average IECV show that CES-D and PHQ items were essentially unidimensional measures of NPSD. The items on these scales only had strong loadings on the general factor and weak loadings on their respective bifactors. GAD-7 items had strong loadings on the general factor but moderate bifactor loadings that were consistently smaller than the items' general factor loadings. Of the three MBI subscales, EE was very similar to the GAD-7 in terms of performance, except for the doublet formed by items 6 and 16, which was managed by our including correlated residuals. EE items had mostly strong loadings on the general factor but moderate bifactor loadings that were consistently smaller than the loadings on the NPSD factor. In the cases of the responses to the EE and GAD-7 items, the bifactor likely represented common item content on each scale, notably the presence of anxiety-related words on the GAD-7 items and repeated references to work on the EE items.

The PA and DP items, by contrast, were less clearly related to the general factor, as reflected by their IECV values and lower overall communality. Rerunning the analysis with the items associated with PA and DP omitted showed essentially the same pattern and the ECV moved to .75, close to essential unidimensionality, as anticipated given the level of overidentification. In sum, responses to EE items behaved in a manner consistent with responses to items on other measures of NSPD and consistent with specific item content. Responses to EE items reflected a notable but smaller bifactor, suggesting some unique variance, reflecting the common variance induced by references to work. By contrast, PA and DP items reflected more unique variance than did EE items.

Correlational analyses revealed that the MBI's EE subscale was more highly related to the PHQ (-9 and -7), the CES-D (-10 and -9), and the GAD-7 than to the MBI's own DP and PA subscales. In terms of nomological networks, the two depression scales, the anxiety scale, and the MBI's EE subscale were similarly related to measures of job adversity and workplace support. Overall, these findings question the claim that the burnout construct represents a syndrome that consists of EE, DP, and diminished PA

and excludes (or does not primarily include) depressive and anxiety symptoms.

The results of the ESEM bifactor analysis indicated that the CES-D-10, PHQ-9, GAD-7, and EE items loaded substantially on the prime factor, which can be thought of as NSPD. Most EE and anxiety items loaded less strongly on a secondary factor. The DP items loaded about equally on two different constructs, NSPD and a depersonalization factor. The PA items reflected two different constructs, professional efficacy and, to a lesser extent, NSPD. On balance, the evidence suggests that depressive, anxiety, and EE scales measure the same construct. That construct can be conceived of as NSPD, consistent with the hypothesis of a general anxio-depressive dimension of ill-being (Kotov et al., 2017; Lang & McTeague, 2009).

Other Research

Leiter and Durup (1994) found burnout and depression to be separate constructs. However, problems in their CFA call their conclusions into question. First, their best model did not fit the data well (highest AGFI ≈ .810). Second, depression and EE were as closely correlated as EE and DP although nearly half the depression items were dropped from the CFA. Although Bakker et al. (2000) also found burnout and depression to be separate constructs, problems call their CFA findings into question as well. EE was more closely related to depression ($r = .68$) than to DP or PA despite the MBI subscales and depression having very different response formats. Surprisingly, the authors chose *not* to let EE and depression items load on the same latent variable and did not report the factor correlations in their CFA. The authors additionally found that, compared to latent depression, latent burnout was more closely related to lack of workplace equity. This finding, however, likely reflects measurement circularity (i.e., content overlap between independent and dependent variables), a problem that has strongly affected burnout research (Schaufeli & Enzmann, 1998). Finally, the CFAs by Leiter and Durup (1994) and Bakker et al. (2000) were limited given the challenges of fitting ordinal SEM in the era when the studies were conducted.

The ESEM approach we adopted has methodological advantages. ESEM helped us identify methodological artifacts of scale construction that, in addition to the NSPD construct, contributed to covariation among items. For example, most EE items that referenced work loaded on the second factor (although they loaded more strongly on the NSPD factor). This happens because the shared reference to work induces some additional local dependence among the items. Target rotation provides a principled means to represent the hypotheses about the dimensionality of the data and the ECV and IECV measures are informative measures about these hypotheses.

Limitations

While our study has strengths, such as the use of advanced factor analytic techniques and a relatively large sample size, it also has several limitations. First, the sample comprised only teachers. Other occupational groups need to be studied to learn if the findings generalize. However, we note that research conducted with dentists (Ahola et al., 2014) and physicians (Wurm et al., 2016) is consistent with our findings (also see Bianchi & Brisson,

2017). Second, the study was limited to one measure of burnout. Research conducted with other instruments, such as the Shirom-Melamed Burnout Measure (Shirom & Melamed, 2006), also revealed substantial correlations with the PHQ-9 (e.g., Schonfeld & Bianchi, 2016). A third limitation is that because we relied strictly on the wording of the prominent measures we selected, we did not include "catch" items designed to identify aberrant responders. Instead we relied on Mokken analyses to check for aberrant responding (see online supplemental materials, S1). A fourth limitation is that, because the data were cross-sectional, we could not investigate the relationships over time. Although he did not assess burnout longitudinally, Schonfeld (2001), in a 1-year, 3-wave longitudinal study of first-year teachers, found that adverse job conditions exerted a substantial effect on the teachers' responses to the CES-D. Depressive symptom scales are in fact highly sensitive to the quality of the work environment (e.g., Schonfeld & Chang, 2017).

Conclusions

The MBI's EE subscale and depression and anxiety scales are likely measuring NSPD. At high levels of distress, the risk of mental disorder increases. In some individuals, high levels of distress can turn into clinical depression. Dimensional measures of NSPD thus provide a glimpse into the process by which job stressors give rise to clinical forms of depression and thus can alert health care specialists to a potential onset of a diagnosable disorder in a worker, paving the way for an appropriate treatment.

References

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Bakker, A. B., Schaufeli, W. B., Demerouti, E., Janssen, P. P. M., Van der Hulst, R., & Brouwer, J. (2000). Using equity theory to examine the difference between burnout and depression. *Anxiety, Stress, and Coping, 13,* 247–268. http://dx.doi.org/10.1080/10615800008549265

Bianchi, R., & Brisson, R. (2017). Burnout and depression: Causal attributions and construct overlap. *Journal of Health Psychology.* Advance online publication. http://dx.doi.org/10.1177/1359105317740415

Bianchi, R., & Schonfeld, I. S. (2018). Burnout-depression overlap: Nomological network examination and factor-analytic approach. *Scandinavian Journal of Psychology, 59,* 532–539. http://dx.doi.org/10.1111/sjop.12460

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

Browne, M. W. (1972). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology, 25,* 207–212. http://dx.doi.org/10.1111/j.2044-8317.1972.tb00492.x

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Moffitt, T. E. (2014). The P factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science, 2,* 119–137. http://dx.doi.org/10.1177/2167702613497473

Cole, J. C., Rabin, A. S., Smith, T. L., & Kaufman, A. S. (2004). Development and validation of a Rasch-derived CES-D short form. *Psychological Assessment, 16,* 360–372. http://dx.doi.org/10.1037/1040-3590.16.4.360

DeMars, C. E. (2013). A tutorial on interpreting bi-factor model scores. *International Journal of Testing, 13,* 354–378. http://dx.doi.org/10.1080/15305058.2013.799067

Dohrenwend, B. P., Shrout, P. E., Egri, G., & Mendelsohn, F. S. (1980). Nonspecific psychological distress and other dimensions of psychopathology. Measures for use in the general population. *Archives of General Psychiatry, 37,* 1229–1236. http://dx.doi.org/10.1001/archpsyc.1980.01780240027003

Heiserman, N., & Maydeu-Olivares, A. (2018). *Best practices for exploratory factor analysis: Target rotations and correlated errors.* Manuscript submitted for publication.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology, 126,* 454–477. http://dx.doi.org/10.1037/abn0000258

Kristensen, T. S. (1995). The demand-control-support model: Methodological challenges for future research. *Stress Medicine, 11,* 17–26. http://dx.doi.org/10.1002/smi.2460110104

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32,* 509–515. http://dx.doi.org/10.3928/0048-5713-20020901-06

Lang, P. J., & McTeague, L. M. (2009). The anxiety disorder spectrum: Fear imagery, physiological reactivity, and differential diagnosis. *Anxiety, Stress & Coping, 22,* 5–25. http://dx.doi.org/10.1080/10615800802478247

Leiter, M. P., & Durup, J. (1994). The discriminant validity of burnout and depression: A confirmatory factor analytic study. *Anxiety, Stress, and Coping, 7,* 357–373. http://dx.doi.org/10.1080/10615809408249357

Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods, 21,* 369–387. http://dx.doi.org/10.1037/met0000093

MacCallum, R. C., Widaman, K. R., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84–99. http://dx.doi.org/10.1037/1082-989X.4.1.84

Maslach, C., Jackson, S. E., & Leiter, M. P. (2016). *Maslach Burnout Inventory manual* (4th ed.). Palo Alto, CA: Consulting Psychologists Press.

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

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology, 52,* 397–422. http://dx.doi.org/10.1146/annurev.psych.52.1.397

Muthén, L. K., & Muthén, B. O. (2018). *Mplus user's guide, 8th ed.* Los Angeles, CA: Author.

Reise, S. P. (2012). The rediscovery of bi-factor modeling. *Multivariate Behavioral Research, 47,* 667–696. http://dx.doi.org/10.1080/00273171.2012.715555

Ritter, P., Lorig, K., Laurent, D., & Matthews, K. (2004). Internet versus mailed questionnaires: A randomized comparison. *Journal of Medical Internet Research, 6,* e29. http://dx.doi.org/10.2196/jmir.6.3.e29

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

Rössler, W., Hengartner, M. P., Ajdacic-Gross, V., & Angst, J. (2015). Predictors of burnout: Results from a prospective community study. *European Archives of Psychiatry and Clinical Neuroscience, 265,* 19–25. http://dx.doi.org/10.1007/s00406-014-0512-x

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

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

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

Schonfeld, I. S., Bianchi, R., & Luehring-Jones, P. (2017). Consequences of job stress for the psychological well-being of teachers. In T. M. McIntyre, S. E. McIntyre, & D. J. Francis (Eds.), *Educator stress: An occupational health perspective* (pp. 55–75). Cham, Switzerland: Springer International Publishing.

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

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,* 176–200. http://dx.doi.org/10.1037/1072-5245.13.2.176

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

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Ar-chives of Internal Medicine, 166,* 1092–1097. http://dx.doi.org/10.1001/archinte.166.10.1092

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245–251. http://dx.doi.org/10.1037/0033-2909.87.2.245

Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *MMPI–2 Restructured Clinical (RC) scales: Development, validation, and interpretation.* Minneapolis, MN: University of Minnesota Press.

Wurm, W., Vogel, K., Holl, A., Ebner, C., Bayer, D., Mörkl, S., . . . Hofmann, P. (2016). Depression-burnout overlap in physicians. *PLoS ONE, 11,* e0149913. http://dx.doi.org/10.1371/journal.pone.0149913

Zijlstra, W. P., van der Ark, A. L., & Sijtsma, K. (2007). Outlier detection in test and questionnaire data. *Multivariate Behavioral Research, 42,* 531–555. http://dx.doi.org/10.1080/00273170701384340