Assessment
19(2) 198–204

DOI: 10.1177/1073191112438743
http://asm.sagepub.com

Mind Your Words: Positive and Negative Items Create Method Effects on the Five Facet Mindfulness Questionnaire

Nicholas T. Van Dam[1], Andréa L. Hobkirk[1], Sharon Danoff-Burg[1], and Mitch Earleywine[1]

Abstract

Mindfulness, a construct that entails moment-to-moment effort to be aware of present experiences and positive attitudinal features, has become integrated into the sciences. The Five Facet Mindfulness Questionnaire (FFMQ), one popular measure of mindfulness, exhibits different responses to positively and negatively worded items in nonmeditating groups. The current study employed confirmatory factor analysis with a large undergraduate sample to examine the validity of a hierarchical mindfulness model and whether response patterns related to item wording arose from method effects. Results indicated that a correlated facets model better explained the data and that negative and positive wording constituted substantive method effects. This study suggests that the FFMQ measures components that may relate to, but do not seem to directly reflect, a latent variable of mindfulness. The authors recommend against the use of an FFMQ total score, favoring individual scale scores, and further examination of method effects in mindfulness scales.

Keywords

mindfulness, psychometrics, method effects, item wording effects, FFMQ

Mindfulness (a complex construct that entails moment-to-moment effort to be aware of present experiences along with several important positive attitudinal features; see Grossman & Van Dam, 2011) has become widely integrated into the social and basic sciences. This integration includes efficacious therapeutic approaches founded on Buddhist principles (e.g., Hofmann, Sawyer, Witt, & Oh, 2010), exploration of mindfulness's impact on neurobiology (Lutz, Slagter, Dunne, & Davidson, 2008), and multiple instruments designed to measure the construct (see, e.g., Christopher, Charoensuk, Gilbert, Neary, & Pearce, 2009). Although interventions and neurobehavioral associations have received a relatively warm reception in the scientific community, instruments aiming to measure mindfulness have been questioned on multiple grounds (e.g., Christopher et al., 2009; Davidson, 2010; Grossman & Van Dam, 2011; Höfling, Moosbrugger, Schermelleh-Engel, & Heidenreich, 2011).

Examinations of the Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003), the most popular of mindfulness measures, have shown that it does not seem to measure the latent variable it purports to (Van Dam, Earleywine, & Borders, 2010). The scale also lacks expected differences between meditating and nonmeditating samples (Christopher et al., 2009). Initial development of the MAAS assumed that mind*ful* states were less accessible than mind*less* states. Philosophical problems with this argument notwithstanding (see Grossman & Van Dam, 2011; Van Dam, Sheppard, Forsyth, & Earleywine, 2011), the scale's developers stated that an analysis of a positively worded version of the MAAS was not psychometrically as sound as the commonly used, negatively worded version. However, a comparison of positively worded and negatively worded versions of the MAAS suggests comparable psychometric limitations in both scales (Höfling et al., 2011), namely, a strong influence of wording direction (i.e., negative or positive), constituting a method effect—after the multitrait–multimethod approach of Campbell and Fiske (1959). Method effects are unlikely limited to the MAAS; many mindfulness questionnaires rely on reverse-scored, negatively worded items, including the Five Facet Mindfulness Questionnaire (FFMQ)—one subscale of which is mostly composed of items from the MAAS.

[1]University at Albany, SUNY, Albany, NY, USA

Corresponding Author:
Nicholas T. Van Dam, Department of Psychology, University at Albany, SUNY, Social Sciences 399, 1400 Washington Avenue, Albany, NY 12222, USA
Email: ntvandam@gmail.com

Christopher et al. (2009) recently showed that the Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith, & Allen, 2004) lacks configural invariance (an important feature of psychometric validation) across Thai and American samples. In addition, Baum et al. (2010) found that the KIMS lacks a hierarchical "mindfulness" factor. The FFMQ has also shown inconsistent factor structures in meditating and nonmeditating samples (Baer et al., 2008). Furthermore, nonmeditators are more likely to reject items on the FFMQ indicating mind*less*ness than they are to endorse items on the FFMQ indicating mind*ful*ness; meditators show no differences. In one analysis, this pattern of endorsement was related to differential item functioning (DIF) on nearly half of the FFMQ's items (Van Dam, Earleywine, & Danoff-Burg, 2009). However, a comparable analysis in a population of meditators and nonmeditators that was demographically matched showed minimal evidence of DIF (Baer, Samuel, & Lykins, 2011; for an alternative interpretation, see Grossman & Van Dam, 2011). This same study replicated the finding of significantly greater rejection of mind*less*ness items than endorsement of mind*ful*ness items in nonmeditators, with no difference in endorsement rates among meditators. Baer et al. (2011) concluded that comparing positively and negatively worded items is "problematic because it confounds the items' scoring direction with their content." (p. 7). However, many of the inconsistencies with the factor structure of the FFMQ, and even the varying findings of DIF, could arise from the effects of item wording. The idea of method effects (of which item wording is one example) stems from Campbell and Fiske's (1959) publication on the multitrait–multimethod approach. If negative and positive wording constitute factors of their own, previous psychometric problems with the FFMQ might arise from an alternative factor structure than proposed (Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006).

Objectives

The present study explores the validity of the FFMQ in relation to several possible factor configurations. Limitations to the hierarchical model of the closely related KIMS (Baum et al., 2010) and inconsistencies in the factor structure of the FFMQ (e.g., Baer et al., 2006; Baer et al., 2008) led us to first question whether a hierarchical model provides the best fit for the FFMQ. We also examined how method effects influenced endorsement of FFMQ items. In light of recent evidence for method effects on the MAAS (Höfling et al., 2011), method effects for positively and negatively worded items on the FFMQ might provide the most parsimonious explanation for recent psychometric findings (e.g., Baer et al., 2011; Van Dam et al., 2009).

Method

Procedure

Undergraduates (N = 459) at the University at Albany, SUNY, participated for course credit. Cases with missing values were excluded (n = 24). Participants averaged 18.9 years of age (SD = 2.0), and 70.4% were female. Participants self-identified as Caucasian (69%), Asian (10.3%), African American (8.6%), Hispanic or Latino (7.4%), and "Other" (4.6%). Responses were orthogonal to demographic characteristics. Procedures were approved by the institutional review board.

Measures

Mindfulness. Participants completed the 39-item version of the FFMQ (Baer et al., 2006), which has reasonable psychometric properties in students, community members, and meditators (Baer et al., 2006; Baer et al., 2008), though its properties have varied across groups (Baer et al., 2008; Van Dam et al., 2009). Internal consistencies (Cronbach's alpha) among each of the FFMQ subscales were as follows: Observe = .83, Describe = .88, ActAware = .89, NonJudge = .91, and NonReact = .81.

Analysis

We contrasted a hierarchical model (Figure 1, Model 1) with a correlated factors model (Figure 1, Model 1A); the better model served as the base model. We then compared five models: (1) base model (Figure 1, Model 1A), (2) base model plus a negative method factor (Figure 1, Model 2), (3) base model plus a positive method factor (Figure 1, Model 3), (4) base model plus uncorrelated negative and positive method factors (Figure 1, Model 4), and (5) a seven-factor model including the original five factors plus two correlated method factors (positive and negative; Figure 1, Model 5). All models used individual items as indicators (in contrast to prior work; see Baer et al., 2006; Baer et al., 2008).

Cutoff criteria for reasonable and good model fit were based on extant recommendations (Brown, 2006; Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004). The Akaike information criterion (AIC), a relative fit index that considers model parsimony, has no specific value that suggests a "good" model fit. A meaningful decrease in AIC suggests improved fit (Brown, 2006). Similarly, the expected cross-validation index (ECVI) also indicates relative fit considering parsimony. ECVI, however, incorporates sample size as well, penalizing nonparsimonious models more for smaller samples (Brown, 2006). Given that the models were nonnested, AIC and ECVI were considered the primary indicators of model fit

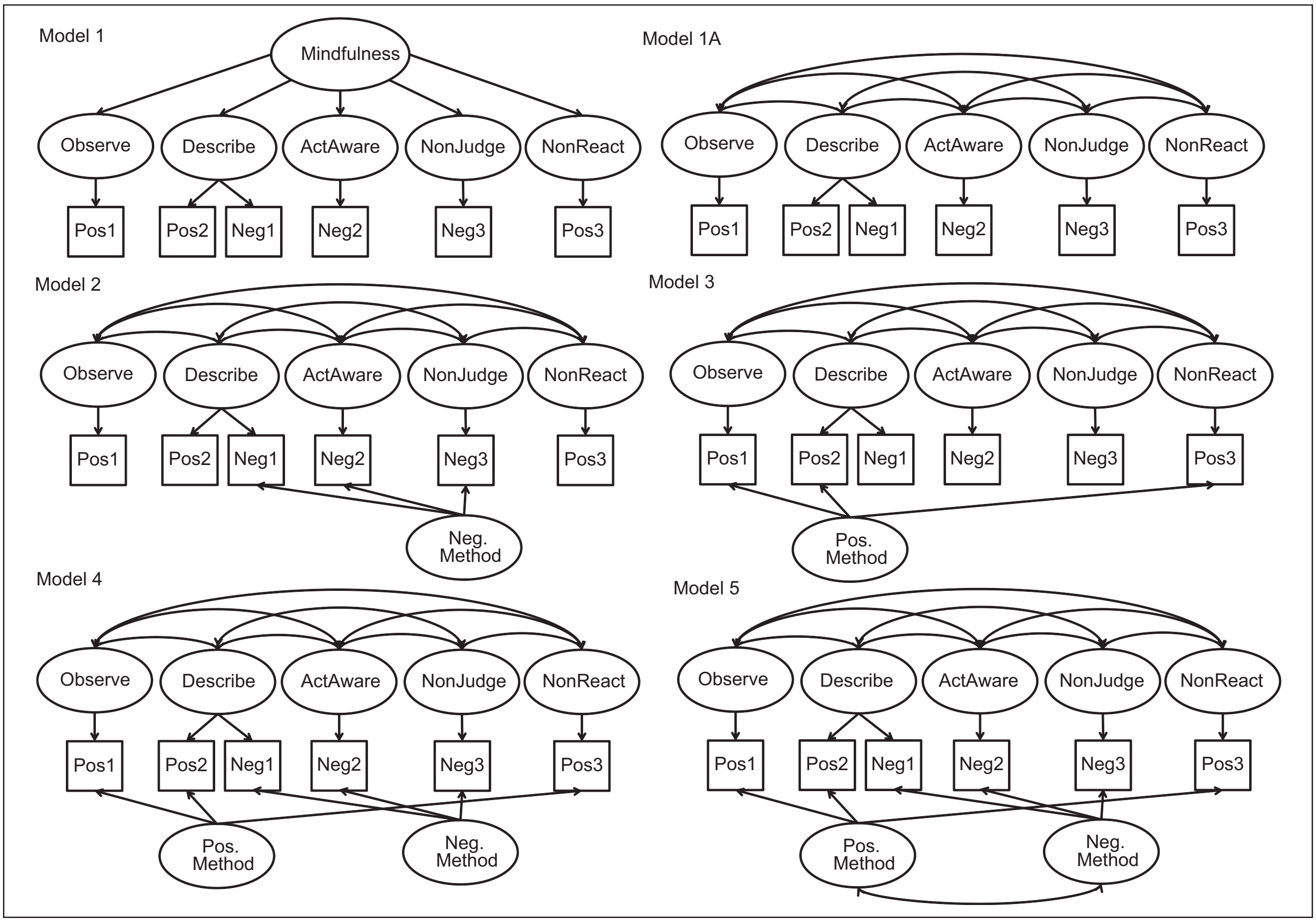

Figure 1. Competing models tested for the Five Facet Mindfulness Questionnaire
Individual items have been grouped (e.g., Pos2 = second cluster of positive items; Neg1 = first cluster of negative items) for ease of viewing. Analyses were conducted with individual items. Model 1—hierarchical five-factor model, Model 1A—correlated five-factor model, Model 2—negative method factor, Model 3—positive method factor, Model 4—uncorrelated negative and positive method factors, Model 5—correlated negative and positive method factors.

improvement. Comparative fit index >.95 was considered good and >.90 was considered reasonable, root mean square error of approximation <.06 was considered good and <.08 was considered reasonable, standardized root mean residual <.08 was considered good and <.10 was considered reasonable. Confirmatory factor analysis computations were made using maximum likelihood estimation in LISREL v8.8 (Jöreskog & Sörbom, 1993). All other statistics were computed in SPSS v18.

Results

Items exhibited normal range, skewness (M = 0.24, Mdn = –0.14), and kurtosis (M = –0.37, Mdn = –0.41). There were no univariate outliers (defined as $|Z| > 3.29$). The Kaiser–Meyer–Olkin statistic of .89 suggested insufficient multicollinearity to impact factor analysis. Out of 435 cases with complete responses, 19 cases (4.4%) exhibited Mahalanobis distance values exceeding $p < .001$. To eliminate their potential influence, these 19 cases were excluded (remaining $n = 417$).

The correlated five-factor model exhibited a better fit than the hierarchical five-factor model, showing improvement across all fit indices (see Table 1). Accordingly, the correlated five-factor model served as the base model for all subsequent comparisons. Each method effect model (positive method—Model 2, negative method—Model 3, uncorrelated positive and negative methods—Model 4, and correlated positive and negative methods—Model 5) showed improvement in fit over the base model and the previous, less specified model (see Table 1). Although the model fit changes from Model 4 to Model 5 were small, the latent method variables exhibited a meaningful

Table 1. Competing Model Fit Criteria

Model	*df*	χ^2	AIC	ECVI	CFI	RMSEA	SRMR
1 Hierarchical	697	1,975	2,141	5.160	.937	.067	.079
1 A Corr 5-F	692	1,898	2,074	4.998	.940	.065	.059
2 Corr 5-F + Neg-M	673	1,547	1,761	4.244	.953	.056	.058
3 Corr 5-F + Pos-M	672	1,600	1,816	4.376	.951	.058	.054
4 Corr 5-F + UnC-M	653	1,269	1,523	3.669	.965	.048	.053
5 Corr 5-F + C-M	652	1,252	1,508	3.633	.965	.047	.052
Excluding the Observe subscale							
Hierarchical	430	1,387	1,519	3.660	.946	.073	.066
Corr 4-F	428	1,367	1,503	3.623	.947	.073	.059
Corr 4-F + C-M	396	756	956	2.303	.976	.047	.050

Note. Corr 5-F = correlated five-factor model; Neg-M = negative method effect; Pos-M = positive method effect; UnC-M = uncorrelated positive and negative method effects; C-M = correlated positive and negative method effects; Corr 4-F = correlated four-factor model; AIC = Akaike information criterion; ECVI = expected cross-validation index; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Table 2. Factor Correlations

	Observe	Describe	ActAware	NonJudge	NonReact
Observe	—	$.27^M$	–.02	$–.11^S$	$.13^S$
Describe	$.27^M$	—	$.34^M$	.05	$.20^S$
ActAware	–.04	$.39^L$	—	$.36^M$	$.25^M$
NonJudge	$–.10^S$	$.10^S$	$.40^L$	—	$.28^M$
NonReact	$.12^S$	$.22^S$	$.31^M$	$.34^M$	—

Note. Superscript S, M, and L denote small, medium, and large effect size, respectively. Lower left triangle represents latent variable correlations from Model 1A. Upper right triangle represents sum score correlations. $|r| > .1$ are significant at $p < .05$.

correlation ($r = .36$, $p < .001$), supporting the latter model. Notably, only the models that included method effects (Models 2-5) met the cutoff criteria for a "good" model fit (see Table 1).

The items (or parcels) of the Observe subscale have exhibited nonsignificant loadings on the hierarchical "mindfulness" factor across several studies (e.g., Baer et al., 2006; Baer et al., 2008), suggesting unique properties for these items across varying groups. Similarly, the items comprising the Observe subscale failed to exhibit a significant loading on the hierarchical factor in the present data ($\lambda = .03$, $t = 0.48$, $p > .05$). To control for an alternative explanation due to the Observe subscale, hierarchical four-factor and correlated four-factor models, excluding items from the Observe subscale, were compared to establish an alternative base model. Much like the five-factor model, the correlated four-factor model exhibited a meaningful improvement in model fit (see Table 1). The correlated four-factor model plus positive and negative method effects improved on base model fit (see Table 1). Again, only the positive and negative method effect models met the cutoff criteria for a "good" model fit (see Table 1).

Factor correlations varied for latent variables and the sum scores (see Table 2). Only the relation between Observe and ActAware was nonsignificant across both latent variable and sum correlations. Nevertheless, half of the latent and sum score correlations were small or nonsignificant. Standardized factor loadings for the correlated five-factor model plus positive- and negative method factors appear in Table 3. Most items (92.3%) exhibited standard factor loadings on the original five factors larger than the recommended value of 0.3. However, 33.3% of the items also exceeded this recommended value for standardized factor loadings on positive or negative method effects (see Table 3). Additionally, the negative method factor accounted for more response variance than two of the original five factors.

Discussion

Interrelated Mindful Constructs

The present findings pose several challenges to current conceptualizations of the FFMQ. A correlated five-factor model fit the data better than a hierarchical five-factor model. This was also the case for a correlated four-factor model relative to a hierarchical four-factor model without Observe subscale items. Support for a correlated factors model (rather than a hierarchical model), modest factor intercorrelations (see Table 2), and contradictory relations of the subscales with relevant external constructs (e.g., Baer et al., 2008) suggest that the facets of the FFMQ may represent different correlated latent constructs rather than a cluster of items essential to a hierarchical notion of mindfulness. This is not to say that mindfulness is not a multidimensional construct with various features and components but rather that the present factor analyses of the FFMQ do not support a single hierarchical model.

Item Wording

The present analyses suggest that wording contributes to endorsement on the FFMQ, possibly creating important

Table 3. Standardized Factor Loadings for the Five Facet Mindfulness Questionnaire, Including Negative and Positive Method Factors From Model 5

Item	Observe	Describe	ActAware	NonJudge	NonReact	Negative method factor	Positive method factor
1	0.61						0.19
6	0.70						0.18
11	0.53						0.20
15	0.77						0.19
20	0.75						0.17
26	0.50						0.20
31	0.64						0.24
36	0.42						0.36
2		0.67					0.37
7		0.68					0.39
12		0.81				0.22	
16		0.82				0.24	
22		0.50				0.25	
27		0.31					0.40
32		0.49					0.57
37		0.51					0.52
5			0.86			0.29	
8			0.71			0.43	
13			0.74			0.31	
18			0.37			0.56	
23			0.19			0.66	
28			0.22			0.60	
34			0.14			0.64	
38			0.31			0.66	
3				0.68		0.24	
10				0.73		0.18	
14				0.78		0.25	
17				0.72		0.11	
25				0.82		0.19	
30				0.79		0.26	
35				0.70		0.25	
39				0.68		0.17	
4					0.31		0.05
9					0.46		0.08
19					0.76		0.22
21					0.45		0.27
24					0.70		0.06
29					0.61		0.09
33					0.72		0.10
Estimated % variance	8.04	7.91	5.46	11.21	6.33	7.30	4.09

differences among relevant comparison groups (e.g., Baer et al., 2008; Baer et al., 2011; Van Dam et al., 2009). Previous DIF analyses may not necessarily have been confounded on the basis of item content, as recently suggested (Baer et al., 2011), but may instead result from different contributions of method effects. Method effects seem a more parsimonious explanation than item impact (see Ackerman, 1992) for recent findings of differential item endorsement by group (Baer et al., 2011; Van Dam et al., 2009). The significant correlation between positive and negative method factors may suggest that individuals susceptible to negative method effects are also susceptible to positive method effects and vice versa, a finding that helps explain recent endorsement patterns (Baer et al., 2011; Van Dam et al., 2009). Of note, method effects might be particularly strong when item content is less familiar. Differences in endorsement of negative and positive items may suggest greater susceptibility to method effects in nonmeditators relative to meditators (e.g., Baer et al., 2011; Van Dam et al., 2009), potentially supporting an interaction of method effects by content knowledge.

This conclusion is an empirical question worthy of further exploration.

Implications for the FFMQ

The lack of a superordinate "mindfulness" factor in the present analyses suggests that the subscales may be related but cannot be considered as being subsumed (at least statistically) by a hierarchical factor. This finding leads us to recommend against the use of an FFMQ total scale score. The present data are consistent with findings that certain (sub)scales of the FFMQ seem to have greater predictive utility than others (see, e.g., Coffey, Hartman, & Fredrickson, 2010). The individual scales might thus be considered on their own and in conjunction with other measures that have support on both theoretical and empirical grounds (see, e.g., Grossman & Van Dam, 2011; Van Dam et al., 2011).

Additionally, recent work reveals that the (reverse scored and negatively worded) ActAware and NonJudge subscales were primarily responsible for total scale association with repetitive thoughts (Evans & Segerstrom, 2011). Perhaps, as has been suggested in personality research, the method effect of negative item wording relates to the behavioral inhibition system (DiStefano & Motl, 2006; Quilty, Oakman, & Risko, 2006). The positive method effect could similarly covary with impression management. Future research should explore the relations of the FFMQ facets to external latent variables while attempting to control for method effects. The issue of method effects is not unique to assessment of mindfulness (e.g., DiStefano & Motl, 2006); however, it may be particularly important in this controversial area of assessment. Examining various method effects may improve our understanding of the facets of the FFMQ (and other measures of mindfulness). Additionally, including other constructs and scales consistent with traditional (Buddhist) notions of mindfulness may increase the construct and predictive validity of the scales (e.g., Coffey et al., 2010), while also permitting continued exploration of the most psychometrically valid and contextually appropriate operationalization(s) of mindfulness.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement,29*, 67-91. doi:10.1111/j.1745-3984.1992.tb00368.x

Baer, R. A., Samuel, D. B., & Lykins, E. L. B. (2011). Differential item functioning on the Five Facet Mindfulness Questionnaire is minimal in demographically matched meditators and nonmeditators. *Assessment,18*, 3-10. doi:10.1177/1073191110392498

Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment,3*, 191-206. doi:10.1177/1073191104268029

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment,13*, 27-45. doi:10.1177/1073191105283504

Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., . . . Williams, J. M. G. (2008). Construct validity of the Five Facet Mindfulness Questionnaire in meditating and nonmeditating samples. *Assessment,15*, 329-342. doi:10.1177/1073191107313003

Baum, C., Kuyken, W., Bohus, M., Heidenreich, T., Michalak, J., & Steil, R. (2010). The psychometric properties of the Kentucky Inventory of Mindfulness Skills in clinical populations. *Assessment,17*, 220-229. doi:10.1177/1073191109356525

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology,84*, 822-848.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York, NY: Guilford Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin,56*, 81-105. doi:10.1037/h0046016

Christopher, M. S., Charoensuk, S., Gilbert, B. D., Neary, T. J., & Pearce, K. L. (2009). Mindfulness in Thailand and the United States: A case of apples versus oranges? *Journal of Clinical Psychology,65*, 590-612. doi:10.1002/jclp.20580

Coffey, K. A., Hartman, M., & Fredrickson, B. L. (2010). Deconstructing mindfulness and constructing mental health: Understanding mindfulness and its mechanisms of action. *Mindfulness,1*, 235-253. doi:10.1007/s12671-010-0033-2

Davidson, R. J. (2010). Empirical explorations of mindfulness: Conceptual and methodological conundrums. *Emotion,10*, 8-11. doi:10.1037/a0018480

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling,13*, 440-464.

Evans, D., & Segerstrom, S. (2011). Why do mindful people worry less? *Cognitive Therapy and Research,35*, 505-510. doi:10.1007/s10608-010-9340-0

Grossman, P., & Van Dam, N. T. (2011). Mindfulness, by any other name . . .: Trials and tribulations of Sati in Western psychology and science. *Contemporary Buddhism,12*, 219-239. doi:10.1080/14639947.2011.564841

Höfling, V., Moosbrugger, H., Schermelleh-Engel, K., & Heidenreich, T. (2011). Mindfulness or mindlessness? A modified version of the Mindful Attention and Awareness Scale (MAAS). *European Journal of Psychological Assessment,27*, 59-64. doi:10.1027/1015-5759/a000045

Hofmann, S. G., Sawyer, A. T., Witt, A. A., & Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression:

A meta-analytic review. *Journal of Consulting and Clinical Psychology,78*, 169-183. doi:10.1037/a0018555

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternations. *Structural Equation Modeling,6*, 1-55.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Lincolnwood, IL: Scientific Software International.

Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences,12*, 163-169. doi:10.1016/j.tics.2008.01.005

Marsh, H. W., Hau, K. T., & Wen, Z. L. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler (1999) findings. *Structural Equation Modeling,11*, 320-341.

Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg Self-Esteem Scale method effects. *Structural Equation Modeling,13*, 99-117.

Van Dam, N. T., Earleywine, M., & Borders, A. (2010). Measuring mindfulness? An item response theory analysis of the Mindful Attention Awareness Scale. *Personality and Individual Differences,49*, 805-810. doi:10.1016/j.paid.2010.07.020

Van Dam, N. T., Earleywine, M., & Danoff-Burg, S. (2009). Differential item function across meditators and non-meditators on the five facet mindfulness questionnaire. *Personality and Individual Differences,47*, 516-521. doi:10.1016/j.paid.2009.05.005

Van Dam, N. T., Sheppard, S. C., Forsyth, J. P., & Earleywine, M. (2011). Self-compassion is a better predictor than mindfulness of symptom severity and quality of life in mixed anxiety and depression. *Journal of Anxiety Disorders,25*, 123-130. doi:10.1016/j.janxdis.2010.08.011