Psychological Assessment

Does the Five Facet Mindfulness Questionnaire Measure What We Think It Does? Construct Validity Evidence From an Active Controlled Randomized Clinical Trial

Simon B. Goldberg, Joseph Wielgosz, Cortland Dahl, Brianna Schuyler, Donal S. MacCoon, Melissa Rosenkranz, Antoine Lutz, Chad A. Sebranek, and Richard J. Davidson

Online First Publication, October 12, 2015. http://dx.doi.org/10.1037/pas0000233

Psychological Assessment

1040-3590/15/$12.00 http://dx.doi.org/10.1037/pas0000233

BRIEF REPORT

Does the Five Facet Mindfulness Questionnaire Measure What We Think It Does? Construct Validity Evidence From an Active Controlled Randomized Clinical Trial

Simon B. Goldberg, Joseph Wielgosz, Cortland Dahl, Brianna Schuyler, Donal S. MacCoon, and Melissa Rosenkranz
University of Wisconsin–Madison

Antoine Lutz
University of Wisconsin–Madison and Lyon Neuroscience Research Center

Chad A. Sebranek and Richard J. Davidson
University of Wisconsin–Madison

The current study attempted a rigorous test of the construct validity of a widely used self-report measure of dispositional mindfulness, the Five Facet Mindfulness Questionnaire (FFMQ), within the context of an active controlled randomized trial (n = 130). The trial included three arms: mindfulness-based stress reduction (MBSR), an active control condition that did not include instruction in mindfulness meditation (Health Enhancement Program [HEP]), and a waitlist control condition. Partial evidence for the convergent validity of the FFMQ was shown in correlations at baseline between FFMQ facets and measures of psychological symptoms and psychological well-being. In addition, facets of the FFMQ were shown to increase over the course of an MBSR intervention relative to a waitlist control condition. However, the FFMQ failed to show discriminant validity. Specifically, facets of the FFMQ were shown to increase over the course of the HEP intervention relative to the waitlist control condition. MBSR and HEP, in contrast, did not differ in changes in FFMQ score over time. Implications of these findings for the measurement and theory of mindfulness and MBSR are discussed.

Keywords: mindfulness, mindfulness-based stress reduction, Five Facet Mindfulness Questionnaire, construct validity, measurement

Supplemental materials: http://dx.doi.org/10.1037/pas0000233.supp

Simon B. Goldberg, Center for Investigating Healthy Minds at Waisman Center and Department of Counseling Psychology, University of Wisconsin–Madison; Joseph Wielgosz, Center for Investigating Healthy Minds at Waisman Center, Department of Psychology, and Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin–Madison; Cortland Dahl, Center for Investigating Healthy Minds at Waisman Center and Department of Psychology, University of Wisconsin–Madison; Brianna Schuyler, Center for Investigating Healthy Minds at Waisman Center and Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin–Madison; Donal S. MacCoon, Center for Investigating Healthy Minds at Waisman Center and Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin–Madison; Melissa Rosenkranz, Center for Investigating Healthy Minds at Waisman Center and Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin–Madison; Antoine Lutz, Center for Investigating Healthy Minds at Waisman Center and Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin–Madison and Lyon Neuroscience Research Center; Chad A. Sebranek, Center for Investigating Healthy Minds at Waisman Center and Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin–Madison; Richard J. Davidson, Center for Investigating Healthy Minds at Waisman Center, Department of Psychology, and Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin–Madison.

This work was supported by the National Center for Complementary and Alternative Medicine (NCCAM) P01AT004952 to Richard J. Davidson. NCCAM was not directly involved in study design; collection, analysis, or interpretation of data; in writing of the report; or in the decision to submit the article for publication. Richard J. Davidson serves on the board of directors for the following nonprofit organizations: The Mind and Life Institute and the Center for Investigating Healthy Minds, Inc. No donors, either anonymous or identified, have participated in the design, conduct, or reporting of research results in this article.

Correspondence concerning this article should be addressed to Simon B. Goldberg, Center for Investigating Healthy Minds, 1500 Highland Avenue, Suite S119, Madison, WI 53705-2280. E-mail: sbgoldberg@wisc.edu

In recent decades, mindfulness-based interventions such as mindfulness-based stress reduction (MBSR; Kabat-Zinn, 1990) and mindfulness-based cognitive therapy (Segal, Williams, & Teasdale, 2002) have become increasingly widespread (Khoury et al., 2013; Zoogman, Goldberg, Hoyt, & Miller, 2015). As such, it likewise becomes increasingly important to have a clear articulation of what mindfulness is and how it can be measured (Lutz, Jha, Dunne, & Saron, in press). The most popular strategy for measuring mindfulness has been to employ self-report questionnaires (Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006; Del Re, Flückiger, Goldberg, & Hoyt, 2013).

One of the most widely used self-report measures of mindfulness that has been employed in hundreds of studies is the Five Facet Mindfulness Questionnaire (FFMQ; Baer et al., 2006). In a series of validation studies, Baer et al. (2006) subjected items from five available mindfulness questionnaires to factor analysis, ultimately creating a 39-item measure intended to capture five facets of mindfulness (Acting With Awareness, Describe, Nonjudge, Nonreact, and Observe). The FFMQ was shown to correlate in expected directions with related constructs (e.g., emotion regulation, neuroticism, alexithymia), and FFMQ facets predicted psychological symptoms (Baer et al., 2006). Support for the validity of the FFMQ has included demonstrating associations between the measure and amount of meditation experience, with facets of the FFMQ shown to mediate the link between meditation experience and psychological well-being (Baer et al., 2008). Key evidence for the validity of the FFMQ has also been found in the measure's responsiveness to various forms of mindfulness training, and changes in mindfulness (indexed using the FFMQ and similar self-report measures of mindfulness) have been shown to moderate effect sizes on clinical outcomes in meta-analyses of mindfulness-based interventions (Khoury et al., 2013). Furthermore, amount and quality of mindfulness practice have both been associated with changes in the FFMQ (Goldberg, Del Re, Hoyt, & Davis, 2014).

Despite these promising psychometric properties, some have questioned the validity of self-report measures of mindfulness (Lutz et al., in press). Grossman (2008, 2011) has raised several vital issues, including the importance of defining and measuring mindfulness in Western psychology in a way that does justice to the term's original intention, avoiding biases in measurement development, and recognizing and addressing semantic confusion with item meanings, alongside the typical known biases inherent to self-report methodologies (e.g., social desirability bias). Indeed, self-report measures of dispositional mindfulness have failed to differentiate groups with theoretically distinct levels of mindfulness (e.g., expert meditators vs. controls; Antonova, Chadwick, & Kumari, in press). Item wording (e.g., positive vs. negative) has also been shown to affect response tendencies (Van Dam, Earleywine, & Danoff-Burg, 2009), although perhaps not when samples are demographically matched (e.g., Baer, Samuel, & Lykins, 2011). Based on these and other reports, some have questioned the construct validity of these self-report measures.

Construct Validity: Convergent and Discriminant

Construct validity is perennially a key concern in the development of psychological tests (Crocker & Algina, 2008). In their classic treatise on construct validity, Cronbach and Meehl (1955) note that "construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not 'operationally defined'" (p. 282). The construct of mindfulness arguably remains without an agreed-upon operational definition (although some have begun attempting to operationalize mindfulness with behavioral measures; Levinson, Stoll, Kindy, Merry, & Davidson, 2014).

Most simply, construct validity reflects the extent to which a test measures what it is intended to measure (Crocker & Algina, 2008). Cronbach and Meehl (1955) identify several means by which the construct validity of a psychological test can be established. Of note, the FFMQ has passed many of these tests (at least in some studies): group differences shown between meditating and nonmeditating samples, correlations with related psychological constructs, and adequately fitting internal factor structure (Baer et al., 2006, 2008). Cronbach and Meehl (1955) give particular emphasis to experimental manipulation as a means of determining construct validity, and the FFMQ has been shown to be sensitive to training in mindfulness (Khoury et al., 2013).

However, a key test of the FFMQ's construct validity remains largely ignored, specifically the measure's discriminant validity in the context of experimental manipulation. Although convergent validity involves demonstrating that a measure is associated with theoretically related constructs, discriminant validity involves demonstrating that a measure is *not* associated with theoretically *unrelated* constructs. Campbell and Fiske (1959) identify the importance of discriminant validity, noting that "for the justification of novel trait measures, for the validation of test interpretation, or for the establishment of construct validity, discriminant validation as well as convergent validation is required" (p. 81).

The current study attempts to provide a standard convergent validity test of the FFMQ alongside a rigorous test of its discriminant validity within the context of experimental manipulation. Specifically, we examined changes in FFMQ scores within a randomized controlled trial (RCT) that included a standard MBSR condition, a waitlist (WL) control condition, and an active control condition—Health Enhancement Program (HEP; MacCoon et al., 2012). HEP is a recently designed active control intervention specifically created to control for the nonmindfulness-specific components of MBSR. Unlike MBSR, which teaches mindfulness practices such as sitting meditation and the body scan, HEP includes music therapy, nutrition education, and physical activity. HEP is matched to MBSR in many nonspecific features: both interventions have a group format, meet weekly for 2.5 h, include an "all-day" component, and assign the same amount of home practice (MacCoon et al., 2012). The development of HEP attempted to match with MBSR nonspecific components of treatment (e.g., social support, therapeutic alliance, positive expectancy) proposed to be powerful in facilitating psychological change, including in the context of mindfulness training (Goldberg, Davis, & Hoyt, 2013).

The inclusion of both an active and inactive control condition allows us to address two discriminant validity questions. We can assess directly whether the FFMQ is responsive to change in a mindfulness-based intervention relative to either a WL control or a nonmindfulness-based active intervention (HEP). We hypothesized that as a valid measure of the mindfulness construct, (a) the FFMQ should increase after mindfulness-specific training (i.e., MBSR) relative to WL, and (b) the FFMQ should show a greater

response to mindfulness-specific training than a control intervention composed of nonspecific elements (HEP).

In combination with an experimentally manipulated discriminant validity test, a standard test of the FFMQ's convergent validity was possible through examining associations between this measure and related constructs at baseline. It was hypothesized that the FFMQ would correlate at baseline with measures of psychological well-being and psychological symptoms.

Method

This study used a randomized controlled design that has been described elsewhere (MacCoon et al., 2012). Participants were recruited for a "health and well-being" study and were randomly assigned to a mindfulness intervention, an active control intervention, or a WL control group. The mindfulness intervention consisted of a standard MBSR course (Kabat-Zinn, 1990) and was provided by experienced MBSR instructors. MBSR is an 8-week intervention offered in a group format that includes instruction in both formal (e.g., sitting meditation) and informal (e.g., attentiveness during daily activities) mindfulness practices. These practices are aimed at cultivating a mindful relationship with one's experience, encouraging attentiveness, nonjudgment, and curiosity (Kabat-Zinn, 1990). The active control intervention consisted of the HEP (MacCoon et al., 2012) course and was provided by experienced HEP instructors with no background in mindfulness. HEP was designed specifically to match MBSR as closely as possible while not including mindfulness as an active ingredient. HEP, like MBSR, is administered in a group setting and assigns homework practice. Unlike MBSR, HEP includes mild physical activity, functional movement, nutrition education, diet planning, and music and imagery designed to increase well-being. Both MBSR and HEP were designed to enhance psychological health, while relying on different theoretical pathways of action. Participants randomized to the WL control condition received no intervention during the course of the study. Participants completed a battery of questionnaires and were assessed on a number of physiological measures as part of a larger study of which this experiment was a part. The three measures of interest here are the FFMQ, the Psychological Well-Being Scale (PWB), and the Symptom Checklist (SCL-90R).

In total, 130 participants were initially enrolled in the study (see Consort diagram, Supplemental Figure 1). Of this full sample, 43 were randomized to MBSR, 42 to HEP, and 36 to the control condition. Baseline data were available for 125 and pre-post data for 106. The sample was predominantly female ($n = 79$, 63.2%) with a mean age of 48.05 years ($SD = 10.74$). Participants were excluded if they had used medication for anxiety, depression, or other psychological issues or had a psychiatric diagnosis in the past year (see MacCoon et al., 2012). Participants were also excluded if they had any history of bipolar or schizophrenic disorders, brain damage, or seizures. Baseline psychological symptoms reflect the nonclinical range of functioning (SCL-90R, $M = 0.17$, $SD = 0.18$).

The FFMQ (Baer et al., 2006) served as our measure of trait mindfulness. The measure includes 39 items that are rated on a 1- to 5-point Likert-type scale assessing five facets. As discussed above, the FFMQ has several strong psychometric features (Baer et al., 2008). Internal consistency reliability estimates for the FFMQ in the current sample were adequate: $\alpha = .89, .92, .89, .79, .80$, and .92 for Acting With Awareness, Describe, Nonjudge, Nonreact, Observe, and total scores, respectively. All five subscales as well as a total score composed of all items were used in analyses. A higher score on the FFMQ indicates higher levels of mindfulness.

Psychological well-being was assessed using the 18-item version of the PWB (Ryff, 1989). This measure comprised items from six domains of well-being, including autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance, and has previously shown desirable psychometric properties (Ryff & Keyes, 1995). As a means of data reduction, a total score was computed ($\alpha = .81$), with a higher score indicating higher levels of well-being.

Psychological symptoms were assessed using the SCL-90R (Derogatis, 1994), a widely used symptom checklist. Items focus on specific symptoms associated with various categories of psychological distress. The measure has been shown to possess adequate psychometric properties (Derogatis & Lazarus, 1994) and demonstrated acceptable internal consistency reliability in the current sample ($\alpha = .94$). The global severity index (GSI), with higher scores reflecting greater distress, was used by averaging all 90 items.

As a test of convergent validity, correlations were examined at baseline between the FFMQ and measures of psychological well-being (PWB) and psychological symptoms (SCL-90R), two constructs theoretically linked to mindfulness. Discriminant validity was tested by examining change over the course of time in the context of experimental manipulation (as recommended by Cronbach & Meehl, 1955). Within-group paired t tests were first used to examine pre-post change on the FFMQ for the MBSR, HEP, and WL groups separately. Analysis of covariance models (ANCOVAs) were then used to assess the relative changes in FFMQ across groups, the most direct test of discriminant validity. In these models, posttest FFMQ scores were predicted by group status controlling for pretest FFMQ scores. Three sets of models were constructed comparing MBSR with the WL control, HEP with the WL control, and MBSR with HEP.

Results

At baseline, the FFMQ total score and all subscales showed moderately sized correlations in the expected direction with the PWB (rs = .23–.50, $p < .05$). The FFMQ total score was significantly correlated with the SCL-90R GSI in the expected direction, $r = -.23$, $p = .008$, as was the Nonjudge subscale, $r = -.36$, $p < .001$, although the remaining subscales were not significantly associated with the SCL-90R (see Supplemental Table 1).

Pre- and posttest means and standard deviations for FFMQ scores are reported for each group in Supplemental Table 2 along with results of within-group paired t tests. No significant differences were observed between groups at baseline ($p > .05$). Significant increases in all five FFMQ subscales and total score were noted in the MBSR condition (Cohen's ds = 0.32–0.54, $p < .05$). The HEP group showed increases on the Acting With Awareness, Describe, and Nonjudge subscales as well as on the FFMQ total score (ds = 0.29–0.66, $p < .05$). The WL control condition showed an increase in the Nonjudge subscale as well as the FFMQ total score (ds = 0.61 and 0.24, $p < .05$).

ANCOVA models were used to examine change in FFMQ scores in active groups (MBSR and HEP) relative to the WL condition (i.e., MBSR vs. WL, HEP vs. WL), as well as in active groups relative to one another (i.e., MBSR vs. HEP, Table 1). When controlling for pretest levels, the MBSR group showed increases on the FFMQ relative to the WL group on the Describe, Nonreact, and Observe facets, as well as a marginally significant increase in the total score (*d*s = .44, .52, .56, and .39 for Describe, Nonreact, Observe, and total score, respectively). When controlling for pretest levels, the HEP group showed increases on the FFMQ relative to the WL group on the Describe and Observe subscales and on the total score, as well as marginally significant increases on the Acting With Awareness and Nonreact subscales (*d*s = .47, .41, .53, .38, and .39 for Describe, Observe, total, Acting With Awareness, and Nonreact, respectively).

In contrast, no statistically significant differences were observed between the two intervention groups (MBSR, HEP) in posttest scores controlling for pretest (see Table 1). Cohen's *d*s reflecting these differences ranged in absolute value from .02–.18 (Supplemental Materials).

An anonymous reviewer suggested examining whether the intercorrelation between FFMQ subscales changes in response to intervention. In particular, this reviewer noted the possibility that correlations with the Observe subscale (a subscale with known differences in meditating and nonmeditating samples; Baer et al., 2008) may increase in the MBSR condition relative to HEP. To examine this possibility in a parsimonious way, four regression models were constructed predicting posttreatment Observe scores from the interaction between group (MBSR vs. HEP) and each of the four remaining FFMQ facets (Acting With Awareness, Describe, Nonjudge, and Nonreact) at posttreatment. A significant interaction was detected between group and Acting With Awareness as well as between group and Nonjudge, with a direction indicating a stronger positive association between these subscales and the Observe subscale at posttreatment in the MBSR group relative to the HEP group. This finding provides some evidence that responses on the FFMQ are influenced by group membership, despite the absence of mean differences. Further work in a larger sample could productively examine changes in the intercorrelations of FFMQ subscales (and potentially the measure's factor structure) in response to MBSR versus active comparison conditions.

Discussion

The current study attempted to examine both convergent and discriminant construct validity evidence for a commonly used self-report measure of mindfulness, the FFMQ. Consistent with several previous reports, evidence of convergent validity was seen in moderate-sized positive correlations between all FFMQ subscales and total score with a measure of psychological well-being. In partial support of the measure's convergent validity, the FFMQ total score and Nonjudge subscale were shown to correlate negatively with a measure of psychological symptoms.

No evidence was found for the FFMQ's discriminant validity, however. Discriminant validity was assessed in the FFMQ's sensitivity to specificity of change in one intervention designed to increase mindfulness (MBSR), one intervention specifically designed to have no effect on this construct (HEP), and a WL control condition. Again consistent with previous reports, facets of the FFMQ were shown to increase from pretest to posttest in response to MBSR, both when examined in the MBSR group alone (i.e., paired *t* tests) as well as when examined relative to a WL control condition (further evidence of convergent validity). However, it

Table 1
Between-Group ANCOVAs Contrasting Change in FFMQ in MBSR, HEP, and WL Control Groups

FFMQ subscale	Group	Estimate	*SE*	*t*	*p*	*d*
Act Aware	MBSR vs. WL	.83	.97	.86	.395	.26
Act Aware	HEP vs. WL	1.63	.87	1.88	.064†	.38
Act Aware	MBSR vs. HEP	−1.01	.9	−1.12	.265	−.1
Describe	MBSR vs. WL	1.48	.74	2	.049*	.44
Describe	HEP vs. WL	1.8	.61	2.95	.004**	.47
Describe	MBSR vs. HEP	−.31	.73	−.43	.67	.06
Nonjudge	MBSR vs. WL	.2	1.14	.17	.864	−.03
Nonjudge	HEP vs. WL	1.49	.94	1.59	.117	.12
Nonjudge	MBSR vs. HEP	−1.44	1.06	−1.37	.176	−.13
Nonreact	MBSR vs. WL	2.4	.98	2.44	.017*	.52
Nonreact	HEP vs. WL	1.7	.89	1.91	.060†	.39
Nonreact	MBSR vs. HEP	.69	.81	.86	.395	.18
Observe	MBSR vs. WL	1.71	.85	2.02	.048*	.41
Observe	HEP vs. WL	2.41	.93	2.6	.012*	.56
Observe	MBSR vs. HEP	.65	.9	.72	.473	.18
Total	MBSR vs. WL	4.69	2.75	1.71	.092†	.39
Total	HEP vs. WL	6.56	2.02	3.24	.002**	.53
Total	MBSR vs. HEP	−2.58	2.65	−.98	.333	−.02

Note. The *t* value is from analysis of covariance (ANCOVA) models. Cohen's *d* reflects the standardized mean difference in within-group pre-post difference scores (i.e., MBSR vs. WL, MBSR vs. HEP). MBSR vs. WL with 67 degrees of freedom, HEP vs. waitlist with 69 degrees of freedom, and MBSR vs. HEP with 67 degrees of freedom. *n*s = 34, 36, and 36 for MBSR, HEP, and WL groups, respectively. Group listed second was coded as the reference group in each set of models (e.g., WL in MBSR vs. WL models). Act Aware = Acting with Awareness; FFMQ = Five Facet Mindfulness Questionnaire; MBSR = mindfulness-based stress reduction; HEP = Health Enhancement Program; WL = waitlist control.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

did not appear that this responsiveness was unique to training in mindfulness. Indeed, the HEP condition, designed explicitly with mindfulness not included, likewise showed pre-post increases on facets of the FFMQ, both when examined in the HEP group alone as well as when examined relative to the WL control. Most striking, there were no differences noted in changes on the FFMQ between the MBSR and HEP conditions when compared directly.

The lack of specificity suggested in the current study may not be an outlying occurrence. A recent meta-analysis included $k = 20$ studies that compared a mindfulness intervention with comparison conditions defined as specific active controls (i.e., control interventions "known or expected to change clinical outcomes," p. 359; Goyal et al., 2014). Somewhat surprisingly, only four of these actively controlled studies reported inclusion of a measure of mindfulness.

Of these, none reported differences in response between mindfulness and active control interventions, as measured on the FFMQ, Mindful Attention Awareness Scale (MAAS), or Cognitive Affective Mindfulness Scale - Revised (CAMS-R) (Goyal et al., 2014).

Cronbach and Meehl (1955) address the complex "implications of negative evidence" (i.e., failing to find evidence of construct validity, p. 295). The three possible interpretations they offer include the following: (a) the test does not measure the construct, (b) the theory from which the hypothesis arose is flawed, or (c) the experiment is flawed. It is worth considering these three possibilities in the context of the current study.

Regarding the first interpretation, it is possible that the FFMQ does not truly measure the construct of mindfulness, despite the construct validity previously reported for this measure (e.g., Baer et al., 2008). It is possible that behavioral and physiological measures collected in the current design and currently under analysis could be more sensitive to differentially detect changes in MBSR as opposed to HEP. However, independent of these alternative measures, the failure of a well-validated (and widely used) measure of self-reported mindfulness to differentiate between these two interventions, which are intended to differ precisely in their effects on this construct, is important information. The apparent sensitivity of the FFMQ to changes induced by intervention generally (i.e., HEP) contradicted our initial prediction. This may be due to the FFMQ detecting changes in constructs beyond those intended or, as discussed below, due to the HEP intervention producing changes in constructs beyond those intended.

Cronbach and Meehl's (1955) second and third possibilities implicate our theoretical understanding of mindfulness (Cronbach and Meehl point out that the third possibility may be considered a special case of the second, as may be the case here). The relevant interpretation in the current study would be that our interventions somehow failed to relate as expected to the theoretical construct of mindfulness. This could have occurred in one of two ways—either MBSR failed to induce mindfulness or HEP did induce mindfulness. It seems unlikely that MBSR failed to teach mindfulness. MBSR is a well-studied, structured course explicitly designed to teach the practice and cultivation of mindfulness. Furthermore, instructors in the current study were previously involved in an RCT study that successfully demonstrated that an MBSR intervention on healthy participants protects more than HEP against social stress, as measured by immune and inflammatory biomarkers (Rosenkranz et al., 2013). In addition, effects noted on the FFMQ in the MBSR condition relative to the WL are comparable to those derived via meta-analysis (i.e., $g = .53$ for WL-controlled studies; Khoury et al., 2013).

The possibility that HEP induces mindfulness is intriguing, although also perhaps not the most likely explanation. It is theoretically conceivable that mindfulness, construed as a set of cognitive, affective, and behavioral tendencies toward present-moment awareness, is less dependent on explicit instruction and can be enhanced in more diverse ways than the literature on mindfulness interventions has assumed. Indeed, it may be the case that many psychological interventions increase this mental capacity. In the present design, however, it does not seem unreasonable to suggest that opportunities for learning mindfulness were greater in the MBSR group than in the HEP group. The HEP instructors had no background in mindfulness themselves, nor did the treatment include any implicit or explicit mindfulness instruction. One would therefore not expect HEP to induce mindfulness any more than a standard aerobics class. Yet, if it is indeed the case that HEP induced mindfulness, the FFMQ may be both valid and unable to differentiate changes in MBSR versus HEP.

Limitations of the current work include the sample size that, although sufficient to detect medium-sized effects, does not rule out the presence of small differences in results. Further assessment of changes in mindfulness in large intervention studies, as well as use of multiple converging measures of the mindfulness construct, would therefore be valuable. In addition, in testing mean differences, the current work assumes measurement invariance from pretest to posttest (Kaplan, 2009); a future study with a larger sample could examine this assumption.

In conclusion, the current study points out an important inconsistency between a widely used measure of mindfulness and the standard theories of mindfulness and MBSR within the context of a moderately large, well-controlled randomized trial. A complication with Cronbach and Meehl's (1955) "negative evidence" is that one cannot definitively determine the source of the difficulty. Cronbach and Meehl (1955) therefore conclude, as we must, that "a fresh body of data" (p. 295) is required to further assess validity.

Our results suggest several fertile areas for future work. Future validation of self-report and behavioral measures of mindfulness should be examined in the context of active control comparisons to disentangle mindfulness from related cognitive and affective constructs and processes. It is important to employ control conditions that contain specific active ingredients intended to be therapeutic and to assess mindfulness in these studies (Goyal et al., 2014). Validity of self-report measures should be further assessed by incorporating convergent measures of the mindfulness construct—behavioral tasks (Levinson et al., 2014); significant other-ratings, observer ratings, and teacher ratings; and non-self-report measures of theoretically related constructs (e.g., emotion and attention regulation). More broadly, it will be important to better characterize whether and how trait mindfulness is influenced by nonmindfulness-based psychological interventions (e.g., psychotherapy). As the current study employed a nonclinical population, it would be important to assess whether the discriminant validity of the FFMQ is detected in a clinical sample. It may also be of interest to explore whether the interactions between FFMQ subscales that have been shown to relate to other constructs (Eisenlohr-Moul, Walsh, Charnigo, Lynam, & Baer, 2012; Peters,

Eisenlohr-Moul, Upton, & Baer, 2013) function differently following mindfulness training relative to active comparison groups.

References

Antonova, E., Chadwick, P., & Kumari, V. (2015). More meditation, less habituation? The effect of intensive mindfulness practice on the acoustic startle reflex. *PLoS ONE, 10,* e0123512. http://dx.doi.org/10.1371/journal.pone.0123512

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

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. http://dx.doi.org/10.1177/1073191105283504

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

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

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, OH: Cengage Learning.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. http://dx.doi.org/10.1037/h0040957

Del Re, A. C., Flückiger, C., Goldberg, S. B., & Hoyt, W. T. (2013). Monitoring mindfulness practice quality: An important consideration in mindfulness practice. *Psychotherapy Research, 23,* 54–66. http://dx.doi.org/10.1080/10503307.2012.729275

Derogatis, L. R. (1994). *SCL-90-R (Symptom Checklist-90-R): Administration, scoring, and procedures manual*. Minnetonka, MN: National Computer Systems.

Derogatis, L. R., & Lazarus, L. (1994). SCL-90, Brief Symptom Inventory, and matching clinical rating scales. In M. E. Maruish (Ed.), *The use of psychological testing in treatment planning and outcome assessment* (pp. 217–248). Hillsdale, NJ: Erlbaum.

Eisenlohr-Moul, T. A., Walsh, E. C., Charnigo, R. J., Jr., Lynam, D. R., & Baer, R. A. (2012). The "what" and the "how" of dispositional mindfulness: Using interactions among subscales of the Five-Facet Mindfulness Questionnaire to understand its relation to substance use. *Assessment, 19,* 276–286. http://dx.doi.org/10.1177/1073191112446658

Goldberg, S. B., Davis, J. M., & Hoyt, W. T. (2013). The role of therapeutic alliance in mindfulness interventions: Therapeutic alliance in mindfulness training for smokers. *Journal of Clinical Psychology, 69,* 936–950. http://dx.doi.org/10.1002/jclp.21973

Goldberg, S. B., Del Re, A. C., Hoyt, W. T., & Davis, J. M. (2014). The secret ingredient in mindfulness interventions? A case for practice quality over quantity. *Journal of Counseling Psychology, 61,* 491–497. http://dx.doi.org/10.1037/cou0000032

Goyal, M., Singh, S., Sibinga, E. M., Gould, N. F., Rowland-Seymour, A., Sharma, R., . . . Haythornthwaite, J. A. (2014). Meditation programs for psychological stress and well-being: A systematic review and meta-analysis. *Journal of the American Medical Association Internal Medicine, 174,* 357–368. http://dx.doi.org/10.1001/jamainternmed.2013.13018

Grossman, P. (2008). On measuring mindfulness in psychosomatic and psychological research. *Journal of Psychosomatic Research, 64,* 405–408. http://dx.doi.org/10.1016/j.jpsychores.2008.02.001

Grossman, P. (2011). Defining mindfulness by how poorly I think I pay attention during everyday awareness and other intractable problems for psychology's (re)invention of mindfulness: Comment on Brown et al. *Psychological Assessment, 23,* 1034–1040. http://dx.doi.org/10.1037/a0022713

Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness*. New York, NY: Delta.

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.). Thousand Oaks, CA: Sage.

Khoury, B., Lecomte, T., Fortin, G., Masse, M., Therien, P., Bouchard, V., . . . Hofmann, S. G. (2013). Mindfulness-based therapy: A comprehensive meta-analysis. *Clinical Psychology Review, 33,* 763–771. http://dx.doi.org/10.1016/j.cpr.2013.05.005

Levinson, D. B., Stoll, E. L., Kindy, S. D., Merry, H. L., & Davidson, R. J. (2014). A mind you can count on: Validating breath counting as a behavioral measure of mindfulness. *Frontiers in Psychology, 5,* 1202.

Lutz, A., Jha, A. P., Dunne, J. D., & Saron, C. D. (in press). Investigating the phenomenological matrix of mindfulness-related practices from a neurocognitive perspective. *American Psychologist*.

MacCoon, D. G., Imel, Z. E., Rosenkranz, M. A., Sheftel, J. G., Weng, H. Y., Sullivan, J. C., . . . Lutz, A. (2012). The validation of an active control intervention for Mindfulness Based Stress Reduction (MBSR). *Behaviour Research and Therapy, 50,* 3–12. http://dx.doi.org/10.1016/j.brat.2011.10.011

Peters, J. R., Eisenlohr-Moul, T. A., Upton, B. T., & Baer, R. A. (2013). Nonjudgment as a moderator of the relationship between present-centered awareness and borderline features: Synergistic interactions in mindfulness assessment. *Personality and Individual Differences, 55,* 24–28. http://dx.doi.org/10.1016/j.paid.2013.01.021

Rosenkranz, M. A., Davidson, R. J., Maccoon, D. G., Sheridan, J. F., Kalin, N. H., & Lutz, A. (2013). A comparison of mindfulness-based stress reduction and an active control in modulation of neurogenic inflammation. *Brain, Behavior, and Immunity, 27,* 174–184. http://dx.doi.org/10.1016/j.bbi.2012.10.013

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57,* 1069–1081. http://dx.doi.org/10.1037/0022-3514.57.6.1069

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69,* 719–727. http://dx.doi.org/10.1037/0022-3514.69.4.719

Segal, Z., Williams, J. W., & Teasdale, J. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. New York, NY: Guilford Press.

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. http://dx.doi.org/10.1016/j.paid.2009.05.005

Zoogman, S., Goldberg, S. B., Hoyt, W. T., & Miller, L. (2015). Mindfulness interventions with youth: A meta-analysis. *Mindfulness, 6,* 290–302. http://dx.doi.org/10.1007/s12671-013-0260-4

Received June 22, 2015
Revision received August 17, 2015
Accepted August 18, 2015 ■