Mindfulness
https://doi.org/10.1007/s12671-019-01235-2

ORIGINAL PAPER

Evaluating the Factor Structure of Each Facet of the Five Facet Mindfulness Questionnaire

William E. Pelham III[1] · Oscar Gonzalez[2] · Stephen A. Metcalf[3] · Cady L. Whicker[3] · Katie Witkiewitz[4] · Lisa A. Marsch[3] · David P. Mackinnon[1]

Abstract

Objective Nearly all studies treat the Five Facet Mindfulness Questionnaire as five independent scales (one measuring each of the five facets), yet almost no methodological work has examined the psychometric structure of the facets independently. We address this gap using factor analytic methods.

Methods Exploratory and confirmatory factor models were fit to item response data from a sample of 522 adults recruited online. Findings were replicated in a sample of 454 adults receiving aftercare for substance use disorder.

Results The parallel analysis suggested multiple factors for all five facets, in both samples. Exploratory factor models suggested the presence of method factors on the acting with awareness (items using the term "distraction") and describing facets (items that were reverse-scored). Confirmatory factor models fit poorly for all facets, in both samples. In follow-up analyses, model fit improved substantially on the acting with awareness and describing facets when method factors were included in a bifactor model. Model fit was also better for the facets of FFMQ short forms than for the full-length facets. The short-form facets and original facets correlated similarly with external criteria in both samples.

Conclusions None of the FFMQ facets fit a unidimensional factor model; yet, follow-up analyses suggested that each can be considered substantively unidimensional. Initial tests suggest that the facets' multidimensionality did not materially impact their relation to other psychological constructs, suggesting that multidimensionality can be ignored for some purposes. The short-form facets and latent variable models (e.g., bifactor specifications) are both viable solutions for addressing multidimensionality when desired.

Keywords Mindfulness · Dimensionality · Factor analysis

The Five Facet Mindfulness Questionnaire (FFMQ; Baer et al. 2006) is one of the most popular self-report questionnaires for measuring mindfulness, having been cited more than 2100 times (per Web of Science on August 26, 2019). The FFMQ conceptualizes mindfulness as having five different dimensions—acting with awareness, describing, nonjudging, nonreactivity, and observing. In accordance with this conceptualization, almost all studies treat the FFMQ as five independent scales (one measuring each of the five facets) rather than as a unified, 39-item form. However, we are aware of only one prior investigation of the psychometric structure of the individual facets (Tran et al. 2013).

The FFMQ was originally developed by factor-analyzing a total of 112 items from five different existing mindfulness questionnaires: the Mindful Attention Awareness Scale (MAAS; Brown and Ryan 2003), the Freiburg Mindfulness Inventory (FMI; Buchheld et al. 2001), the Kentucky Inventory of Mindfulness Skills (KIMS; Baer et al. 2004), the Cognitive and Affective Mindfulness Scale (CAMS;

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12671-019-01235-2) contains supplementary material, which is available to authorized users.

✉ William E. Pelham, III
wp8000@gmail.com

[1] Department of Psychology, Arizona State University, Tempe, AZ 85281, USA

[2] Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[3] Department of Psychiatry, Dartmouth College, Lebanon, NH 03766, USA

[4] Center on Alcoholism, Substance Abuse, and Addictions, University of New Mexico, Albuquerque, NM 87131, USA

Feldman et al. 2004), and the Mindfulness Questionnaire (MQ; Chadwick et al., 2008). All five questionnaires were administered to a sample of 613 undergraduate students. A five-factor solution was chosen in exploratory factor analysis (EFA), with each factor corresponding to a facet on the resulting FFMQ. The authors retained the eight items with the highest factor loadings for each of the acting with awareness, describing, nonjudging, and observing facets; only seven items loaded on the nonreactivity factor. A confirmatory factor analysis (CFA) with five intercorrelated factors exhibited acceptable fit in a replication sample of 268 undergraduate students (root mean square error of approximation [RMSEA] = 0.06, comparative fit index [CFI] = 0.96).

Since its initial publication, the five-factor psychometric structure of the FFMQ has been replicated in many samples, and in diverse contexts (e.g., Christopher et al. 2012; Curtiss and Klemanski 2014; Veehof et al. 2011; Williams et al. 2014). However, nearly all psychometric work has evaluated the scale as a unified, 39-item form (e.g., in a five-factor CFA), contrary to how the FFMQ is typically used. Most investigators treat the FFMQ facets as *five separate scales* that can be used to test how different dimensions of the mindfulness construct may play different roles in the psychological phenomenon of interest. For example, Clerkin et al. (2017) tested each of the five facets in a separate model when examining how mindfulness may mediate the relation between social anxiety and drinking problems, reporting findings for only the acting with awareness, nonjudging, and describing facets. To characterize the frequency of this practice, we reviewed all articles citing the FFMQ that were published in 2017 in the journal *Mindfulness*. 33 of 36 articles (92%) used one or more facet scores as separate variables in an analysis, and only 15 of 36 articles (42%) used a total score (i.e., created by adding together more than one facet) in analysis (see supplement for article coding).

If the facets are to be used as independent scales, then it will be important to evaluate the psychometric properties of each scale separately. No evidence is provided in the initial FFMQ publication that the facets independently display a good fit to a unidimensional factor model (Baer et al. 2006). In fact, because the FFMQ was created by factor-analyzing several existing scales and retaining items with the highest factor loadings, it might be especially vulnerable to psychometric issues such as retention of redundant items and narrow construct definition (Podsakoff, MacKenzie, & Podsakoff, 2012; Smith et al. 2000). Because the items on different scales were written by different authors and at different points in time, method factors (e.g., item wording) may be confounded with substantive factors (e.g., the nonjudging construct) when the items are analyzed together in the same factor analyses. Moreover, item-level issues may have been obscured during scale development because the items were parceled (i.e., grouped together and then averaged) before factor analyses (Christopher et al. 2012; Little et al. 2002).

Three lines of evidence suggest the FFMQ facets may not each be unidimensional. First, van Dam et al. (2012) found that adding method factors for item valence (i.e., positive vs. negative valence) significantly improved the fit of the FFMQ in confirmatory factor analyses. Thus, the facets that contain a mix of positive and negative items may not fit a unidimensional model. Second, an inspection of the FFMQ items (Table 1) suggests the presence of item duplication and items with shared wording, both of which can compromise unidimensionality. The acting with awareness facet provides an example of item duplication. The text of item 13 ("I am easily distracted") is completely subsumed within the text of item 5 ("When I do things, my mind wanders off and *I'm easily distracted*"). Such duplicative items may have been introduced to the FFMQ because its items were selected solely on the basis of maximizing factor loadings in a large EFA (Baer et al. 2006), rather than written and selected to fit well together. If an item pool contains duplicative items, the duplicative items may correlate highly and take over the factor definition, obtain high factor loadings, and thus be selected for retention. The nonreactivity facet provides an example of items with shared wording. Four of its items were drawn from the FMI and begin with the identical stem, "When I have distressing thoughts or images…", but the remaining three items were drawn from the MQ and do not contain this phrase. When items are duplicated or share wording, the variance common across items will be explained not only by the latent factor (i.e., mindfulness) but also by extraneous method factors. Indeed, local dependence (Yen, 1993), or the covariance of items for reasons other than the common factor, has been found on the FFMQ in multiple samples (Medvedev et al. 2017; Tran et al. 2013).

Third, the only published evaluation of the psychometric structure of the individual FFMQ facets yielded a pattern of results consistent with the presence of unmodeled method factors. In the process of developing a short form, Tran et al. (2013) fit unidimensional factor models to each of the five facets in an Austrian community sample ($N = 640$). Model fit for three of the five facets met conventional guidelines for the CFI (all above 0.96) but not the RMSEA (ranging from 0.07 to 0.12) (Lai and Green 2016). However, fit for both the acting with awareness and nonreactivity facets was unacceptable, with RMSEAs above 0.10 and CFIs below 0.90. The fit could be improved to an acceptable level only by correlating a series of residual item variances (i.e., addressing items with similar content and wording). If replicated, such a result would suggest that these two facets of the FFMQ may reflect more than one underlying construct.

In summary, almost all applications of the FFMQ treat the five facets as separate scales, but almost no psychometric work has evaluated them as separate scales. The only published evaluation of their structure as independent scales indicated that multiple facets fit the unidimensional factor model t1.1 **Table 1** Descriptive statistics for FFMQ item responses in the primary sample

	Facet	Item	OnFFMQ-24	OnFFMQ-23	Source	Mean	SD	% per response value	Item label
t1.2									
t1.3	Acting with awareness	5		✓	KIMS	3.42	1.08	0.06/0.15/0.27/0.37/0.15	When I do things, my mind wanders off and I'm easily distracted
t1.4		8		✓	KIMS	3.86	0.96	0.01/0.07/0.25/0.38/0.29	I do not pay attention to what I'm doing because I'm daydreaming, worrying, or otherwise…
t1.5		13		✓	CAMS	3.52	1.10	0.05/0.13/0.26/0.36/0.20	I am easily distracted
t1.6		18	✓	✓	MAAS	3.82	0.98	0.01/0.09/0.24/0.38/0.28	I find it difficult to stay focused on what's happening in the present
t1.7		23	✓		MAAS	3.78	1.00	0.02/0.10/0.24/0.39/0.26	It seems I am 'running on automatic' without much awareness of what I'm doing
t1.8		28	✓		MAAS	4.03	0.90	0.01/0.05/0.21/0.39/0.35	I rush through activities without being really attentive to them
t1.9		34	✓		MAAS	3.82	0.94	0.01/0.07/0.28/0.38/0.27	I do jobs or tasks automatically without being aware of what I'm doing
t1.10		38	✓		MAAS	3.86	0.98	0.02/0.07/0.23/0.39/0.29	I find myself doing things without paying attention
t1.11	Describing	2	✓		KIMS	3.48	1.10	0.05/0.14/0.28/0.34/0.19	I'm good at finding words to describe my feelings
t1.12		7	✓		KIMS	3.67	1.02	0.02/0.11/0.26/0.38/0.23	I can easily put my beliefs, opinions, and expectations into words
t1.13		12	✓		KIMS	3.74	1.08	0.04/0.10/0.20/0.39/0.26	It's hard for me to find the words to describe what I'm thinking
t1.14		16		✓	KIMS	3.68	1.11	0.04/0.13/0.21/0.36/0.26	I have trouble thinking of the right words to express how I feel about things
t1.15		22	✓	✓	KIMS	3.82	1.01	0.02/0.10/0.20/0.40/0.28	When I have a sensation in my body, it's difficult for me to describe it because I cannot…
t1.16		27	✓		KIMS	3.41	1.10	0.06/0.14/0.29/0.35/0.16	Even when I'm feeling terribly upset, I can find a way to put it into words
t1.17		32		✓	KIMS	3.27	1.10	0.06/0.20/0.29/0.33/0.13	My natural tendency is to put my experiences into words
t1.18		37		✓	CAMS	3.37	1.14	0.07/0.16/0.27/0.34/0.16	I can usually describe how I feel at the moment in considerable detail
t1.19	Nonjudging	3			KIMS	3.50	1.18	0.06/0.15/0.25/0.31/0.24	I criticize myself for having irrational or inappropriate emotions
t1.20		10	✓		KIMS	3.50	1.06	0.03/0.15/0.30/0.33/0.19	I tell myself I should not be feeling the way I'm feeling
t1.21		14		✓	KIMS	3.86	1.07	0.02/0.11/0.22/0.31/0.35	I believe some of my thoughts are abnormal or bad and I should not think that way
t1.22		17	✓		KIMS	3.27	1.13	0.05/0.22/0.30/0.26/0.16	I make judgments about whether my thoughts are good or bad
t1.23		25	✓	✓	KIMS	3.57	1.05	0.02/0.15/0.26/0.35/0.20	I tell myself that I should not be thinking the way I'm thinking
t1.24		30	✓	✓	KIMS	3.75	1.06	0.02/0.12/0.21/0.36/0.28	I think some of my emotions are bad or inappropriate and I should not feel them
t1.25		35		✓	MQ	3.65	1.09	0.02/0.15/0.28/0.28/0.28	When I have distressing thoughts or images, I judge myself as good or bad, depending what…
t1.26		39	✓		KIMS	3.42	1.14	0.04/0.20/0.27/0.28/0.21	I disapprove of myself when I have irrational ideas
t1.27	Nonreactivity	4			FMI	3.18	0.93	0.05/0.16/0.43/0.31/0.06	I perceive my feelings and emotions without having to react to them
t1.28		9	✓	✓	FMI	3.24	0.94	0.05/0.13/0.41/0.34/0.07	I watch my feelings without getting lost in them
t1.29		19	✓	✓	MQ	3.16	0.99	0.07/0.16/0.39/0.32/0.07	When I have distressing thoughts or images, I 'step back' and am aware of the thought or image…
t1.30		21		✓	FMI	3.40	0.94	0.03/0.12/0.37/0.37/0.11	In difficult situations, I can pause without immediately reacting
t1.31		24	✓	✓	MQ	3.02	1.00	0.08/0.21/0.38/0.29/0.05	When I have distressing thoughts or images, I feel calm soon after
t1.32		29	✓		MQ	3.16	0.92	0.05/0.14/0.46/0.28/0.06	

Table 1 (continued)

Facet	Item	OnFFMQ-24	OnFFMQ-23	Source	Mean	SD	% per response value	Item label
								When I have distressing thoughts or images, I am able just to notice them without reacting
	33	✓		MQ	3.05	0.96	0.07/0.18/0.45/0.25/0.06	When I have distressing thoughts or images, I just notice them and let them go
Observing	1			KIMS	2.84	1.04	0.11/0.26/0.37/0.22/0.05	When I'm walking, I deliberately notice the sensations of my body moving
	6			KIMS	3.20	1.10	0.06/0.20/0.34/0.26/0.13	When I take a shower or bath, I stay alert to the sensations of water on my body
	11			KIMS	2.92	1.11	0.12/0.22/0.35/0.24/0.07	I notice how foods and drinks affect my thoughts, bodily sensations, and emotions
	15	✓	✓	KIMS	3.35	1.02	0.04/0.14/0.36/0.32/0.13	I pay attention to sensations, such as the wind in my hair or sun on my face
	20	✓	✓	KIMS	3.48	1.02	0.04/0.11/0.35/0.33/0.17	I pay attention to sounds, such as clocks ticking, birds chirping, or cars passing
	26	✓	✓	KIMS	3.86	0.96	0.02/0.05/0.25/0.39/0.28	I notice the smells and aromas of things
	31	✓	✓	KIMS	3.58	1.01	0.03/0.09/0.33/0.36/0.19	I notice visual elements in art or nature, such as colors, shapes, textures, or patterns of light…
	36			KIMS	3.47	0.92	0.02/0.12/0.34/0.41/0.11	I pay attention to how my emotions affect my thoughts and behavior

Note. *KIMS*, Kentucky Inventory of Mindfulness Skills; *CAMS*, Cognitive and Affective Mindfulness Scale; *MAAS*, Mindfulness Attention and Awareness Scale; *MQ*, Mindfulness Questionnaire; *FMI*, Freiburg Mindfulness Inventory (see the "Method" section for citations). "Source" indicates which original mindfulness questionnaire the item was drawn from. "% per response value" indicates the percentage of participants responding in the first, second, third, fourth, and fifth category on the response scale. Citations for short forms are as follows: FFMQ-24 (Bohlmeijer et al. 2011); FFMQ-23 (Burzler et al. 2019). Items were reverse-scored as indicated prior to calculating descriptive statistics. $N = 522$ for all items

poorly, and there are several reasons to expect this poor fit to replicate (e.g., presence of nearly duplicate items). If the facets contain unmodeled dimensions or locally dependent items, this can have several negative consequences. First, the latent factor score (or total score) will be skewed toward measuring the items that are locally dependent. When two of eight items measure whether the individual is "easily distracted," for example, then this concept will be overrepresented in the resulting factor or total score, relative to other concepts reflected in the item pool. Second, the loadings across items will no longer reflect the true relationship of the items with the construct. Typically, loadings for locally dependent items will be positively biased, and loadings for other items will be negatively biased. Third, reliability coefficients (e.g., alpha) will be artificially inflated by the inclusion of common variance explained by the factors unrelated to the underlying construct of interest (e.g., unmodeled dimension for item wording similarity). Finally, relations of the measure with external criteria can be either positively or negatively biased (e.g., Neal and Carey 2005). Thus, the presence of unmodeled dimensions or locally dependent items on the FFMQ facets could lead to overestimates of reliability, narrow construct definition, inaccurate relations of items to the overall construct, and biased relations of the facet to external criteria.

The purpose of this study was to investigate the psychometric structure of the individual facets of the FFMQ using both confirmatory and exploratory methods. We applied all psychometric analyses in two different samples: adults recruited via Amazon Mechanical Turk ($N = 522$) and individuals receiving aftercare treatment for substance use disorder ($N = 456$). We evaluated whether each facet displayed sound, unidimensional structure, and attempted to remedy any model misfit that was revealed.

Method

Participants

Primary Sample (MTurk) The primary sample consisted of 522 adults living in the USA who completed the FFMQ on Amazon Mechanical Turk (MTurk). These adults were participating in a larger study involving the administration of many different measures of self-regulation (Eisenberg et al. 2018, 2019; Enkavi et al. 2019). These 522 participants passed a number of quality checks to promote valid responding (Eisenberg et al. 2018), and there were no missing data. The mean age of participants was 34 years ($SD = 8$, range = [20, 59]) and most were Caucasian (86%). Approximately half were female (51%) and nearly half had a college degree (44%).

Replication Sample (Clinical) The replication sample was created by combining data from two trials of mindfulness-based relapse prevention for individuals with substance use disorders (Bowen et al. 2009, 2014). Participants in both trials were randomized to different aftercare conditions after completing inpatient or intensive outpatient treatment. The FFMQ was completed prior to randomization, and thus prior to receipt of mindfulness-based relapse prevention training. The first study (Bowen et al. 2009) included 168 adults with a mean age of 41 years ($SD = 10$). 64% of participants were male, 54% were non-Hispanic white, 30% were African American, 15% were Native American, 5% were Hispanic or Latino/a, 41% were unemployed, and 72% had a high school degree. The second study (Bowen et al. 2014) included 286 adults with a mean age of 38 years ($SD = 11$). 75% of participants were male, 53% were non-Hispanic white, 25% were African American, 6% were Native American, 7% were Hispanic or Latino/a, 66% were unemployed, and 66% had a high school degree. These two studies were pooled into a single dataset in order to produce a large replication dataset. 324 of 454 participants (71%) had complete data across all 39 items, and no item had more than 4% of responses missing.

Procedures

In the primary (MTurk) sample, participants completed the FMMQ online using the Experiment Factory platform (Sochat et al., 2016). In the replication (clinical) sample, participants completed the FFMQ on a web-based survey platform (DatStat Illume, DatStat, Incorporated, Seattle, WA). In both samples, the FFMQ was one questionnaire within a larger battery of completed measures.

Measures

The FFMQ consists of 39 items, with seven items measuring nonreactivity and eight items measuring each of the other four facets. Table 1 lists the items, indicates which of the five facets they are supposed to load on, and reports descriptive statistics in the primary sample. Respondents rate each item on a scale from 1 (*never or rarely true*) to 5 (*very often or always true*). Total scores on a facet are calculated by summing item responses after reverse-scoring items that indicate lower, rather than higher, levels of mindfulness.

Data Analyses

In preliminary analyses, we fit a series of factor models to the *entire* FFMQ to ensure that the five-factor psychometric structure observed in the larger literature on the FFMQ replicated in these samples. Next, we fit a series of factor models to *each facet* of the FFMQ in order to examine the psychometric structure of each facet separately.

Confirmatory Factor Models Confirmatory factor analyses were fit in *Mplus* version 7.4 (Muthén and Muthén 2015). Although the FFMQ items are categorical (i.e., have discrete response categories), much of the previous literature has analyzed them as if continuous. To facilitate comparison to existing literature, we report results for these analyses both (a) when modeling items as categorical and (b) when modeling items as continuous. The weighted least squares mean and variance adjusted (WLSMV) estimator was used when modeling items as categorical. The maximum likelihood (ML) estimator was used when modeling items as continuous. For identification, means of factors were fixed to zero, variances of factors were fixed to 1, and factor intercorrelations, factor loadings, and item intercepts (or thresholds) were freely estimated.

For each model, we report several fit statistics: the chi-square test of model fit, the root mean square error of approximation (RMSEA), the comparative fit index (CFI), the weighted root mean square residual (WRMR), and the standardized root mean square residual (SRMR). Standard recommendations are that a well-fitting model is indicated by a non-significant chi-square statistic (i.e., $p > .05$), an RMSEA below 0.06 (Hu and Bentler 1999), a CFI above 0.95 (Hu and Bentler 1999), a WRMR below 1.0 (Yu, 2002), and an SRMR below 0.08 (Hu and Bentler 1999). These cutoffs are not strict requirements but rather benchmarks against which to compare our results (Millsap 2007; West and Taylor 2015).

Exploratory Factor Models Parallel analysis (Horn 1965) was used to evaluate the dimensionality of each facet. This procedure consists of comparing the observed eigenvalues in the dataset to a set of reference eigenvalues simulated under the null hypothesis that the items are unrelated to each other (i.e., there are no common factors). We used bootstrapping to simulate the reference eigenvalues (B = 100; Revelle 2017), so standard errors for the random references eigenvalues could be calculated using the standard deviation of each eigenvalue across bootstrap resamples. The number of factors was then determined in two ways (Crawford et al. 2010). First, we counted the number of observed eigenvalues that exceeded the corresponding mean eigenvalues observed across resamples ("mean eigenvalue criterion"). Second, we counted the number of observed eigenvalues that exceeded the 95th percentile of the resamples ("95th percentile eigenvalue criterion"). The latter is a more conservative procedure in that it requires stronger evidence for the presence of additional factors. All models were fit to the polychoric correlation matrix.

Next, exploratory factor analysis was conducted using the number of factors indicated by the results of the parallel analysis procedure. This model was fit to the polychoric correlation matrix using oblimin rotation. Both parallel analysis and exploratory factor analysis models were conducted using the *psych* package (Revelle 2017) in the R statistical software environment (v3.5.3) (R Core Team, 2019).

Attempts to Improve Fit When misfit was detected, we considered two strategies to improve fit. First, we examined the fit of an alternative model specification: bifactor models. Bifactor models can improve fit by incorporating orthogonal factors that account for method effects (Chen et al. 2006). The model specification was guided by inspection of the items and the results of the EFA and CFA.

Second, we examined the fit of unidimensional models for short versions of the FFMQ facets. Short versions of the facets may improve fit by dropping items that contribute to misfit. We examined the short versions defined by two previously developed short forms that differed in their item selection: the FFMQ-24 (Bohlmeijer et al. 2011) and the FFMQ-23 (Burzler et al. 2019). Table 1 indicates which items are present on each form. These forms shared 15 items in common; the FFMQ-24 includes nine items not present on the FFMQ-23 and the FFMQ-23 includes eight items not present on the FFMQ-24.

If the use of short-form facets improves fit to a unidimensional model, it is important to also demonstrate that this improvement is obtained without materially affecting the facets' relation to other psychological constructs. This was investigated by collecting extrinsic convergent validity evidence (Fiske 1971; Lubinski, 2004) comparing (a) the correlation of a short-form facet with an external criterion to (b) the correlation of the full-length facet with that same external criterion. To be comprehensive, we tested the difference in these dependent correlations with multiple external criteria. In the primary sample, the external criteria were cigarettes smoked per day, risky alcohol use, psychological distress, self-control, behavioral inhibition, sensation seeking, future time perspective, conscientiousness, and openness (see Table 6). In the replication sample, the external criteria were craving, dependence severity, drinking problems, and acceptance (see Table S4). For each criterion, we used the *cocor* package (Diedenhofen & Musch, 2015) to compare the dependent correlations using the Williams (1959) test and characterize the size of the difference using Cohen's q (Cohen 1988). All p-values were adjusted for multiple testing using the False Discovery Rate correction (Benjamini and Hochberg 1995).

Results

In each section, we first report those obtained in the primary sample then compare them to those obtained in the replication sample.

Preliminary Analyses

Before proceeding to one-facet models, preliminary analyses verified that the five-factor structure of the full FFMQ found in previous literature was also observed in these samples. Confirmatory factor analyses indicated an acceptable fit of the five-factor model in both the primary and replication sample (Table 2). When items were modeled as continuous (ML estimator), in the primary sample the RMSEA was 0.066, CFI was 0.885, and SRMR was 0.067. The fit was worse in the replication sample (RMSEA = 0.069, CFI = 0.791, SRMR = 0.097). Although not always meeting recommended cutoffs, fit statistic values were similar to those obtained in past work, with CFI in particular being slightly lower (e.g., Baer et al. 2006: RMSEA = 0.06 and CFI = 0.96; Veehof et al. 2011: RMSEA = 0.080, CFI = 0.908, SRMR = 0.098; Bohlmeijer et al. 2011: RMSEA = 0.074, CFI = 0.914, SRMR = 0.073; Christopher et al. 2012: RMSEA = 0.06, CFI = 0.97, SRMR = 0.09; Williams et al. 2014: RMSEA = 0.074, CFI = 0.94 SRMR = 0.058). Parallel analysis suggested a five-factor solution in both samples, matching the expected dimensionality of the questionnaire (see supplement). Exploratory factor analyses with five factors generally recovered the familiar five facets (see supplement). In summary, both confirmatory and exploratory factor analyses replicated previous work in supporting a five-factor model when the FFMQ was analyzed as a whole.

One-Facet Models

We next examined the psychometric structure of each of the five facets separately.

Parallel Analysis Figure 1 shows the observed and resampled eigenvalues for each facet in each sample (see Fig. S1 for a zoomed-out version). Under the mean eigenvalue criterion, the parallel analysis suggested a 2-factor solution for acting with awareness, 2-factor solution for describing, 2-factor solution for nonjudging, 4-factor solution for nonreactivity, and 3-factor solution for observing. Under the more conservative 95th percentile eigenvalue criterion, the parallel analysis suggested a 2-factor solution for acting with awareness, a 2-factor solution for describing, a 1-factor solution for nonjudging, a 2-factor solution for nonreactivity, and 3-factor solution for observing. The first eigenvalues extracted in parallel analysis were 13 to 33 times larger than the second eigenvalues, which were uniformly small (0.17 to 0.42).

Exploratory Factor Analysis Exploratory factor models were fit to each facet using the number of factors identified using the more conservative 95th percentile eigenvalue criterion. Results are shown in Table 3. For some facets, the factors that emerged beyond the first factor were substantively interpretable. For example, on the describing facet, the first factor was indicated by the five positively worded items, and the second factor was indicated by the three negatively worded items (i.e., there was a method effect for item valence). On the acting with awareness facet, the first factor was mostly indicated by items invoking the concepts of awareness and attention, and the

Table 2 Fit of confirmatory factor models

Estimator	Sample	Model	# params	χ^2	df	χ^2/df	$p(\chi^2$,df)	RMSEA	CFI	WRMR	SRMR
WLSMV	Primary sample (MTurk)	5-Factor	205	2707	692	3.9	$p<.001$	0.075	0.938	1.908	–
		1-Factor: acting with awareness	40	516	20	25.8	$p<.001$	0.218	0.964	2.209	–
		1-Factor: describing	40	463	20	23.1	$p<.001$	0.206	0.969	1.674	–
		1-Factor: nonjudging	40	208	20	10.4	$p<.001$	0.134	0.986	1.084	–
		1-Factor: nonreactivity	35	174	14	12.4	$p<.001$	0.148	0.945	1.280	–
		1-factor: observing	40	221	20	11.1	$p<.001$	0.139	0.964	1.344	–
	Replication sample (clinical)	5-Factor	205	3628	692	5.2	$p<.001$	0.097	0.753	2.562	–
		1-Factor: acting with awareness	40	368	20	18.4	$p<.001$	0.197	0.915	1.835	–
		1-Factor: describing	40	689	20	34.5	$p<.001$	0.273	0.804	2.565	–
		1-Factor: nonjudging	40	131	20	6.5	$p<.001$	0.111	0.962	1.054	–
		1-Factor: nonreactivity	35	54	14	3.8	$p<.001$	0.079	0.970	0.762	–
		1-Factor: observing	40	181	20	9	$p<.001$	0.134	0.938	1.265	–
ML	Primary sample (MTurk)	5-Factor	127	2242	692	3.2	$p<.001$	0.066	0.885	–	0.067
		1-Factor: acting with awareness	24	393	20	19.6	$p<.001$	0.189	0.878	–	0.055
		1-Factor: describing	24	212	20	10.6	$p<.001$	0.136	0.939	–	0.040
		1-Factor: nonjudging	24	109	20	5.4	$p<.001$	0.092	0.971	–	0.027
		1-Factor: nonreactivity	21	97	14	6.9	$p<.001$	0.106	0.932	–	0.041
		1-Factor: observing	24	139	20	7	$p<.001$	0.107	0.936	–	0.043
	Replication sample (clinical)	5-Factor	127	2162	692	3.1	$p<.001$	0.069	0.791	–	0.097
		1-Factor: acting with awareness	24	213	20	10.6	$p<.001$	0.147	0.874	–	0.067
		1-Factor: describing	24	244	20	12.2	$p<.001$	0.158	0.825	–	0.088
		1-Factor: nonjudging	24	84	20	4.2	$p<.001$	0.084	0.943	–	0.039
		1-Factor: nonreactivity	21	30	14	2.2	$p<.01$	0.051	0.973	–	0.030
		1-Factor: observing	24	117	20	5.8	$p<.001$	0.104	0.911	–	0.050

Note. WLSMV, weighted least squares mean and variance adjusted estimator; *ML*, maximum likelihood estimator; *# params*, number of free parameters; χ^2, chi-square statistic of model fit; *df*, degrees of freedom for chi-square statistic; *p*, *p* value for chi-square test of model fit; *RMSEA*, root mean square error of approximation; *CFI*, comparative fit index; *WRMR*, weighted root mean square residual; *SRMR*, standardized root mean square residual. Five-factor models included all 39 items, each loading on their respective facet, and all five factors intercorrelated. One-factor models included only the items on the indicated facet

second factor was mostly indicated by items invoking the concepts of distraction and focus. On the observing facet, the third factor was indicated by the three-item prompts using the word "sensations." For other facets (e.g., nonreactivity), it was difficult to interpret the meaning of the secondary factors.

Confirmatory Factor Analysis Confirmatory factor analyses conducted separately by facet indicated poor model fit for all five facets in the primary sample (Table 2). The chi-square test of model fit was significant for all facets, indicating the model fit was worse than a saturated model. When items were modeled as categorical (WLSMV estimator), the RMSEA (0.075–0.218) and WRMR (1.084–2.209) were above the suggested cutoffs for all facets, while the CFI was above the suggested cutoff for four of five facets (0.945–0.986). When items were modeled as continuous (ML estimator), RMSEA was generally better but still substantially above the suggested cutoff, whereas CFI was generally worse and below the suggested cutoff for four of five facets.

Modification Indices For each facet, more than half of the modification indices were statistically significant (i.e., exceeded 3.84 at 1 *df*), suggesting that correlating the corresponding items' residuals would improve model fit. The largest modification indices linked very similar pairs of items. For example, the largest modification index (MI = 147.35) was observed on the acting with awareness facet, where the text of item number 13 ("I am easily distracted") was completely subsumed within the text of item number 5 ("When I do things, my mind wanders off and *I'm easily distracted*"). The second-largest modification index (MI = 117.64) was observed on the describe facet, for two items that were again very similar in wording and content: "It's hard for me to find the words to describe what I'm thinking" (#12) and "I have trouble thinking of the right words to express how I feel about things" (#16).

Replication Sample Results in the replication sample were generally consistent with those in the primary sample. As in

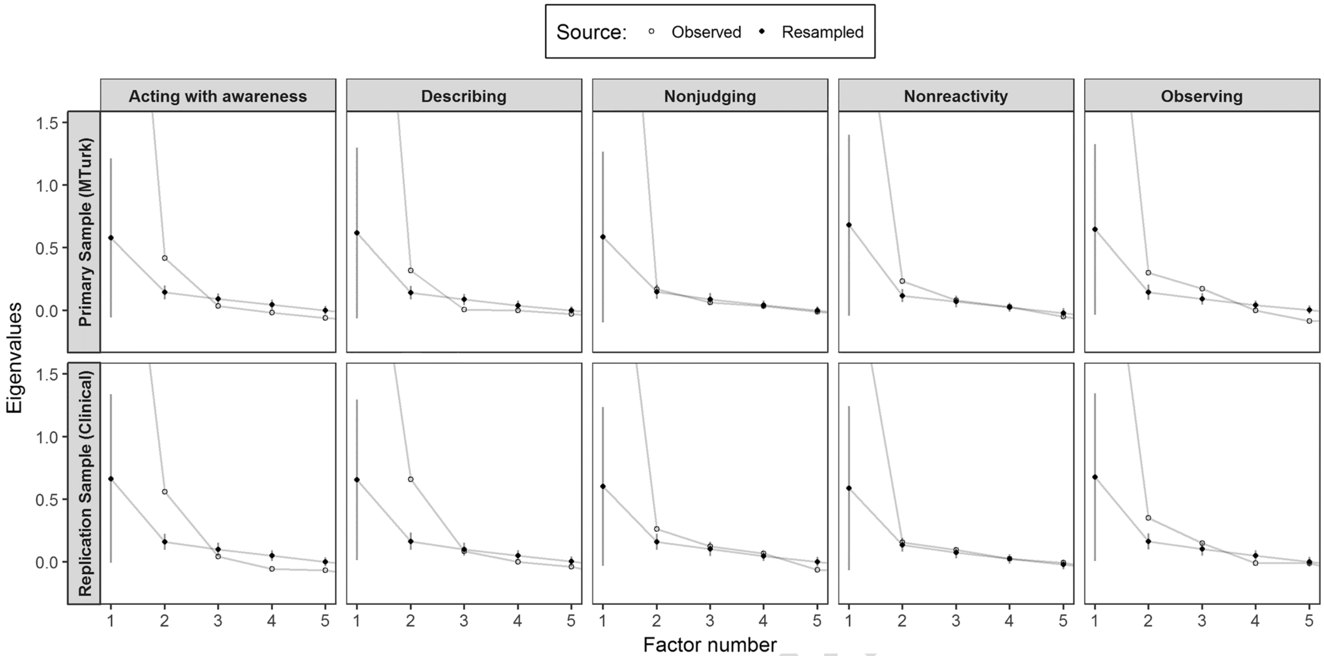

Fig. 1 Parallel analysis of individual facets of FFMQ. Each panel indicates eigenvalues for a specific combination of facet (columns) and sample (rows). "Observed" points are the eigenvalues observed in the dataset. "Resampled" points are the mean eigenvalues observed in 100 bootstrap resamples, as described in methods. For the resampled points, vertical error bars indicate plus or minus two standard deviations across bootstrap resamples. The first observed eigenvalue exceeded the first resampled eigenvalue in all cases, but the *y*-axis is shrunk to focus on the point at which observed and resampled lines cross. See Fig. S1 for the version of the plot that shows full *y*-axis

the primary sample, parallel analysis suggested that all five facets were multidimensional under the mean eigenvalue criterion and four of five facets were multidimensional under the 95th percentile eigenvalue criterion (the exception was nonreactivity, which was found to be unidimensional). Confirmatory factor models revealed that four out of five facets exhibit poor fit (the exception was again nonreactivity). When items were modeled as continuous (ML estimator), four of five facets exceeded the suggested cutoffs for RMSEA, four of five facets were below the suggested cutoff for CFI, and two of five facets were above the suggested cutoff for SRMR. When items were modeled as categorical (WLSMV estimator), the fit was similarly poor. Finally, more than half of modification indices were statistically significant, and the highest indices were observed for the residual correlation between duplicative items.

Summary The parallel analysis suggested that all five facets were multidimensional, although factors beyond the first one accounted for a small proportion of variance. In some cases, exploratory factor analyses suggested that the additional dimensions could be readily interpreted (e.g., the second factor on describing was a method factor for item valence). Confirmatory factor analyses indicated that all five facets of the FFMQ exhibit poor to unacceptable fit, with the acting with awareness and describing facets demonstrating particularly poor fit. The nonjudging and nonreactivity facets appeared to have better overall structure and fit than the other facets. Results were generally consistent across primary and replication samples.

Attempts to Improve the Fit of One-Facet Models

The fit of the individual facets to a unidimensional factor model was in many cases quite poor (Table 2). In the context of one-factor models, model fit can be improved (a) by modeling additional sources of covariance among the items beyond the common factor (e.g., correlating residual item variances, adding method factors) or (b) by dropping items contributing to misfit. Since the eigenvalues of the second factors that emerged in the exploratory factor models were very small relative to the first factors, they might reflect subsets of the items sharing wording or content similarity that leads to correlated residual variances.

Bifactor Models Per parallel analysis, the nonjudging facet was found to be unidimensional in the primary sample and the nonreactivity facet was found to be unidimensional in the replication sample. In the EFAs, both facets produced difficult-to-interpret second factors in the sample in which parallel analysis suggested they might be multidimensional. Thus, we did not consider these facets further for bifactor modeling.

We fit bifactor models to the acting with awareness, describing, and observing facets. For the describing facet, we fit a bifactor model with items 12, 16, and 22 (those that are

Table 3 Loadings in exploratory factor models fit separately to each facet

Facet	Source	Item	Label	Primary sample			Replication sample	
				F1	F2	F3	F1	F2
Acting with awareness	MAAS	38	I find myself doing things without paying attention	+ 0.95	–		–	+ 0.78
	MAAS	34	I do jobs or tasks automatically without being aware of what I'm doing	+ 0.90	–		–	+ 0.80
	MAAS	28	I rush through activities without being really attentive to them	+ 0.81	–		–	+ 0.67
	MAAS	23	It seems I am 'running on automatic' without much awareness of what I'm doing	+ 0.63	+ 0.24		+ 0.20	+ 0.61
	KIMS	8	I do not pay attention to what I'm doing because I'm daydreaming, worrying, or otherwise distracted	+ 0.26	+ 0.61		+ 0.75	–
	MAAS	18	I find it difficult to stay focused on what's happening in the present	–	+ 0.67		+ 0.66	–
	KIMS	5	When I do things, my mind wanders off and I'm easily distracted	–	+ 0.93		+ 0.74	–
	CAMS	13	I am easily distracted	–	+ 0.94		+ 0.92	–
Describing	KIMS	32	My natural tendency is to put my experiences into words	+ 0.92	–		+ 0.86	–
	CAMS	37	I can usually describe how I feel at the moment in considerable detail	+ 0.90	–		+ 0.85	–
	KIMS	2	I'm good at finding words to describe my feelings	+ 0.74	–		+ 0.59	+ 0.25
	KIMS	7	I can easily put my beliefs, opinions, and expectations into words	+ 0.68	+ 0.24		+ 0.69	–
	KIMS	27	Even when I'm feeling terribly upset, I can find a way to put it into words	+ 0.64	–		+ 0.66	–
	KIMS	12	It's hard for me to find the words to describe what I'm thinking	–	+ 0.80		–	+ 0.76
	KIMS	16	I have trouble thinking of the right words to express how I feel about things	–	+ 0.88		–	+ 0.80
	KIMS	22	When I have a sensation in my body, it's difficult for me to describe it because I cannot find the right words	–	+ 0.88		–	+ 0.61
Nonjudging	KIMS	39	I disapprove of myself when I have irrational ideas				+ 0.99	–
	MQ	35	When I have distressing thoughts or images, I judge myself as good or bad, depending what the thought or image is about				+ 0.31	+ 0.49
	KIMS	3	I criticize myself for having irrational or inappropriate emotions				+ 0.25	+ 0.45
	KIMS	30	I think some of my emotions are bad or inappropriate and I should not feel them				–	+ 0.72
	KIMS	14	I believe some of my thoughts are abnormal or bad and I should not think that way				–	+ 0.74
	KIMS	25	I tell myself that I should not be thinking the way I'm thinking				–	+ 0.81
	KIMS	10	I tell myself I should not be feeling the way I'm feeling				–	+ 0.58
	KIMS	17	I make judgments about whether my thoughts are good or bad				–	+ 0.65
Nonreactivity	FMI	9	I watch my feelings without getting lost in them	+ 1.00	–			
	FMI	4	I perceive my feelings and emotions without having to react to them	+ 0.37	+ 0.36			
	MQ	19	When I have distressing thoughts or images, I 'step back' and am aware of the thought or image without getting taken over by it	–	+ 0.48			
	FMI	21	In difficult situations, I can pause without immediately reacting	–	+ 0.62			
	MQ	29	When I have distressing thoughts or images, I am able just to notice them without reacting	–	+ 0.94			
	MQ	33	When I have distressing thoughts or images, I just notice them and let them go	–	+ 0.70			
	MQ	24	When I have distressing thoughts or images, I feel calm soon after	–	+ 0.59			
Observing	KIMS	20	I pay attention to sounds, such as clocks ticking, birds chirping, or cars passing	+ 0.79	–	–	+ 0.37	+ 0.34
	KIMS	26	I notice the smells and aromas of things	+ 0.78	–	–	–	+ 0.94
	KIMS	31	I notice visual elements in art or nature, such as colors, shapes, textures, or patterns of light and shadow	+ 0.72	–	–	+ 0.36	+ 0.40
	KIMS	15	I pay attention to sensations, such as the wind in my hair or sun on my face	+ 0.50	–	+ 0.42	+ 0.79	–
	KIMS	36	I pay attention to how my emotions affect my thoughts and behavior	+ 0.22	+ 0.46	–	–	+ 0.47
	KIMS	1	When I'm walking, I deliberately notice the sensations of my body moving	+ 0.22	+ 0.28	+ 0.39	+ 0.68	–
	KIMS	11	I notice how foods and drinks affect my thoughts, bodily sensations, and emotions	–	+ 0.94	–	+ 0.54	–
	KIMS	6	When I take a shower or bath, I stay alert to the sensations of water on my body	–	–	+ 0.93	+ 0.84	–

Note. Based on the number of factors suggested by 95th percentile eigenvalue criterion in the parallel analysis. Loadings below 0.20 are omitted ("–") for readability. Nonjudging facet was found to be unidimensional in the primary sample and nonreactivity facet was found to be unidimensional in replication sample: these two solutions are omitted. In the primary sample, correlations among factors ranged from 0.61 (between F1 and F3 in observing) to 0.83 (between F1 and F2 in describing). In the replication sample, correlations among factors ranged from 0.52 (between F1 and F2 in describing) to 0.66 (between F1 and F2 in acting with awareness)

reverse-scored) loading onto an orthogonal method factor in addition to the facet-wide common factor. For the acting with awareness facet, we fit a bifactor model with items 5, 8, 13, and 18 loading onto an orthogonal method factor in addition the facet-wide common factor. For the observing facet, we fit three bifactor models that included (a) an orthogonal method factor for items beginning with the "I pay attention…" stem (items 15, 20, 36), (b) an orthogonal method factor for items beginning with the "I notice…" stem (items 11, 26, 31), and (c) orthogonal method factors for both those groups of items. Estimation did not converge for these models, so they are not reported.

Table 4 reports the fit of unidimensional versus bifactor confirmatory factor models for the acting with awareness

Table 4 Fit of bifactor models for facets

Estimator	Sample	Facet	Model	# params	χ^2	df	χ^2/df	$p(\chi^2$,df)	RMSEA	CFI	WRMR	SRMR
WLSMV	Primary sample (MTurk)	Acting with awareness	Unidimensional	40	516	20	25.8	$p < .001$	0.218	0.964	2.209	–
			Bifactor: items 5, 8, 13, 18	44	117	16	7.3	$p < .001$	0.110	0.993	0.792	–
		Describing	Unidimensional	40	463	20	23.1	$p < .001$	0.206	0.969	1.674	–
			Bifactor: items 12, 16, 22	43	105	17	6.2	$p < .001$	0.100	0.994	0.634	–
	Replication sample (clinical)	Acting with awareness	Unidimensional	40	368	20	18.4	$p < .001$	0.197	0.915	1.835	–
			Bifactor: items 5, 8, 13, 18	44	93	16	5.8	$p < .001$	0.104	0.981	0.758	–
		Describing	Unidimensional	40	689	20	34.5	$p < .001$	0.273	0.804	2.565	–
			Bifactor: items 12, 16, 22	43	116	17	6.8	$p < .001$	0.114	0.971	0.888	–
ML	Primary sample (MTurk)	Acting with awareness	Unidimensional	24	393	20	19.6	$p < .001$	0.189	0.878	–	0.055
			Bifactor: items 5, 8, 13, 18	28	81	16	5.1	$p < .001$	0.088	0.979	–	0.024
		Describing	Unidimensional	24	212	20	10.6	$p < .001$	0.136	0.939	–	0.040
			Bifactor: items 12, 16, 22	27	44	17	2.6	$p < .001$	0.055	0.992	–	0.019
	Replication sample (clinical)	Acting with awareness	Unidimensional	24	213	20	10.6	$p < .001$	0.147	0.874	–	0.067
			Bifactor: items 5, 8, 13, 18	28	46	16	2.9	$p < .001$	0.065	0.980	–	0.033
		Describing	Unidimensional	24	244	20	12.2	$p < .001$	0.158	0.825	–	0.088
			Bifactor: items 12, 16, 22	27	53	17	3.1	$p < .001$	0.069	0.972	–	0.030

Note. WLSMV, weighted least squares mean and variance adjusted estimator; *ML*, maximum likelihood estimator; *# params*, number of free parameters; χ^2, chi-square statistic of model fit; *df*, degrees of freedom for chi-square statistic; *p*, *p*-value for chi-square test of model fit; *RMSEA*, root mean square error of approximation; *CFI*, comparative fit index; *WRMR*, weighted root mean square residual; *SRMR*, standardized root mean square residual. In bifactor models, the listed items each loaded onto a method factor that was uncorrelated with the latent mindfulness construct

and describing facets. The fit of the acting with awareness facet was substantially improved ($p < .001$ in likelihood ratio tests) by adding a method factor for the items invoking distraction (items 5, 8, 13, and 18). The facet met recommended cutoffs for CFI, WRMR, and SRMR in all models; RMSEA ranged from 0.065 to 0.114 across estimators and samples. The fit of the describing facet was substantially improved ($p < .001$ in likelihood ratio tests) by adding a method factor for the items that are reverse-scored (i.e., items that indicate having difficulty describing). The facet met recommended cutoffs for CFI, WRMR, and SRMR in all models; RMSEA ranged from 0.055 to 0.114 across estimators and samples.

Short Forms Table 5 reports fit of unidimensional confirmatory factor models fit to short versions of the facets, as defined by the FFMQ-24 (Bohlmeijer et al. 2011) and FFMQ-23 (Burzler et al. 2019). The fit of the facets to a unidimensional factor model was somewhat better on the FFMQ-24 than on the FFMQ-39. In the primary sample, when treating items as categorical (estimator = WLSMV), the chi-square test of exact model fit remained statistically significant ($p < .05$) for all five facets. The acting with awareness and observing facets both exhibited clear improvements in RMSEA, CFI, and WRMR, with each meeting recommended cutoffs for at least two of the three indices. The fit of the nonjudging and nonreactivity facets exhibited more modest improvements, meeting some recommended cutoffs while remaining above the cutoff for others (e.g., RMSEA = 0.083 for nonjudging; RMSEA = 0.103 for nonreactivity). The fit of the describing facet exhibited still less improvement, remaining similar in RMSEA (0.216 vs. 0.206) and χ^2/df (25.4 vs. 23.1) while improving on CFI (0.984 vs. 0.969) and WRMR (1.056 vs. 1.674). A similar pattern was observed when modeling items as continuous (estimator = ML) and in the replication sample: fit improved substantially for the acting with awareness and observing facets, less so for the nonjudging and nonreactivity facets, and least for the describing facet. In summary, the FFMQ-24 facets generally exhibited better fit to the unidimensional model than did the FFMQ-39 facets, with some misfit remaining.

The fit of the facets to a unidimensional factor model was substantially better on the FFMQ-23 than on the FFMQ-39. In the primary sample, when treating items as categorical (estimator = WLSMV), the chi-square test of exact model fit was non-significant ($p > .05$) for the describing, nonjudging, and observing facets, indicating close fit. The fit of the acting with awareness facet improved substantially on RMSEA (0.157 vs. 0.218), CFI (0.998 vs. 0.964), and WRMR (0.427 vs. 2.209), meeting the recommended cutoff for two of the three indices. The fit of the nonreactivity facet did not change; the FFMQ-23 and FFMQ-39 include the same items on this facet. A similar pattern was observed when modeling items as continuous (estimator = ML) and in the replication sample: fit to a unidimensional factor model was substantially better in the FFMQ-23 facets (vs. FFMQ-39 facets), with some misfit remaining on the acting with awareness facet.

Extrinsic Convergent Validity of Short-Form Facets and FFMQ-39 Facets

Use of the short-form facets appeared to significantly improve fit to the unidimensional model, but it was important to verify that they still related similarly to external criteria. Table 6 reports tests of extrinsic convergent validity in the primary sample. Each test evaluated the difference in the correlation of a criterion variable with (a) the FFMQ-39 facet and (b) the short form (i.e., FFMQ-24 or FFMQ-23) facet. After applying the FDR correction, 14 of 81 differences were statistically significant ($p < .05$), with the correlation of the criterion with the short-form facet being weaker than that with the FFMQ-39 facet in each instance. Effect sizes for the difference were all below Cohen's (1988) threshold for a "small" effect (i.e., $q = 0.10$), with absolute differences in the r metric ranging from 0 to 0.05. Thus, results suggest that the FFMQ-39 facets and their short-form equivalents are roughly empirically interchangeable for the prediction of smoking, risky alcohol use, psychological distress, self-control, behavioral inhibition, sensation seeking, future time perspective, conscientiousness, and openness.

Table S4 reports tests of extrinsic convergent validity in the replication sample. After applying the FDR correction, 6 of 30 differences were statistically significant ($p < .05$), with the correlation of the criterion with the short-form facet being weaker than that with the FFMQ-39 facet in each instance. Effect sizes for the difference were all below Cohen's (1988) threshold for a "small" effect (i.e., $q = 0.10$), with absolute differences in the r metric ranging from 0 to 0.09. Thus, results suggest that the FFMQ-39 facets and their short-form equivalents were roughly empirically for the prediction of craving, dependence severity, drinking problems, and acceptance.

Discussion

We evaluated the factor structure of the individual facets of the FFMQ in two samples. The parallel analysis suggested all five facets were multidimensional. Exploratory factor models suggested the presence of method factors on the acting with awareness (items using the term "distraction") and describing facets (items that were reverse-scored). Confirmatory factor models indicated the poor model fit for all facets, in both samples. Misfit was obscured when the facets were analyzed simultaneously (as in prior work) rather than independently. In follow-up analyses, model fit improved substantially on the acting with awareness and describing facets when method factors were accounted for in a bifactor model. Model fit was also better for the facets of FFMQ short forms (FFMQ-24 and FFMQ-23) than for the full-length facets. Thus, the use of short-form facets or latent variable models (e.g., bifactor specifications) are both viable solutions for addressing multidimensionality when desired.

FFMQ Facets Do Not Fit the Unidimensional Factor Model

Confirmatory factor models showed that the FFMQ facets do not fit the unidimensional factor model, and exploratory factor models suggested they are multidimensional. Poor fit to a unidimensional factor model has now been found in all three samples in which the facets' fit has been evaluated (Tran et al. 2013, plus our primary and replication samples). The pattern of misfit could be traced to the presence of construct-irrelevant common variance (e.g., method factor for item valence on the describing facet). This implies that when investigators include an individual facet from the FFMQ as a variable in their analyses, they are in part measuring latent mindfulness construct and in part measuring method factors introduced by the FFMQ items. Conflating method and mindfulness factors limits the ability to precisely test theories about how different aspects of mindfulness explain psychological phenomena, undermining the original goal of the FFMQ.

The acceptable fit of the five facet model does not justify the use of the individual facets. To see why, suppose that a manuscript reporting the development of a new measure of mindfulness is submitted for review. Confirmatory factor analysis of the new measure yields poor model fit statistics, and exploratory factor analysis suggests that the measure is multidimensional. However, the authors argue that when the measure is evaluated in larger confirmatory factor analysis with several additional, correlated scales, then this model fits the data acceptably. Reviewers are unlikely to agree that this argument supports the psychometric adequacy of the new mindfulness measure. In the same way, the fact that the acting with awareness facet shows acceptable fit in a model with four other facets does not excuse its poor fit when modeled alone.

When Might the Multidimensionality of the Facets Matter?

The empirical (or mathematical) multidimensionality of the facets does not imply that they each measure multiple *substantive* mindfulness constructs. In fact, the results of our follow-up analyses suggest that each facet can be conceptualized as substantively unidimensional. For the acting with awareness and describing facets, accounting for method factors in a bifactor specification yielded a reasonable fit to a substantively unidimensional model. For all five facets, the use of short-form facets improved fit to a unidimensional model while appearing to minimally affect the nature of the construct per the extrinsic convergent validity tests. Thus, the current findings support the conceptualization of each facet as measuring a single mindfulness construct, with empirical t5.1 **Table 5** Fit of confirmatory factor models for one-facet models of FFMQ-39, FFMQ-24, and FFMQ-23

t5.2	Estimator	Sample	Facet	Form	# items	# params	χ^2	df	χ^2/df	$p(\chi^2$,df)	RMSEA	CFI	WRMR	SRMR
t5.3	WLSMV	Primary sample (MTurk)	Acting with awareness	FFMQ-39 (Baer)	8	40	516	20	25.8	$p<.001$	0.218	0.964	2.209	–
t5.4				FFMQ-24 (Bohlmeijer)	5	25	19	5	3.9	$p<.01$	0.074	0.998	0.427	–
t5.5				FFMQ-23 (Burzler)	4	20	28	2	13.9	$p<.001$	0.157	0.996	0.596	–
t5.6			Describing	FFMQ-39 (Baer)	8	40	463	20	23.1	$p<.001$	0.206	0.969	1.674	–
t5.7				FFMQ-24 (Bohlmeijer)	5	25	127	5	25.4	$p<.001$	0.216	0.984	1.056	–
t5.8				FFMQ-23 (Burzler)	4	20	1	2	0.6	ns	0.000	1.000	0.101	–
t5.9			Nonjudging	FFMQ-39 (Baer)	8	40	208	20	10.4	$p<.001$	0.134	0.986	1.084	–
t5.10				FFMQ-24 (Bohlmeijer)	5	25	23	5	4.6	$p<.001$	0.083	0.998	0.428	–
t5.11				FFMQ-23 (Burzler)	4	20	0	2	0.1	ns	0.000	1.000	0.049	–
t5.12			Nonreactivity	FFMQ-39 (Baer)	7	35	174	14	12.4	$p<.001$	0.148	0.945	1.280	–
t5.13				FFMQ-24 (Bohlmeijer)	5	25	33	5	6.5	$p<.001$	0.103	0.985	0.645	–
t5.14				FFMQ-23 (Burzler)	7	35	174	14	12.4	$p<.001$	0.148	0.945	1.280	–
t5.15			Observing	FFMQ-39 (Baer)	8	40	221	20	11.1	$p<.001$	0.139	0.964	1.344	–
t5.16				FFMQ-24 (Bohlmeijer)	4	20	1	2	0.3	ns	0.000	1.000	0.105	–
t5.17				FFMQ-23 (Burzler)	4	20	1	2	0.3	ns	0.000	1.000	0.105	–
t5.18		Replication sample (clinical)	Acting with awareness	FFMQ-39 (Baer)	8	40	368	20	18.4	$p<.001$	0.197	0.915	1.835	–
t5.19				FFMQ-24 (Bohlmeijer)	5	25	36	5	7.3	$p<.001$	0.118	0.984	0.697	–
t5.20				FFMQ-23 (Burzler)	4	20	12	2	6.0	$p<.01$	0.105	0.996	0.383	–
t5.21			Describing	FFMQ-39 (Baer)	8	40	689	20	34.5	$p<.001$	0.273	0.804	2.565	–
t5.22				FFMQ-24 (Bohlmeijer)	5	25	173	5	34.6	$p<.001$	0.274	0.877	1.704	–
t5.23				FFMQ-23 (Burzler)	4	20	13	2	6.5	$p<.01$	0.110	0.994	0.439	–
t5.24			Nonjudging	FFMQ-39 (Baer)	8	40	131	20	6.5	$p<.001$	0.111	0.962	1.054	–
t5.25				FFMQ-24 (Bohlmeijer)	5	25	15	5	3.1	$p<.01$	0.068	0.992	0.513	–
t5.26				FFMQ-23 (Burzler)	4	20	0	2	0.1	ns	0.000	1.000	0.055	–
t5.27			Nonreactivity	FFMQ-39 (Baer)	7	35	54	14	3.8	$p<.001$	0.079	0.970	0.762	–
t5.28				FFMQ-24 (Bohlmeijer)	5	25	30	5	5.9	$p<.001$	0.104	0.969	0.686	–
t5.29				FFMQ-23 (Burzler)	7	35	54	14	3.8	$p<.001$	0.079	0.970	0.762	–
t5.30			Observing	FFMQ-39 (Baer)	8	40	181	20	9.0	$p<.001$	0.134	0.938	1.265	–
t5.31				FFMQ-24 (Bohlmeijer)	4	20	10	2	5.0	$p<.01$	0.095	0.992	0.410	–
t5.32				FFMQ-23 (Burzler)	4	20	10	2	5.0	$p<.01$	0.095	0.992	0.410	–
t5.33	ML	Primary sample (MTurk)	Acting with awareness	FFMQ-39 (Baer)	8	24	393	20	19.6	$p<.001$	0.189	0.878	–	0.055
t5.34				FFMQ-24 (Bohlmeijer)	5	15	13	5	2.6	$p<.05$	0.056	0.995	–	0.014
t5.35				FFMQ-23 (Burzler)	4	12	20	2	9.8	$p<.001$	0.130	0.986	–	0.016
t5.36			Describing	FFMQ-39 (Baer)	8	24	212	20	10.6	$p<.001$	0.136	0.939	–	0.040
t5.37				FFMQ-24 (Bohlmeijer)	5	15	45	5	9.1	$p<.001$	0.124	0.975	–	0.024
t5.38				FFMQ-23 (Burzler)	4	12	0	2	0.2	ns	0.000	1.000	–	0.003
t5.39			Nonjudging	FFMQ-39 (Baer)	8	24	109	20	5.4	$p<.001$	0.092	0.971	–	0.027
t5.40				FFMQ-24 (Bohlmeijer)	5	15	9	5	1.9	ns	0.040	0.997	–	0.010
t5.41				FFMQ-23 (Burzler)	4	12	0	2	0.1	ns	0.000	1.000	–	0.002
t5.42			Nonreactivity	FFMQ-39 (Baer)	7	21	97	14	6.9	$p<.001$	0.106	0.932	–	0.041

t5.44 **Table 5** (continued)

	Estimator	Sample	Facet	Form	# items	# params	χ^2	df	χ^2/df	$p(\chi^2$,df)	RMSEA	CFI	WRMR	SRMR
t5.43				FFMQ-24 (Bohlmeijer)	5	15	16	5	3.3	$p < .01$	0.066	0.985	–	0.025
t5.44				FFMQ-23 (Burzler)	7	21	97	14	6.9	$p < .001$	0.106	0.932	–	0.041
t5.45			Observing	FFMQ-39 (Baer)	8	24	139	20	7.0	$p < .001$	0.107	0.936	–	0.043
t5.46				FFMQ-24 (Bohlmeijer)	4	12	0	2	0.2	ns	0.000	1.000	–	0.004
t5.47				FFMQ-23 (Burzler)	4	12	0	2	0.2	ns	0.000	1.000	–	0.004
t5.48		Replication sample (clinical)	Acting with awareness	FFMQ-39 (Baer)	8	24	213	20	10.6	$p < .001$	0.147	0.874	–	0.067
t5.49				FFMQ-24 (Bohlmeijer)	5	15	27	5	5.3	$p < .001$	0.099	0.969	–	0.032
t5.50				FFMQ-23 (Burzler)	4	12	5	2	2.4	ns	0.057	0.996	–	0.012
t5.51			Describing	FFMQ-39 (Baer)	8	24	244	20	12.2	$p < .001$	0.158	0.825	–	0.088
t5.52				FFMQ-24 (Bohlmeijer)	5	15	67	5	13.4	$p < .001$	0.167	0.882	–	0.066
t5.53				FFMQ-23 (Burzler)	4	12	7	2	3.6	$p < .05$	0.076	0.991	–	0.017
t5.54			Nonjudging	FFMQ-39 (Baer)	8	24	84	20	4.2	$p < .001$	0.084	0.943	–	0.039
t5.55				FFMQ-24 (Bohlmeijer)	5	15	11	5	2.1	ns	0.051	0.989	–	0.020
t5.56				FFMQ-23 (Burzler)	4	12	0	2	0.1	ns	0.000	1.000	–	0.003
t5.57			Nonreactivity	FFMQ-39 (Baer)	7	21	30	14	2.2	$p < .01$	0.051	0.973	–	0.030
t5.58				FFMQ-24 (Bohlmeijer)	5	15	14	5	2.8	$p < .05$	0.064	0.975	–	0.027
t5.59				FFMQ-23 (Burzler)	7	21	30	14	2.2	$p < .01$	0.051	0.973	–	0.030
t5.60			Observing	FFMQ-39 (Baer)	8	24	117	20	5.8	$p < .001$	0.104	0.911	–	0.050
t5.61				FFMQ-24 (Bohlmeijer)	4	12	7	2	3.6	$p < .05$	0.076	0.988	–	0.019
t5.62				FFMQ-23 (Burzler)	4	12	7	2	3.6	$p < .05$	0.076	0.988	–	0.019

Note. WLSMV, weighted least squares mean and variance adjusted estimator; *ML*, maximum likelihood estimator; *# items*, number of items on facet; *# params*, number of free parameters; χ^2, chi-square statistic of model fit; *df*, degrees of freedom for chi-square statistic; *p*, *p*-value for chi-square test of model fit; *RMSEA*, root mean square error of approximation; *CFI*, comparative fit index; *WRMR*, weighted root mean square residual; *SRMR*, standardized root mean square residual. Forms are the FFMQ-39 (Baer et al. 2006); the FFMQ-24 (Bohlmeijer et al. 2011); and the FFMQ-23 (Burzler et al. 2019)

multidimensionality arising from other sources of common variance (e.g., duplication in item wording, reverse-scoring). Investigators may benefit from considering the ways in which the facets' empirical multidimensionality might affect inferences in their own specific research applications.

One way that multidimensionality may be important is by affecting the relation of the facets to other psychological constructs (Reise et al. 2013a, b). We probed this possibility indirectly via the tests of extrinsic convergent validity of the short-form and full-length facets. The short-form facets had better fit to a unidimensional model (i.e., were less multidimensional) and yet correlated similarly to the full-length facets with other measures in the primary and replication samples (Table 6, Table S4). This is reassuring because it suggests that multidimensionality in the facets would not materially affect scientific conclusions about their relation to the external measures we evaluated. However, we tested only a limited set of external criteria and so cannot establish that this pattern will apply to all possible applications. The impact may be larger when the external criterion is related to the method factor—for example, the correlation between the acting with awareness facet and an external measure of distractibility might be especially inflated by the presence of the "distraction" method factor.

Another way that multidimensionality may be important is in affecting the apparent fit of structural equation models that include an FFMQ facet. Suppose one is testing the substantive theory that acting with awareness facet of mindfulness fully mediates the relationship between social anxiety and drinking problems (Clerkin et al. 2017). In such an application, the poor fit of the measurement component of the model (i.e., eight-item responses loading on a single latent variable for acting with awareness) may result in apparently poor fit statistics for the model as whole (e.g., RMSEA > 0.10, CFI < 0.90). The investigator could mistakenly interpret this poor model fit as evidence against the structural component of his or her theory (i.e., full mediation), when in fact it simply reflects multidimensionality in the FFMQ facet.

t6.1 **Table 6** Tests of extrinsic convergent validity of facets of FFMQ-39, FFMQ-24, and FFMQ-23 in primary sample

	Form	Facet	Criterion	*N*	*r*(full,crit)	*r*(short,crit)	*r*(full,short)	Cohen's *q*	*p*
t6.2									
t6.3	FFMQ-24 (Bohlmeijer)	Acting with awareness	Cigarettes per day	522	0.13	0.12	0.96	+ 0.02	ns
t6.4			Risky alcohol use (AUDIT)	522	− 0.16	− 0.16	0.96	+ 0.01	ns
t6.5			Psychological distress (K6)	522	− 0.36	− 0.35	0.96	− 0.01	ns
t6.6			Self-control (BSCS)	522	0.65	0.62	0.96	+ 0.05	$p < .05$
t6.7			Behavioral inhibition (BIS/BAS)	522	− 0.29	− 0.27	0.96	− 0.03	ns
t6.8			Sensation seeking (UPPS-P)	522	− 0.05	− 0.08	0.96	+ 0.03	ns
t6.9			Future time perspective (ZTPI)	522	0.44	0.41	0.96	+ 0.04	$p < .05$
t6.10			Conscientiousness (TIPI)	522	0.53	0.49	0.96	+ 0.05	$p < .05$
t6.11			Openness (TIPI)	522	0.17	0.18	0.96	− 0.02	ns
t6.12		Describing	Cigarettes per day	522	0.10	0.09	0.98	+ 0.01	ns
t6.13			Risky alcohol use (AUDIT)	522	− 0.05	− 0.05	0.98	+ 0.00	ns
t6.14			Psychological distress (K6)	522	− 0.29	− 0.30	0.98	+ 0.01	ns
t6.15			Self-control (BSCS)	522	0.42	0.43	0.98	− 0.01	ns
t6.16			Behavioral inhibition (BIS/BAS)	522	− 0.27	− 0.29	0.98	+ 0.01	ns
t6.17			Sensation seeking (UPPS-P)	522	0.07	0.06	0.98	+ 0.00	ns
t6.18			Future time perspective (ZTPI)	522	0.25	0.24	0.98	+ 0.01	ns
t6.19			Conscientiousness (TIPI)	522	0.27	0.27	0.98	− 0.00	ns
t6.20			Openness (TIPI)	522	0.28	0.25	0.98	+ 0.02	$p < .05$
t6.21		Nonjudging	Cigarettes per day	522	0.03	0.04	0.98	− 0.00	ns
t6.22			Risky alcohol use (AUDIT)	522	− 0.15	− 0.14	0.98	− 0.01	ns
t6.23			Psychological distress (K6)	522	− 0.41	− 0.39	0.98	− 0.02	ns
t6.24			Self-control (BSCS)	522	0.34	0.31	0.98	+ 0.03	$p < .05$
t6.25			Behavioral inhibition (BIS/BAS)	522	− 0.32	− 0.31	0.98	− 0.01	ns
t6.26			Sensation seeking (UPPS-P)	522	− 0.09	− 0.08	0.98	− 0.01	ns
t6.27			Future time perspective (ZTPI)	522	0.05	0.02	0.98	+ 0.02	$p < .05$
t6.28			Conscientiousness (TIPI)	522	0.20	0.18	0.98	+ 0.03	$p < .05$
t6.29			Openness (TIPI)	522	0.12	0.09	0.98	+ 0.02	$p < .05$
t6.30		Nonreactivity	Cigarettes per day	522	0.10	0.09	0.97	+ 0.01	ns
t6.31			Risky alcohol use (AUDIT)	522	− 0.05	− 0.04	0.97	− 0.01	ns
t6.32			Psychological distress (K6)	522	− 0.32	− 0.33	0.97	+ 0.01	ns
t6.33			Self-control (BSCS)	522	0.39	0.38	0.97	+ 0.02	ns
t6.34			Behavioral inhibition (BIS/BAS)	522	− 0.41	− 0.41	0.97	− 0.01	ns
t6.35			Sensation seeking (UPPS-P)	522	0.12	0.13	0.97	− 0.01	ns
t6.36			Future time perspective (ZTPI)	522	0.20	0.17	0.97	+ 0.02	ns
t6.37			Conscientiousness (TIPI)	522	0.26	0.23	0.97	+ 0.03	$p < .05$
t6.38			Openness (TIPI)	522	0.17	0.18	0.97	− 0.01	ns
t6.39		Observing	Cigarettes per day	522	0.08	0.05	0.92	+ 0.03	ns
t6.40			Risky alcohol use (AUDIT)	522	− 0.00	− 0.02	0.92	+ 0.01	ns
t6.41			Psychological distress (K6)	522	0.03	0.03	0.92	− 0.00	ns
t6.42			Self-control (BSCS)	522	0.23	0.22	0.92	+ 0.00	ns
t6.43			Behavioral inhibition (BIS/BAS)	522	0.02	0.04	0.92	− 0.02	ns
t6.44			Sensation seeking (UPPS-P)	522	0.03	0.03	0.92	+ 0.01	ns
t6.45			Future time perspective (ZTPI)	522	0.25	0.25	0.92	+ 0.00	ns
t6.46			Conscientiousness (TIPI)	522	0.18	0.16	0.92	+ 0.02	ns
t6.47			Openness (TIPI)	522	0.27	0.26	0.92	+ 0.01	ns
t6.48	FFMQ-23 (Burzler)	Acting with awareness	Cigarettes per day	522	0.13	0.13	0.94	+ 0.01	ns
t6.49			Risky alcohol use (AUDIT)	522	− 0.16	− 0.14	0.94	− 0.02	ns
t6.50			Psychological distress (K6)	522	− 0.36	− 0.34	0.94	− 0.03	ns
t6.51			Self-control (BSCS)	522	0.65	0.63	0.94	+ 0.04	ns

Table 6 (continued)

Form	Facet	Criterion	N	r(full,crit)	r(short,crit)	r(full,short)	Cohen's q	p
		Behavioral inhibition (BIS/BAS)	522	− 0.29	− 0.31	0.94	+ 0.02	ns
		Sensation seeking (UPPS-P)	522	− 0.05	− 0.02	0.94	− 0.04	ns
		Future time perspective (ZTPI)	522	0.44	0.43	0.94	+ 0.02	ns
		Conscientiousness (TIPI)	522	0.53	0.50	0.94	+ 0.04	ns
		Openness (TIPI)	522	0.17	0.13	0.94	+ 0.04	ns
	Describing	Cigarettes per day	522	0.10	0.11	0.96	− 0.01	ns
		Risky alcohol use (AUDIT)	522	− 0.05	− 0.04	0.96	− 0.01	ns
		Psychological distress (K6)	522	− 0.29	− 0.24	0.96	− 0.05	$p < .05$
		Self-control (BSCS)	522	0.42	0.37	0.96	+ 0.06	$p < .05$
		Behavioral inhibition (BIS/BAS)	522	− 0.27	− 0.23	0.96	− 0.05	$p < .05$
		Sensation seeking (UPPS-P)	522	0.07	0.08	0.96	− 0.01	ns
		Future time perspective (ZTPI)	522	0.25	0.24	0.96	+ 0.01	ns
		Conscientiousness (TIPI)	522	0.27	0.24	0.96	+ 0.03	ns
		Openness (TIPI)	522	0.28	0.27	0.96	+ 0.01	ns
	Nonjudging	Cigarettes per day	522	0.03	0.03	0.97	+ 0.00	ns
		Risky alcohol use (AUDIT)	522	− 0.15	− 0.15	0.97	+ 0.00	ns
		Psychological distress (K6)	522	− 0.41	− 0.42	0.97	+ 0.02	ns
		Self-control (BSCS)	522	0.34	0.37	0.97	− 0.03	$p < .05$
		Behavioral inhibition (BIS/BAS)	522	− 0.32	− 0.30	0.97	− 0.02	ns
		Sensation seeking (UPPS-P)	522	− 0.09	− 0.09	0.97	+ 0.00	ns
		Future time perspective (ZTPI)	522	0.05	0.08	0.97	− 0.03	$p < .05$
		Conscientiousness (TIPI)	522	0.20	0.22	0.97	− 0.02	ns
		Openness (TIPI)	522	0.12	0.10	0.97	+ 0.01	ns
	Observing	Cigarettes per day	522	0.08	0.05	0.92	+ 0.03	ns
		Risky alcohol use (AUDIT)	522	− 0.00	− 0.02	0.92	+ 0.01	ns
		Psychological distress (K6)	522	0.03	0.03	0.92	− 0.00	ns
		Self-control (BSCS)	522	0.23	0.22	0.92	+ 0.00	ns
		Behavioral inhibition (BIS/BAS)	522	0.02	0.04	0.92	− 0.02	ns
		Sensation seeking (UPPS-P)	522	0.03	0.03	0.92	+ 0.01	ns
		Future time perspective (ZTPI)	522	0.25	0.25	0.92	+ 0.00	ns
		Conscientiousness (TIPI)	522	0.18	0.16	0.92	+ 0.02	ns
		Openness (TIPI)	522	0.27	0.26	0.92	+ 0.01	ns

Note. r(full,crit), correlation between FFMQ-39 facet and criterion variable; *r(short,crit)*, correlation between short form (i.e., FFMQ-24 or FFMQ-23) facet and criterion variable; *r(full,short)*, correlation between FFMQ-39 facet and short-form facet; *p*, is p value for test of difference between *r*(full,crit) and *r*(short,crit) using the Williams (1959) method. All *p* values in extrinsic convergent validity analyses were corrected with the False Discovery Rate
Q11 correction (Benjamini and Hochberg 1995). Risky alcohol use is the total score from the Alcohol Use Disorders Identification Test (AUDIT; Saunders,
Q12 Aasland, Babor, Fuente, & Grant, 1993). Psychological distress was measured via the Kessler Psychological Distress (K6) Scale (Kessler et al., 2002). Self-control was measured via the Brief Self-Control Scale (Tangney et al. 2004). Behavioral inhibition was measured via the BIS/BAS (Carver and White 1994). Sensation seeking was measured via the UPPS-P Scale (Lynam et al. 2006). Conscientiousness and openness were measured via the Ten
Q13 Item Personality Inventory (Gosling, Rentfrow, & Swann, 2003)

More generally, multidimensionality on the FFMQ facets will convey the negative consequences outlined in the introduction. The latent factor score will be skewed toward items linked by method effects (e.g., wording similarity); the loadings across items will no longer reflect the true relations of the items to the construct (Edwards et al. 2017); reliability coefficients will be artificially inflated (Green and Yang 2009). These consequences can be illustrated when considering the bifactor models fit to the acting with awareness facet. Regarding bias in item-construct relations, after adding a method factor for items asking about "distraction," the median absolute change in item factor loading was 0.07 (range = 0–0.12) for the primary sample and 0.13 (range = 0.06–0.20) for the replication sample. Regarding bias in reliability coefficients, in the primary sample, 86% of the common variance was explained by the latent acting with awareness construct and 9% was explained by the orthogonal method factor—these numbers were 77% and 12% in the replication sample (see supplement). Thus, the presence of multidimensionality affected estimates of item-construct relations and of score reliability.

Options for Addressing the Multidimensionality of the Facets

The fact that the short-form and full-length facets correlated similarly with external criteria the multidimensionality of the facets may be safe to ignore in many applications, with its impact varying on a case-by-case basis. When it is desired, there are at least two ways to address multidimensionality in the facets. The first way is to use short-form versions of the facets (e.g., FFMQ-24 or FFMQ-23). This has the advantage of being a simple solution to implement: the user can simply calculate the facet sum-score using a subset of the items and then proceed with their planned analysis. However, there are two downsides to using the short-forms to address multidimensionality. The first downside is the reduced content span. Facet unidimensionality may be obtained by dropping items that contribute to misfit but simultaneously reduces content span. For example, the FFMQ-23 (Burzler et al. 2019) short version of the acting with awareness facet excludes all the items invoking "automaticity" so that the remaining four just ask about distraction (i.e., the developers chose the second of the two factors that emerge in EFA in our data). The second downside is reduced score reliability. In the primary sample, relative to the FFMQ-39 facets, coefficient alpha was on average 0.04 lower for the FFMQ-24 facets and 0.08 lower for the FFMQ-15 facets (cf. average decrements of 0.07 and 0.15 in the replication sample). The decrement in score reliability is even larger for individuals in the lower and upper ranges of the mindfulness constructs (Pelham III et al. 2019).

The second way to address multidimensionality in the facets is the use of measurement models in a structural equation modeling framework (Podsakoff et al. 2003; Podsakoff et al., 2012). For example, one could fit a bifactor model to the acting with awareness facet (as we did here) to account for the method factor while examining the relationship of the acting with awareness factor with another measure (Chen et al. 2006). Relative to the use of short-form facets, addressing multidimensionality through measurement models is more difficult but allows retention of all items, improving score reliability.

Implications for Future Methodological Evaluations of the FFMQ

The multidimensionality of the individual facets is masked when all five facets are modeled simultaneously, as five intercorrelated factors. For example, in both of our samples, the RMSEA and χ^2 / df of the five facet model were superior to those of each of the constituent one-facet models (Table 2). This seeming paradox may be explained by the fact that fit indices pool misfit from across the entire model, allowing good fit in one component of the model (e.g., elements of the covariance matrix relating an item from the first factor to items from a second factor) to obscure poor fit in another component of the model (e.g., elements of the covariance matrix relating an item from the first factor to other items from the first factor) (Anderson and Gerbing 1988).

A clear implication is that future psychometric evaluations of the FFMQ can benefit from including analyses of each facet separately. Just as the five facet model masked poor fit in the individual facets, other psychometric issues may similarly be masked when analyzing all five facets in a single model. For example, a scale's metric invariance is commonly tested by constraining the factor loadings to be equal across two groups (e.g., males and females), and then testing the significance in the change in the chi-square statistic before and after imposing the invariance constraints (Millsap and Olivera-Aguilar 2015). In the context of the full FFMQ, the observed change in the chi-square statistic would reflect misfit across *all five* facets, such that substantial measurement bias on one of the five facets could be masked by invariance on the other four. If the facets serve as independent measures of different aspects of mindfulness, then we should always evaluate their performance both individually and as a whole.

Limitations

Some limitations arise from the nature of the samples. Neither sample included a subgroup expected to be especially high on mindfulness (e.g., experienced meditators). Participants in the primary sample completed a series of quality checks to promote valid responding (Eisenberg et al. 2018), but these data still retain the limitations endemic to an MTurk sample. Participants in the second sample had all received inpatient or intensive outpatient care for Substance Use Disorder, a population that may exhibit lower mindfulness than the general population (Karyadi et al. 2014).

Additional limitations pertain to our methods. First, we were unable to attain convergence in bifactor models fit to the observing facet. The bayesian estimation might be an alternative approach that could address convergence problems. Second, while we strove to include a range of criterion variables in our assessment of extrinsic convergent validity, we still omitted a number of constructs that investigators have been interested in relating to mindfulness facets (e.g., empathy; MacDonald and Price 2017).

In summary, the purpose of this study was to investigate the psychometric structure of the individual facets of the FFMQ. Contrary to expectations based on published analyses of the scale as a whole (i.e., all five facets), we found that the FFMQ facets generally do not fit the unidimensional factor model. Follow-up analyses suggested that the facets can still be considered *substantively* unidimensional, with empirical multidimensionality arising from method effects (e.g., shared item wording). Investigators may benefit from being aware of the facets' multidimensionality and the ways in which it might affect inferences in their own specific research application. The use of short-form facets or latent variable models (e.g., bifactor specifications) were both viable solutions for addressing multidimensionality.

Author Contributions LM and DM designed and executed the study from which data for the primary sample were drawn. KW designed and executed the larger study from which data for the secondary sample were drawn. WP, OG, and DM conceptualized the current research question and planned analyses. WP and OG carried out analyses and wrote the first draft of the manuscript. SM, CW, KW, LM, and DM critically reviewed and revised the manuscript. All authors approved the final version of the manuscript for submission.

Funding Information This work was supported by the National Institutes of Health (NIH) Science of Behavior Change Common Fund Program through an award administered by the National Institute on Drug Abuse (NIDA) UH2 DA041713. Additional support was provided by NIDA P30 DA029926 and R37 DA009757, National Science Foundation grant DGE-1311230, and National Institute on Alcohol Abuse and Alcoholism grant F31 AA026768. Finally, data for the replication sample were collected using support from NIDA R21 DA010562 and R01 DA025764.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. IRB approval for the primary sample study was received at Stanford University. IRB approval for the secondary sample studies was received at the University of Washington.

Informed Consent Informed consent/assent was obtained from all participants in this study.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin, 103*, 411–423.

Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: the Kentucky Inventory of Mindfulness Skills. *Assessment, 11*, 191–206.

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.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*, 289–300.

Bohlmeijer, E., ten Klooster, P. M., Fledderus, M., Veehof, M., & Baer, R. (2011). Psychometric properties of the five facet mindfulness questionnaire in depressed adults and development of a short form. *Assessment, 18*, 308–320.

Bowen, S., Chawla, N., Collins, S. E., Witkiewitz, K., Hsu, S. H., Grow, J., et al. (2009). Mindfulness-based relapse prevention for substance use disorders: a pilot efficacy trial. *Substance Abuse, 30*, 295–305.

Bowen, S., Witkiewitz, K., Clifasefi, S. L., Grow, J., et al. (2014). Relative efficacy of mindfulness-based relapse prevention, standard relapse prevention, and treatment as usual for substance use disorders: a randomized clinical trial. *JAMA Psychiatry, 71*, 547–556.

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.

Buchheld, N., Grossman, P., & Walach, H. (2001). Measuring mindfulness in insight meditation (Vipassana) and meditation-based psychotherapy: the development of the Freiburg mindfulness inventory (FMI). *Journal for Meditation and Meditation Research, 1*, 11–34.

Burzler, M. A., Voracek, M., Hos, M., & Tran, U. S. (2019). Mechanisms of mindfulness in the general population. *Mindfulness, 10*, 469–480.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319–333.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*, 189–225.

Christopher, M. S., Neuser, N. J., Michael, P. G., & Baitmangalkar, A. (2012). Exploring the psychometric properties of the five facet mindfulness questionnaire. *Mindfulness, 3*, 124–131.

Clerkin, E. M., Sarfan, L. D., et al. (2017). Mindfulness facets, social anxiety, and drinking to cope with social anxiety: testing mediators of drinking problems. *Mindfulness, 8*, 159–170.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale: Lawrence Erlbaum Associates.

Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement, 70*, 885–901.

Curtiss, J., & Klemanski, D. (2014). Factor analysis of the five facet mindfulness in a heterogeneous clinical sample. *Journal of Psychopathology and Behavioral Assessment, 36*, 683–694.

Edwards, M. C., Houts, C. R., & Cai, L. (2017). A diagnostic procedure to detect departures from local independence in item response theory models. *Psychological Methods, 23*, 138–149.

Eisenberg, I. W., Bissett, P. G., Canning, J. R., Dallery, J., Enkavi, A. Z., Whitfield-Gabrieli, S., et al. (2018). Applying novel technologies and methods to inform the ontology of self-regulation. *Behaviour Research and Therapy, 101*, 46–57.

Eisenberg, I. W., Bissett, P. G., Enkavi, A. Z., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications, 10*, 2319.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences, 116*, 5472–5477.

Feinn, R., Tennen, H., & Kranzler, H. R. (2006). Psychometric properties of the short index of problems as a measure of recent alcohol-related problems. *Alcoholism: Clinical and Experimental Research, 27*, 1436–1441.

Feldman, G. C., Hayes, A. M., Kumar, S. M., & Greeson, J. M. (2004). Development, factor structure, and initial validation of the cognitive and affective mindfulness scale.

Fiske, D. W. (1971). *Measuring the concepts of personality.* Oxford: Aldine.

Flannery, B. A., Volpicelli, J. R., & Petinnati, M. H. (2006). Psychometric properties of the Penn alcohol craving scale. *Alcoholism: Clinical and Experimental Research, 23*, 1289–1295.

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: a cautionary tale. *Psychometrika, 74*, 121–135.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55.

Karyadi, K. A., VanderVeen, J. D., & Cyders, M. A. (2014). A meta-analysis of the relationship between trait mindfulness and substance use behaviors. *Drug and Alcohol Dependence, 143*, 1–10.

Lai, K., & Green, S. B. (2016). The problem with having two watches: assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research, 51*, 220–239.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 151–173.

Lynam, D., Smith, G., Whiteside, S., & Cyders, M. (2006). *The UPPS-P: assessing five personality pathways to impulsive behavior*. West Lafayette: Purdue University.

MacDonald, H. Z., & Price, J. L. (2017). Emotional understanding: examining alexithymia as a mediator of the relationship between mindfulness and empathy. *Mindfulness, 8*, 1644–1652.

Medvedev, O. N., Siegert, R. J., Kersten, P., & Krägeloh, C. U. (2017). Improving the precision of the five facet mindfulness questionnaire using a Rasch approach. *Mindfulness, 8*, 995–1008.

Millsap, R. E. (2007). Structural equation modeling made difficult. *Personality and Individual Differences, 42*, 875–881.

Millsap, R. E., & Olivera-Aguilar, M. (2015). Investigating measurement invariance using confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 380–392). New York: Guilford Press.

Muthén, L. K., & Muthén, B. (2015). *Mplus user's guide (7th ed.)*. Los Angeles: Muthén & Muthén.

Neal, D. J., & Carey, K. B. (2005). A follow-up psychometric analysis of the self-regulation questionnaire. *Psychology of Addictive Behaviors, 19*, 414–422.

Pelham, W. E., III, Gonzalez, O., Metcalf, S. A., Whicker, C. L., Scherer, E. A., Witkiewitz, K., et al. (2019). Item response theory analysis of the five facet mindfulness questionnaire and its short forms. *Mindfulness, 10*, 1615–1628.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *The Journal of Applied Psychology, 88*, 879–903.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013a). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*, 129–140.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013b). Multidimensionality and structural coefficient bias in structural equation modeling: a bifactor perspective. *Educational and Psychological Measurement, 73*, 5–26.

Revelle, W. R. (2017). *Psych: procedures for personality and psychological research*. Evanston: Northwestern University.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*, 102–111.

Stockwell, T., Murphy, D., & Hodgson, R. (2010). The severity of alcohol dependence questionnaire: its use, reliability and validity. *British Journal of Addiction, 78*, 145–155.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*, 271–324.

Tran, U. S., Glück, T. M., & Nader, I. W. (2013). Investigating the five facet mindfulness questionnaire (FFMQ): construction of a short form and evidence of a two-factor higher order structure of mindfulness. *Journal of Clinical Psychology, 69*, 951–965.

van Dam, N. T., Hobkirk, A. L., Danoff-Burg, S., & Earleywine, M. (2012). Mind your words: positive and negative items create method effects on the five facet mindfulness questionnaire. *Assessment, 19*, 198–204.

Veehof, M. M., ten Klooster, P. M., Taal, E., Westerhof, G. J., & Bohlmeijer, E. T. (2011). Psychometric properties of the Dutch Five Facet Mindfulness Questionnaire (FFMQ) in patients with fibromyalgia. *Clinical Rheumatology, 30*, 1045–1054.

West, S. G., & Taylor, A. B. (2015). Model fit and model selection in structural equation modeling. In W. Wu (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York: Guilford Press.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society. Series B (Methodological), 21*, 396–399.

Williams, M. J., Dalgleish, T., Karl, A., & Kuyken, W. (2014). Examining the factor structures of the five facet mindfulness questionnaire and the self-compassion scale. *Psychological Assessment, 26*, 407–418.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.