Mindfulness (2020) 11:1226–1237
https://doi.org/10.1007/s12671-020-01333-6

ORIGINAL PAPER

The Cross-cultural Validity of the Five-Facet Mindfulness Questionnaire Across 16 Countries

Johannes A. Karl[1] · Silvia Mariela Méndez Prado[2] · Asmir Gračanin[3] · Paul Verhaeghen[4] · Alexandre Ramos[5] · Satchit Prasun Mandal[6] · Johannes Michalak[7] · Chun-Qing Zhang[8] · Carlos Schmidt[9] · Ulrich S. Tran[10] · Elena Druica[11] · Stian Solem[12] · Andreea Astani[13] · Xinghua Liu[14] · Juan V. Luciano[15] · Mladenka Tkalčić[3] · Josefine L. Lilja[16] · Ingrid Dundas[17] · Samuel Y. S. Wong[18] · Ronald Fischer[1]

Published online: 27 March 2020

Abstract

Objectives The goal of the current study was to investigate the universality of the five-factor model of mindfulness and the measurement equivalence of the Five-Facet Mindfulness Questionnaire (FFMQ).

Methods The study used FFMQ data from published and unpublished research conducted in 16 countries (total $N = 8541$). Using CFA, different models, proposed in the literature, were fitted. To test the cross-cultural equivalence of the best fitting model, a multi-group confirmatory factor analysis was used. Further, the equivalence of individual facets of the FFMQ and potential sources of non-equivalence was explored.

Results The best fitting models in most samples were a five-facet model with a higher-order mindfulness factor and uncorrelated positive and negative item-wording factors and a five-facet model with a correlated facets and uncorrelated positive and negative item-wording factors. These models showed structural equivalence, but did not show metric equivalence (equivalent factor loadings) across cultures. Given this lack of equivalent factor loadings, not even correlations or mean patterns can be compared across cultures. A similar pattern was observed when testing the equivalence of the individual facets; all individual facets failed even tests of metric equivalence. A sample size weighted exploratory factor analysis across cultures indicated that a six-factor solution might provide the best fit across cultures with acting with awareness split into two factors. Finally, both the five- and six-factor solution showed substantially better fit in more individualistic and less tight cultures.

Conclusions Overall, the FFMQ has conceptual and measurement problems in a cross-cultural context, raising questions about the validity of the current conceptualization of mindfulness across cultures. The results showed that the fit of the FFMQ was substantially better in individualistic cultures that indicate that further data from non-Western cultures is needed to develop a universal conceptualization and measurement of mindfulness.

Keywords Mindfulness · FFMQ · Culture

Is mindfulness a cultural universal? Mindfulness is a principle originating from Buddhist tradition, which was first exported to the West and subsequently exported back to the East in the form of therapeutic interventions and psychological measurements. How valid are such measures to capture mindfulness across cultures? What can cross-cultural research using these measures reveal about mindfulness as a potentially universal psychological trait? In Western psychology, mindfulness is often defined as "paying attention in a particular way; on purpose, in the present moment, and nonjudgmentally" (Kabat-Zinn 1994, p.4). Such definitions provided ground for the development of a broad range of mindfulness measures. A combined analysis of multiple available measures resulted in the widely used Five-Facet Mindfulness Questionnaire (FFMQ; Baer et al. 2011), which conceptualizes mindfulness as a higher-order factor subsuming five facets: acting with awareness, non-judging, non-reacting, describing, and observing. The measure has been employed in

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12671-020-01333-6) contains supplementary material, which is available to authorized users.

✉ Johannes A. Karl
Johannes.karl@vuw.ac.nz

Extended author information available on the last page of the article

many different cultures, and translations exist in major language groups including German, Spanish, Portuguese, and Chinese. The continued research on mindfulness across cultures using the FFMQ indicates an implicit claim to universalism of the five-facet structure of mindfulness. The current approach to the measurement of mindfulness can be considered an imposed-etic approach (Berry 1989), because mindfulness as a concept originated in a specific Buddhist context, but was transformed into a measurement instrument through a Western lens in measures, such as the FFMQ, and subsequently exported globally to assess mindfulness in different cultures.

A crucial step to support the universality of the construct of mindfulness is to establish measurement equivalence across cultural groups. Equivalence in the current context refers to the comparability of measured scores between cultures and can be broken down into three levels that can be empirically assessed: structural equivalence, metric equivalence, and scalar equivalence (Fontaine 2005; Van De Vijver and Leung 2011). Structural equivalence implies that the same items can be used to measure the same latent constructs across cultures (Fischer and Fontaine 2010). In other words, measures show structural equivalence if the same items are used across cultures and these items form the same dimensional structure in all cultures. For example, the item "I'm good at finding words to describe my feelings" would be associated with the describing facet in all cultures. Metric equivalence implies that items have similar loading strength on the underlying constructs. For example, the item "I'm good at finding words to describe my feelings." would be an equally good indicator of the describing facet in all cultures (the factor loadings are statistically similar). Finally, scalar equivalence implies that the item intercepts are identical. In other words, respondents with the same level of mindfulness overall would answer identically to each individual question in all cultures and their answers are not affected or shifted by response biases such as acquiescence bias (yes-saying), different referent standards (e.g., reference group effects), or differences in social desirability of a construct across groups (Heine et al. 2002; Van De Vijver and Leung 2011).

Importantly, these levels of equivalence address the measurement properties of a scale across groups, but do not provide insight into potential domain under-representation across groups. Domain under-representation is present if a concept differs in conceptual scope across cultures, by missing important theoretical elements of the construct within specific cultural settings. In the case of mindfulness, during its transition from a Buddhist context into a Western secular context, metaphysical elements were often omitted to increase the diffusion of the practice (Kucinskas 2014, 2018).

Whether the FFMQ is equivalent, and at which level, holds important implications for cross-cultural research on mindfulness. Structural equivalence allows exploration of the basic structure of a measure, i.e., if items relate to the proposed theoretical variable. Metric equivalence allows for the cross-cultural comparison of the correlations and score patterns, but no conclusions about cultural differences in mindfulness as a theoretical construct can be made. Only under the condition of scalar equivalence can researchers directly compare mean scores. In other words, researchers can investigate the dimensionality of mindfulness with structural equivalence, can compare the relationship of mindfulness with other measures across cultures with metric equivalence, and can directly compare cultural groups with scalar equivalence. Non-equivalence across a large number of cultural groups indicates that the FFMQ is not a suitable tool for cross-cultural research and that further research is necessary to establish a conceptualization of mindfulness that is valid across cultures.

A crucial part of testing for equivalence is to determine the theoretical structure which can be tested across groups. For the FFMQ, a number of structures have been suggested: a five-facet model in which the individual facets are subsumed under one (Baer et al. 2006) or two higher-order factors (Tran et al. 2013), and five correlated facets without a higher order factor (Van Dam et al. 2012). Further, a number of studies have suggested that the FFMQ should be modeled with positive and negative item-wording factors (Aguado et al. 2015; Van Dam et al. 2012). These item-wording factors model participants' differential responding to positively and negatively worded questions, improving the fit of the structure.

Examining most of the above described possibilities, the FFMQ could be modeled as (1) five correlated facets with no higher order factor, (2) five correlated facets with uncorrelated methods factors for negatively and positively worded items, (3) five facets subsumed under a single higher-order factor, (4) five facets subsumed under a single higher-order factor with uncorrelated methods factors, (5) five facets subsumed under a single higher-order factor with correlated methods factors, or (6) five facets subsumed under a single higher-order factor with correlated method factors which in turn are also correlated with the higher-order factor. A visualization of the proposed models are found in Fig. 1. Overall, the first step in determining whether the FFMQ is equivalent across cultures is determining the best fitting model within each culture. This analysis will provide first insights into the best conceptual representation of mindfulness across the different contexts in which the instrument has been applied. The most common structure can then be directly tested across all the sites for which data is available.

What is needed is an examination of contextual variables that may influence the replicability of the structure of the FFMQ across groups. This can be achieved by focusing on three major cultural dimensions that might be of relevance for mindfulness. First, monumentalism-flexibility (Minkov et al. 2018) captures important aspects of the stability of self. Minkov et al. (2018) described this important culture-level

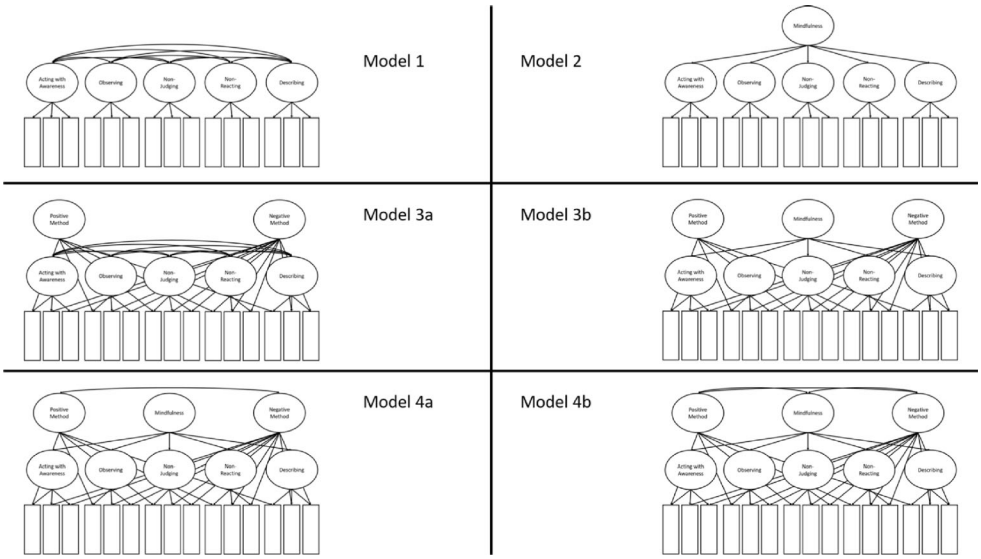

Fig. 1 Models tested in each country

axis as: "Monumentalism is a metaphor for a cultural tendency to encourage people to be like a monolithic monument: proud, stable, and consistent (made of the same substance outside and inside). Flexibility is the opposite cultural tendency, favoring a modest self-regard, duality, and adaptability." (p. 12). In other words, monumentalism can be thought of representing self-consistency (being the same person regardless of context). In contrast, flexibility is related to situation-specific behavioral responses, similar to the concept of 'face' in Asian societies which requires sensitivity to relationships and being a different person depending on the current context (Hwang 1987). In line with this reasoning, Minkov et al. (2018) found that countries in Asia (e.g., China, Korea, Singapore) score high on flexibility, whereas Western cultures cluster around the midpoint of this dimension. Therefore, if the FFMQ is closer to the ideal structure in Asian cultures, this would indicate that the proposed structure of the FFMQ represents an Eastern rather than a Western concept of mindfulness.

Further, individualism-collectivism expresses the level of embeddedness of the individual in the wider society (Hofstede 2001; Triandis 1995). Cultures in the West tend to have looser connections between the individual and the in-group; therefore, they score higher on individualism. In contrast, individuals in Eastern cultures tend to be more embedded in the wider in-group; therefore, those cultures tend to score lower on individualism (Hofstede 2001; Minkov et al. 2017). A positive relationship between individualism and structural fit can be taken as an indication that the structure of mindfulness proposed in the FFMQ reflects a Western concept of mindfulness.

Tightness-looseness represents societal tendencies to be judgmental and punitive to deviations from cultural norms (Gelfand et al. 2011; Uz 2015). The greater tendencies of individuals in looser cultures to be non-judgmental of one's own and others' deviations from norms might be more in line with current definitions of mindfulness underlying the FFMQ that see mindfulness as non-judgmental awareness (Kabat-Zinn 1994). The expectation is therefore that societal looseness is associated with greater fit of the five-facet structure, indicating better fit of the FFMQ in cultures less judgmental of norm deviations.

In summary, the current research had three main goals. First, to determine the best fitting structure of the FFMQ in the individual samples and cultures. Second, to test this model across cultures to determine structural, metric, and scalar equivalence, which provides insights about whether the construct of mindfulness can be compared across cultures and what kind of comparisons can be made. The third and final goal was to examine what contextual variables may influence the stability and replicability of mindfulness measurement.

Method

Participants

The sample contained 8541 participants from 16 countries sourced from previously published and unpublished studies: Australia ($N = 165$; community adults, Beshara et al. 2013), Austria ($N = 973$; community adults and students, Tran et al. 2013), Chile ($N = 398$; students, Schmidt and Vinet 2015), China ($N = 215$; community adults, Ma et al. 2018), Germany ($N = 529$; students, Michalak et al. 2016), Spain ($N = 1155$; adults and students, Aguado et al. 2015), Hong Kong ($N = 536$; adults and students, Chung et al. 2014; Wong et al. 2017), Croatia ($N = 242$; adults and students, Gračanin et al. 2017), India ($N = 300$; community adults, Mandal et al. 2016), Norway ($N = 466$; adults and students, Dundas et al. 2013; Solem et al. 2015), New Zealand ($N = 399$, students, Karl and Fischer 2019), Poland ($N = 702$; students, Radoń 2014), Portugal ($N = 251$; community adults,

Ramos et al. 2018), Romania ($N = 293$; adults and students, Astani 2016; Druica and Ianole-Calin 2018), Sweden ($N = 495$; students, Lilja et al. 2011), and the United States of America ($N = 1422$; students, sample 1,2: Verhaeghen 2018; sample 3,5: Verhaeghen 2019; sample 4: Verhaeghen and Aikman 2020). These studies were identified through a literature search on Google Scholar, PsychInfo, and the Web of science.

Measures

The analysis was conducted on the FFMQ-39 (Baer et al. 2006). All questions were measured on a 1–5 scale with verbal anchors in the respective language of the questionnaire (English: *Never or very rarely true*, *Rarely true*, *Sometimes true*, *Often true*, *Very often or always true*). The FFMQ was administered by the original authors in the language relevant to the cultural context. Across all facets and countries, the FFMQ showed acceptable to excellent reliabilities. All facets showed similar average reliability, with non-reacting showing the lowest average reliability. Due to space constraints, the full reliability table can be found in the supplementary material (Table 1).

Data Analyses

The data analysis has been pre-registered on the OSF (https://osf.io/nftxb/). All deviations from the proposed analysis are indicated where necessary. All CFAs and multi-group confirmatory factor analyses (MGCFAs) were fitted using lavaan (Rosseel 2012) in R (R Core Team 2018), following the procedures set out by Fischer and Karl (2019). Each model was fitted with an MLM estimator to adjust for multivariate non-normality. Further, the variance of all latent variables was fixed to unity to allow estimation of all factor loadings, rather than fixing one item's loading to 1.

Testing the Structure of the FFMQ The first step was to test the individual proposed models of the FFMQ outlined in the introduction (Fig. 1 provides a graphical summary). To determine the best fitting model in each sample, separate CFAs were fitted for each of the models in each sample and compared for the relative fit. For each model, the following fit indices were reported: χ^2, degrees of freedom, χ^2/degrees of freedom (for a discussion see: Rasch 1980), RMSEA (Steiger 2016) with confidence intervals, SRMR, Comparative Fit Index (CFI; Bentler 1990), $\widehat{\gamma}$ (Fan and Sivo 2007), and Bayesian Information Criterion (BIC; Schwarz 1978). Acceptable fit for CFI and $\widehat{\gamma}$ was defined as $> .90$ and good fit was defined as $> .95$ (Marsh et al. 2004), acceptable fit for the SRMR $< .08$ (Hu and Bentler, 1999). The RMSEA was evaluated following MacCallum et al. (1996), with less than 0.01, 0.05, and 0.08 to indicate excellent, good, and mediocre fit, respectively. To compare nested models, changes in $\Delta\text{CFI} > .01$ and $\Delta\widehat{\gamma} > .001$ were used as indicating acceptable fit (Cheung and Rensvold 2002). Further, the Bayesian information criterion (BIC) was used as the deciding criterion. Reductions of 10 between models were taken a strong indication of improvement and a reduction of 5 as a moderate indication of improvement (Berchtold 2019; Raftery 1995). Further, Vuong's test of non-nested model comparison with ML estimator (Vuong 1989) implemented in the nonnest2 package (Merkle and You 2018) was used to supplement the judgement whether models showed improved fit. The model that showed improved fit from the previous model in the majority of samples was selected.

Testing the Equivalence of the FFMQ A commonly employed method to test for measurement equivalence is MGCFA (Fischer and Karl 2019; Milfont and Fischer 2010). To test the equivalence of the ideal structure derived in the individual CFAs a MGCFA was used. If multiple samples were present for a culture, the individual samples that successfully converged in the previous step were merged to obtain an overall sample for each culture. To test for structural equivalence, item loadings and intercepts were allowed to vary between cultures. Structural equivalence was present if the model showed acceptable fit across all cultures. For metric equivalence, item loadings were constrained to be equal but the intercept was allowed to vary between cultures. Metric equivalence was present if the constrained model fits well and there was no substantial drop in model fit from the prior, less restricted model. Substantial drop in fit was defined as $\Delta\text{CFI} < .01$, a more stringent cut-off of $\Delta\text{CFI} < .002$, $\Delta\text{NCI} < .01$, and $\Delta\widehat{\gamma} < .001$ (Cheung and Rensvold 2002). Last, for scalar equivalence, the intercept to be was constrained to be equal between cultures. Scalar equivalence was present if the constrained model showed good fit and no substantial drop in fit from the metrically restrained model. The same criteria for deciding on model fit were used as for metric equivalence.

Exploratory Analyses In addition to these confirmatory analyses, several exploratory analyses were specified in the preregistration. First, in case that the overall FFMQ would not be equivalent, the equivalence of the five facets individually (acting with awareness, non-reacting, non-judging, observing, and describing) would be tested. Further, if no equivalence of the FFMQ in most cultures was found, alternative solutions using an exploratory factor analysis to determine the common factor solution across cultures would be explored. Finally, the effect of culture level variables, such as individualism, monumentalism, and tightness-looseness, on the appropriateness of the five-factor solution in different cultures was investigated. A Procrustes-analysis, examining the congruence of the loadings of each country's five-factor structure to an ideal solution where items' loadings on the factors were defined as

ones and zeros following the original proposed structure of the FFMQ (Baer et al. 2006) was run. Tucker's Φ (Tucker 1951) was extracted as a measure of similarity between the perfect matrix and the loading matrix of each country. Subsequently, the obtained congruence coefficients were correlated with Minkov's (2017, 2018) individualism-collectivism and monumentalism-flexibility axis as well as two indicators of tightness-looseness (Gelfand et al. 2011; Uz 2015) to investigate whether the fit of the FFMQ to the idealized structure differs systematically along these cultural dimensions. Data was obtained on both individualism-collectivism and monumentalism-flexibility for all countries from Minkov et al. (2017, 2018), except for Croatia. Tightness scores were obtained from Gelfand et al. (2011) which had data for all countries except four (Chile, Croatia, Romania, and Sweden). Looseness scores were obtained from Uz (2015), which had data on all countries except five (Australia, China, Hong Kong, Norway, and New Zealand).

Results

Following the prior outlined analysis plan, first the fit of the individual models for each individual sample was tested to determine the best fitting structure for the FFMQ. The FFMQ model with correlated facets (Fig. 1, model 1) converged successfully in all samples, but only showed good fit in the Portuguese and New Zealand samples, which together represented 8.70% of all samples (all results are reported in the supplementary material, Table 2). This indicates that the correlated five-facet model did not represent the underlying structure of the data in most samples.

The FFMQ model with five-facets subsumed under a higher-order factor (Fig. 1, model 2) showed good fit in two samples (8.70% of the samples), insufficient fit in 20 samples (86.96%), and failed to converge in one sample (4.35% of the samples, all results and comparison of fit with model 1 are reported in the supplementary material, model fit: Table 3, comparison Table 8). The change in model fit was examined based on the pre-registered criteria between model 1 and model 2 and which indicated that it did not improve the fit in any of the samples where the models converged (22 out of the 23 samples; 1 sample failed to converge). According to Vuong's test of non-nested models, model 1 showed better fit for 20 out of 23 samples. For two samples, no preferable model could be determined and one sample failed to converge. Overall, this indicates that the FFMQ model with a higher-order factor does not empirically fit the data better compared to the model with correlated facets.

A possibility for the low fit of the FFMQ with a higher-order factor could be the presence of positive and negative method factors identified in previous research (e.g., Aguado et al. 2015; Van Dam et al. 2012). The model with correlated facets and uncorrelated method factors (Fig. 1, model 3a) converged successfully in all samples, and showed good fit in 15 samples, which together represented 62.50% of all samples (all results are reported in the supplementary material, Table 4). The change in model fit was examined based on the pre-registered criteria between model 2 and model 3a and which indicated that it improved the fit in all samples where the models converged (22 out of the 23 samples; 1 sample failed to converge). According to Vuong's test of non-nested models, model 3a showed better fit for all samples that converged.. Therefore, the fit of the FFMQ with one higher-order factor and uncorrelated positive/negative method factors was examined next (Fig. 1, model 3b). This model showed acceptable fit in 12 samples (52.17% of the samples), insufficient fit in eight samples (34.78% of the samples), and failed to converge in three samples (13.04% of the samples). The full results are reported in the supplementary material (Table 5).

Model 3b showed no difference from model 3a based on CFI,$\widehat{\gamma}$, but was favoured by the BIC. This was supported by Vuong's test of non-nested models which indicated that for 15 samples (62.50%) no preferred model could be found. We therefore additionally tested the fit of model 3b against the prior models to determine whether the fit increased.. Because model 2 (facets subsumed under higher-order factor) did not show increased fit compared to model 1 (correlated facets), the fit of model 3b (FFMQ with a higher-order factor and positive/negative methods factors) was first compared against model 1 (correlated facets). Model 3b showed substantially higher CFI, $\widehat{\gamma}$, and a substantial reduction in BIC for 20 samples (86.96% of all samples), but three samples failed to converge (13.04% of all samples). The comparison of model 3b against model 2 yielded similar results, indicating improved fit of the FFMQ with uncorrelated method factors. This result was further supported by Vuong's test of non-nested models which indicated better fit of model 3b for 20 samples (86.96% of all samples) while three samples failed to converge (13.04% of all samples). This indicates that positive and negative wording method factors were present in most samples and should be modeled.

While the previous finding indicated that the introduction of method factors substantially improves the fit of the FFMQ, it was unclear whether these method factors should be correlated or uncorrelated with each other. Therefore, model 4a allowed the method factors to be correlated with each other. This model showed acceptable fit in 11 samples (47.83% of samples), insufficient fit in six samples (26.09% of samples), and failed to converge in 6 samples (26.09% of the samples). The full results are reported in the supplementary material (Table 6). The model with correlated method factors showed improved fit compared to the model with uncorrelated method factors in one sample (4.35% of all samples), no improvement in fit in 16 samples (69.57% of all samples), and six samples did not converge (26.08% of all samples). Vuong's test of non-

nested model yielded similar results with model 4a fitting better in two samples (8.70% of all samples), no clear preference between models for 15 samples (65.22% of all samples), and 6 samples (26.09% of all samples) did not converge. Overall, this indicates that the model with correlated method factors does not fit better than the model with uncorrelated method factors in most samples. Furthermore, the uncorrelated method factor model was preferable as it was conceptually simpler and the most parsimonious model.

Last, the model in which the method factors were not only allowed to correlate with each other, but also with the higher-order factor of mindfulness (model 4b), showed acceptable fit in 11 samples (47.83% of samples), insufficient fit in four samples (17.39% of samples), and failed to converge in eight samples (34.78% of samples). The full results are reported in the supplementary material (Table 7). In summary, the FFMQ models with uncorrelated positive and negative method factors showed the best fit in the individual samples. Because model 3a (correlated facets with methods factors) and model 3b (facets subsumed under a higher order factor with methods factors) could not be differentiated in most samples both models were selected to be tested for the cross-cultural equivalence of the FFMQ.

Cross-cultural Equivalence of the FFMQ

In the next step, the cross-cultural equivalence of the five-facet model with higher-order factor and uncorrelated method factors was examined. As specified in the pre-registration, the equivalence of the model in all samples that converged in the previous analysis and showed acceptable fit in the individual analysis of fit was examined (the analysis was also run for all countries that successfully converged and showed an identical result. The results are available on the OSF page of this project. If multiple samples per country were available, these were merged to obtain an overall dataset for each country. One sample from the US was excluded, because the model did not converge, and one sample from Norway had to be excluded due to bad fit. No data from Australia, India, Hong Kong, China, Poland, Romania, or Chile were included due to all samples either having bad fit or the model not converging. This left data from only Western countries: Portugal, New Zealand, Germany, USA, Austria, Croatia, Spain, Sweden, and Norway.

Initially, an unconstrained model was fitted to formally test for structural equivalence. This model showed good fit ($\chi^2(5922) = 10{,}097.470$, $\chi^2/df = 1.705$, CFI = .943, RMSEA = .038[.036, .039], SRMR = .056, BIC = 526,262.200, $\widehat{\gamma} = .962$), indicating that the model was structurally equivalent across cultures. This was the baseline model for the further comparisons. To test metric equivalence, the same model was fitted across cultures, but with all factor loadings on the substantive factors constrained to be equal across cultures (loadings on the method factors were allowed to vary freely, see Van Dam et al. 2012). While the model showed acceptable fit by itself ($\chi^2(6274) = 11{,}250.640$, $\chi^2/df = 1.793$, CFI = .932, RMSEA = .040[.039, .041], SRMR = .071, BIC = 524,536.000, $\widehat{\gamma} = .954$), it nevertheless showed a substantial drop from the unconstrained model (ΔCFI = − .011, $\Delta\widehat{\gamma} = -.008$) across all countries, indicating that the FFMQ was not metrically equivalent across cultures. We repeated the analysis for model 3a for all samples that showed good individual fit and found identical results. While the model showed acceptable fit by itself ($\chi 2$ (7183) = 11927.759, $\chi 2$ /df = 1.661, CFI = .943, RMSEA = .038[.036, .039], SRMR = .054, BIC = 602826.531,$\widehat{\gamma}$ = .962), it nevertheless showed a substantial drop from the unconstrained model (ΔCFI = − .010, $\Delta\widehat{\gamma} = -.007$). Metric equivalence of a model in which all paths including the method factors were constrained was also tested. This analysis yielded an identical result. The results are available on the OSF page of this project. Based on these results, no further test for scalar equivalence was conducted since the data already failed metric equivalence tests.

Exploratory Analyses

Equivalence of the Individual Facets of the FFMQ The pre-registration specified that in case of poor equivalence of the overall FFMQ, the equivalence of the individual facets would be tested. First, CFAs were ran for the separate facets in each sample to determine which samples should be included in the equivalence analysis (due to space constraints the fit for all samples and all facets is reported in the supplementary material, Table 9). All samples that showed adequate fit across CFI, RMSEA, SRMS, and $\widehat{\gamma}$ were included. If multiple samples in a country showed good fit those were subsequently merged, and the equivalence analysis run across countries.

Acting with Awareness Acting with awareness did not show acceptable fit in any of the samples, indicating that a unidimensional structure of acting with awareness might not be the best fit in most samples.

Observing Observing showed a good fit in 82.61% of all samples, indicating that the uni-dimensional structure of observing fits well in most samples. Because the structure of the observing facet did not fit well in the individual CFA, Australia and Poland were excluded from the equivalence analysis. Across the remaining countries, the model showed good structural equivalence ($\chi^2(280) = 682.084$, $\chi^2/df = 2.436$, CFI = .959, RMSEA = .058 [.052, .063], SRMR = .036, BIC = 165, 751.165, $\widehat{\gamma} = .986$). Nevertheless, when tested for metric equivalence, the model showed a substantial drop in fit (ΔCFI = − .016, $\Delta\widehat{\gamma} = -.006$), indicating that the observing facet was

not metrically equivalent across the samples studied here. Overall, this indicates that while the observing items measure a single construct in most countries, the individual items were not equally good indicators in each country.

Non-Judging The non-judging facet showed good fit of the structure in 43.48% of all samples, indicating that the unidimensional structure of non-judging did not fit well in the majority of samples. For the equivalence analysis, Australia, Austria, Chile, Hong Kong, Spain the USA, and Germany were excluded because no sample from these countries showed good fit. Across countries, the model showed good structural equivalence ($\chi^2(180) = 359.322$, $\chi^2/df = 1.996$, CFI = .976, RMSEA = .061 [.051, .070], SRMR = .033, BIC = 70,728.890, $\widehat{\gamma}$ = .986). Nevertheless, when tested for metric equivalence, a significant drop in model fit (ΔCFI = − .017, $\Delta\widehat{\gamma}$ = − .011) was found, implying that the factor loadings were not identical.

Describing The describing facet showed good fit of the structure only in 4.35% of all samples. The only sample where the describing facet showed good fit was Austria. This excluded any test for measurement equivalence.

Non-Reacting The non-reacting facet showed good fit of the structure in 52.17% of all samples, indicating that the unidimensional structure of non-reacting fits well in the majority of samples. Samples from Germany, Austria, Croatia, Chile, China, Poland, Hong Kong, and Spain were excluded from the further equivalence analysis because the samples did not show acceptable fit in the individual analyses. Across the remaining countries, the model showed good structural equivalence ($\chi^2(112) = 221.483$, $\chi^2/df = 1.978$, CFI = .968, RMSEA = .058 [.047, .069], SRMR = .035, BIC = 58,255.100, $\widehat{\gamma}$ = .990). Nevertheless, when tested for metric equivalence, a significant drop in model fit (ΔCFI = − .013, $\Delta\widehat{\gamma}$ = − .005) was found indicating that the non-reacting facet was not equivalent across cultures.

To summarize the previous analysis, it showed that no single facet of the FFMQ exhibits metric equivalence across all available countries. Further, both acting with awareness and describing did not show good CFA fit when investigated separately from the overall structure of the FFMQ, suggesting that these facets might not be uni-dimensional. Overall, this analysis parallels the finding on the overall structure of the FFMQ. This indicates that neither the FFMQ as a whole nor the individual facets are sufficiently cross-culturally equivalent to allow for cross-cultural comparison of means or even correlations with other constructs.

Alternative Structure of the FFMQ While the FFMQ model with a higher-order factor of mindfulness and uncorrelated methods factors showed good fit in most cultures, a number of cultures, mostly non-Western, still showed below acceptable fit. Overall, both on the level of the total FFMQ and the individual facets, no metric equivalence was found, which indicates that individual items do not load in the same way on the underlying constructs across cultures.

Therefore, an exploratory analysis was conducted to examine the possibility of an alternative structure of the FFMQ across cultures. A sample-size weighted average correlation matrix of the FFMQ items across all cultures was computed and the ideal number of components to be extracted from the correlation matrix calculated using parallel analysis (Dinno 2018; Horn 1965). The parallel analysis indicated that six components should be extracted (adjusted Eigenvalues: 7.288, 4.587, 2.609, 2.172, 1.849, 1.042).

As specified in the pre-registration, two separate PCAs were run, once allowing for correlated components using an oblimin rotation and one forcing components to be orthogonal using a varimax rotation. The results for the varimax rotation are reported in the supplementary material (Table 10); the results of the oblimin rotation are reported on the OSF page of this project, as they were nearly identical. The overall factor structure in the combined sample suggested that four of the five facets emerged, but that the acting with awareness items loaded on two separate factors. One factor was defined by the acting with awareness items focusing on behavior, whereas the second factor was defined by presence items. The results of the cross-cultural PCA indicate that a six-factor structure might fight better across cultures compared to the five-factor structure.

Sources of Incongruence in the Structure in the FFMQ

To test the possibility that the previously proposed FFMQ structure was systematically linked to culture-level variables such as individualism-collectivism and monumentalism-flexibility, the five-factor solution of each country was rotated towards an idealized loading matrix. The average Φ ranged from .850 to .954 for the individual countries (all results are in Table 11 in the supplementary material), where .90 can be considered good fit (Fischer and Fontaine 2010). Overall, 11 countries (Australia, Austria, Chile, China, Germany, Spain, Norway, New Zealand, Portugal, Sweden, USA; 68.75% of all countries) showed good congruence to the ideal structure. To test whether the level of congruence with the ideal matrix can be predicted using country-level cultural information, the average Tucker's Φ in each country was correlated with individualism, monumentalism, and tightness vs looseness scores. Individualism ($r = .77$ [.44, .92], $p < .001$), but not flexibility ($r = -.18$ [− .63, .37], $p = .52$) was significantly related to greater congruence with the proposed structure. Individualism predicted 60.02% of the variance in average congruence. Further, average congruence was significantly related to looseness ($r = .77$ [.32, .94], $p < .05$) measured with

the indicator by Uz (2015), indicating better fit in looser cultures and explaining 59.58% of variance in average congruence. When measured with the tightness indicator by Gelfand et al. (2011), the relationship was significant ($r = -.59$ [−.87, −.03], $p < .05$) and in the same direction indicating greater fit in looser cultures explaining 28.88% of the variance.

Finally, the fit of the individual countries to the pooled solution was examined by rotating the six-factor solution in each country towards the pooled loading matrix. The average Φ ranged from .883 to .980 for the individual countries. Overall, only two countries (Hong Kong and India; 12.50% of all countries) showed below acceptable congruence to the pooled structure (all results can be found in the supplementary material, Table 12).

To test whether the level of congruence with the ideal matrix can be predicted using country-level cultural information, the average Tucker's Φ across all dimensions in each country was correlated with individualism, monumentalism, and tightness-looseness. Individualism ($r = .75$ [.39, .91], $p < .001$), but not flexibility ($r = -.23$[−.66, .32], $p = .42$) was significantly related to greater congruence with the proposed structure. Individualism predicted 56.45% of the variance in average congruence towards the pooled structure. In regard to tightness-looseness, the average congruence was not significantly related to looseness ($r = .35$ [−.31, .79], $p = .29$; data by Uz 2015). For the tightness indicator by Gelfand et al. (2011), the relationship was in the same direction but not significant ($r = -.41$ [−.80, .21], $p = .18$). Overall, this indicates that the five-factor structure of the FFMQ was a better representation of the underlying structure of mindfulness in more individualistic, loose Western rather than more collectivistic, tight non-Western countries. While the six-factor structure of the FFMQ shows no bias based on tightness-looseness, it was still biased in favor of individualistic cultures.

Discussion

The current study used a large multi-national data set to provide a systematic analysis of the cross-cultural equivalence of the FFMQ across a wide range of cultures. The main findings were (a) the FFMQ structure did not adequately fit across cultures, even when including separate method factors, (b) the acting with awareness facet broke apart into a behavioral and a presence factor in an exploratory analysis, and (c) the ideal structure of the FFMQ might be driven by cultural values.

Implications for the Modeling of the FFMQ

The FFMQ showed substantially better fit in most countries if it was modeled with positive and negative item-wording factors. These findings support previous research (e.g., Aguado et al. 2015; Van Dam et al. 2012) on the presence of item-wording factors in the FFMQ. These item-wording effects might be more substantial for non-meditators (Van Dam et al. 2009); for an alternative explanation of these findings, see Baer et al. (2011). Overall, this indicates that the inclusion of item-wording factors substantially improved the fit compared to the model proposed by Baer et al. (2006) in the majority of samples. The current results also suggest that these item-wording factors are most likely orthogonal. The presence of item wording factors echoes concerns in the literature about scales, such as the MAAS, that measure mindfulness with only negatively scored items (Grossman 2011). The use of scales that are not balanced for wording might conflate response tendencies to negatively worded items with substantial variance in mindfulness (for an example of potential variables influencing responses to positive and negative items, see Michaelides et al. (2016).

Across cultures, the FFMQ items were best represented as a six-factor structure, with acting with awareness divided into awareness of thoughts and awareness of actions. This differentiation of the acting with awareness facet suggests that two different processes might underlie this factor and that it should not be treated as a uni-dimensional construct across cultures. The two-factor structure that emerged resembles the distinction made by researchers of consciousness about private cognitive spaces, in other words awareness of the external world and one's behavior in it, and public cognitive spaces, in other words awareness of internal world, e.g., thoughts and images (Gray 2004). The first sub-factor of acting with awareness was characterized by items indicating awareness of one's behavior, aligning with public cognitive spaces. The second sub-factor of acting with awareness was characterized by items specific to one's mental processes, aligning with private cognitive spaces. Previous research on the effect of body focused meditation showed that this meditation practice can impair metacognitive efficiency (Schmidt et al. 2019). Using a two-factor structure of acting with awareness separating thought from action awareness might provide further insight into the relationship of mindfulness, body-awareness, and metacognition.

Implications for Cross-cultural Comparisons

The second aim of the current study was to test the cross-cultural validity of the FFMQ. The cross-cultural equivalence of the FFMQ with uncorrelated methods factors was examined and the results indicated that this model shows good structural equivalence across the different countries in which the individual CFAs showed good fit. The items were related to the proposed theoretical facets (e.g., showed non-trivial loadings). Nevertheless, no support for metric equivalence was found. This indicates that the items of the FFMQ are not equally good indicators of the individual facets across

countries. This non-equivalence precludes both comparisons of correlations between the FFMQ and other variables of interest across countries as well as direct or indirect (profile) mean comparisons between cultures. The FFMQ in its current form is not a suitable tool to assess mindfulness in a cross-cultural context. This does not preclude the use of the FFMQ in mono-cultural studies but highlights the need for a cross-culturally valid measure of mindfulness. Initial steps should start with explicit considerations whether mindfulness is an emic (culture specific) or etic (universal) concept (Berry 1989; Farh et al. 2006). Current research practice treats mindfulness as de-facto etic construct with scales largely developed in a western context and subsequently translated or adapted into other languages.

One potential reason for this de-facto etic approach to mindfulness in the West is the tendency of Westerners to consider themselves to be essentially culturally neutral, meaning they think of themselves as not introducing a cultural bias into a psychological concept (Bellah et al. 1985). However, any psychological tests are potentially shaped by the cultural environment in which they were first proposed. The case of mindfulness research shows that this assumption of cultural neutrality is not warranted and instead the concept became more individualistic, focused on personal freedom, and authenticity during the move of mindfulness practice from Asian contexts to North America (Purser and Milillo 2015; Wilson 2014). This individualization of mindfulness is not only reflected in theory, but also in the measurement of mindfulness. The current study explored whether the fit of the proposed structure of the FFMQ was systematically linked to previously identified cultural dimensions. The results indicated that cultures higher on individualism and looseness showed better congruence to the proposed structure of the FFMQ. While the effects of monumentalism-flexibility were not significant, they still showed an effect in the same direction indicating that more flexible cultures had worse fit to the overall structure. The findings suggest that the FFMQ may capture conceptualizations of mindfulness prevalent in Western and individualistic cultures compared to understandings of mindfulness in more collectivistic cultures, including some of the more collectivistic settings from which the concept originated. This is of concern since it indicates the presence of a systematic Western individualistic bias in the current FFMQ and highlights that to produce a cross-culturally valid measure of mindfulness, a translation or adaptation of currently used mindfulness measures might not be sufficient. More conceptual work is needed to adequately understand mindfulness across cultural contexts.

One promising approach for developing a more valid cross-cultural measure can be found in the development of the internationally validated positive and negative affective schedule (Thompson 2007), which used a mixture of qualitative and quantitative data from a wide range of cultures to determine items and factor structures of the new measure. Overall, to advance cross-cultural research on mindfulness, new measures should be developed utilizing an approach which includes diverse cultural perspectives to minimize the cultural bias of the measure.

Limitations and Future Research

The strength of the current study lies in the wide range of cultures captured in data set. The samples cover all permanently inhabited continents besides Africa which allows for the examination of the equivalence of the FFMQ from a broad perspective. The major limitation of the current study is the reliance on previously published data on the FFMQ rather than on representative samples from each country. Further, the analysis did not control for meditative experience of the participants, barring comparisons between meditators and non-meditators in different cultures.

Coming back to the initial questions, whether the FFMQ is a valid tool of measurement across cultures, results indicated general problems with the cross-cultural comparability even though the individual samples often showed acceptable fit when considered individually. Importantly, the exploratory analysis suggests that mindfulness as a construct might be biased towards individualistic Western interpretations of the construct. Overall, the FFMQ and the conceptualization of mindfulness in terms of five facets subsumed under a single overall construct might not be suitable for cross-cultural comparisons. To further develop the field of mindfulness research, both a closer exploration of the theoretical structure and cross-culturally valid measurement tools are necessary. Future research could collect data from a wide range of cultures on emic perspectives on mindfulness to aid the creation of a cross-culturally valid measure of mindfulness.

Authors' Contributions JK: designed and executed the study, analyzed the data, and wrote the paper. RF: collaborated with the design, analysis of the data, and writing of the study. All authors collected and contributed data, provided feedback on the paper, and approved the final version of the manuscript for submission.

Funding Information J.V.L has a Miguel Servet Type II contract awarded by the Institute of Health Carlos III (CPII19/00003).

Data Availability All raw data, the analytic code, and all materials associated with the study are available on the Open Science Framework (https://osf.io/nftxb/).

Compliance with Ethical Standards

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

Ethical Statement The analysis was based on previously published data.

References

Aguado, J., Luciano, J. V., Cebolla, A., Serrano-Blanco, A., Soler, J., & García-Campayo, J. (2015). Bifactor analysis and construct validity of the five facet mindfulness questionnaire (FFMQ) in non-clinical Spanish samples. *Frontiers in Psychology, 6*, 404. https://doi.org/10.3389/fpsyg.2015.00404.

Astani, A.-I. (2016). Mindfulness and unconditional self-acceptance as protective factors against thin ideal internalization. *Annals of A.I. I. Cuza University. Psychology Series, 25*(1), 37–46.

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*(1), 27–45. https://doi.org/10.1177/1073191105283504.

Baer, R. A., Samuel, D. B., & Lykins, E. L. B. (2011). Differential item functioning on the five facet mindfulness questionnaire is minimal in demographically matched meditators and nonmeditators. *Assessment, 18*(1), 3–10. https://doi.org/10.1177/1073191110392498.

Bellah, R. N., Madsen, R., Sullivan, W. M., Swidler, A., & Tipton, S. M. (1985). *Habits of the heart: individualism and commitment in American life*. Berkeley: University of California Press.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238.

Berchtold, A. (2019). Sequence analysis and transition models. In *Encyclopedia of Animal Behavior* (2nd ed., pp. 506–517). Elsevier: Academic Press. https://doi.org/10.1016/B978-0-12-809633-8.01241-3.

Berry, J. W. (1989). Imposed etics-emics-derived etics: the operationalization of a compelling idea. *International Journal of Psychology, 24*(2–6), 721–735. https://doi.org/10.1080/00207598908247841.

Beshara, M., Hutchinson, A. D., & Wilson, C. (2013). Does mindfulness matter? Everyday mindfulness, mindful eating and self-reported serving size of energy dense foods among a sample of South Australian adults. *Appetite, 67*, 25–29. https://doi.org/10.1016/j.appet.2013.03.012.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5.

Chung, P.-K., Zhang, C.-Q., & Zhang, C.-Q. (2014). Psychometric validation of the Toronto mindfulness scale – trait version in Chinese college students. *Europe's Journal of Psychology, 10*(4), 726–739. https://doi.org/10.5964/ejop.v10i4.776.

Dinno, A. (2018). Paran: Horn's test of principal components/factors. Retrieved 18 February, 2020, from https://cran.r-project.org/package=paran.

Druica, E., & Ianole-Calin, R. (2018). Simply clustering. Making new sense in the five facets mindfulness questionnaire. *Romanian Statistical Review, 66*(1), 61–81.

Dundas, I., Vøllestad, J., Binder, P.-E., & Sivertsen, B. (2013). The five factor mindfulness questionnaire in Norway. *Scandinavian Journal of Psychology, 54*(3), 250–260. https://doi.org/10.1111/sjop.12044.

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research, 42*(3), 509–529. https://doi.org/10.1080/00273170701382864.

Farh, J.-L., Cannella, A. A., & Lee, C. (2006). Approaches to scale development in Chinese management research. *Management and Organization Review, 2*(3), 301–318. https://doi.org/10.1111/j.1740-8784.2006.00055.x.

Fischer, R., & Fontaine, J. R. J. (2010). Methods for investigating structural equivalence. In D. Matsumoto & F. J. R. E. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 179–215). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511779381.010.

Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology, 10*, 1507. https://doi.org/10.3389/fpsyg.2019.01507.

Fontaine, J. (2005). Equivalence. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (Vol. 1, pp. 803–813).

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Ferrer, M., Fischlmayr, I. C., et al. (2011). Differences between tight and loose cultures: A 33-nation study. *Science, 332*(6033), 1100–1104. https://doi.org/10.1126/science.1197754.

Gračanin, A., Gunjača, V., Tkalčić, M., Kardum, I., Bajšanski, I., & Perak, B. (2017). Struktura usredotočene svjesnosti i njezina povezanost s crtama ličnosti i emocionalnim reagiranjem. *Psihologijske Teme, 26*(3), 675–700. https://doi.org/10.31820/pt.26.3.9.

Gray, J. A. (2004). *Consciousness: creeping up on the hard problem*. Oxford: Oxford University Press.

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. (2011). *Psychological Assessment, 23*(4), 1034–1040. https://doi.org/10.1037/a0022713.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: the reference-group effect. *Journal of Personality and Social Psychology, 82*(6), 903–918. https://doi.org/10.1037/0022-3514.82.6.903.

Hofstede, G. (2001). *Culture's consequences: comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks: Sage Publications.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447.

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), 1–55. https://doi.org/10.1080/10705519909540118

Hwang, K. (1987). Face and favor: the Chinese power game. *American Journal of Sociology, 92*(4), 944–974. https://doi.org/10.1086/228588.

Kabat-Zinn, J. (1994). *Wherever you go, there you are: mindfulness meditation in everyday life*. Hyperion. New York, U.S.A

Karl, J. A., & Fischer, R. (2019). Individual difference predictors of mindfulness. *PsyArXiv*. https://doi.org/10.31234/osf.io/z2cx6.

Kucinskas, J. (2014). The unobtrusive tactics of religious movements. *Sociology of Religion, 75*(4), 537–550. https://doi.org/10.1093/socrel/sru055.

Kucinskas, J. (2018). *The mindful elite: mobilizing from the inside out*. Oxford: Oxford University Press.

Lilja, J. L., Frodi-Lundgren, A., Hanse, J. J., Josefsson, T., Lundh, L.-G., Sköld, C., Hansen, E., & Broberg, A. G. (2011). Five facets mindfulness questionnaire—reliability and factor structure: a Swedish version. *Cognitive Behaviour Therapy, 40*(4), 291–303. https://doi.org/10.1080/16506073.2011.580367.

Ma, Y., She, Z., Siu, A. F.-Y., Zeng, X., & Liu, X. (2018). Effectiveness of online mindfulness-based interventions on psychological distress and the mediating role of emotion regulation. *Frontiers in Psychology, 9*, 2090. https://doi.org/10.3389/fpsyg.2018.02090.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure

modeling. *Psychological Methods, 1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130.

Mandal, S. P., Arya, Y. K., & Pandey, R. (2016). Validation of the factor structure of the five facet mindfulness questionnaire. *Indian Journal of Health & Wellbeing, 7*(1), 61–66.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2.

Merkle, E., & You, D. (2018). nonnest2: tests of non-nested models. Retrieved 18 February, 2020, from https://cran.r-project.org/package=nonnest2.

Michaelides, M. P., Koutsogiorgi, C., & Panayiotou, G. (2016). Method effects on an adaptation of the Rosenberg self-esteem scale in Greek and the role of personality traits. *Journal of Personality Assessment, 98*(2), 178–188. https://doi.org/10.1080/00223891.2015.1089248.

Michalak, J., Zarbock, G., Drews, M., Otto, D., Mertens, D., Ströhle, G., Schwinger, M., Dahme, B., & Heidenreich, T. (2016). Erfassung von achtsamkeit mit der deutschen version des five facet mindfulness questionnaires (FFMQ-D). *Zeitschrift Für Gesundheitspsychologie, 24*(1), 1–12. https://doi.org/10.1026/0943-8149/a000149.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: applications in cross. *International Journal of Psychological Research, 3*(1), 111–121. https://doi.org/10.1007/s11135-007-9143-x.

Minkov, M., Dutt, P., Schachner, M., Morales, O., Sanchez, C., Jandosova, J., Khassenbekov, Y., & Mudd, B. (2017). A revision of Hofstede's individualism-collectivism dimension: a new national index from a 56-country study. *Cross Cultural & Strategic Management, 24*(3), 386–404. https://doi.org/10.1108/CCSM-11-2016-0197.

Minkov, M., Bond, M. H., Dutt, P., Schachner, M., Morales, O., Sanchez, C., Jandosova, J., Khassenbekov, Y., & Mudd, B. (2018). A reconsideration of Hofstede's fifth dimension: new flexibility versus monumentalism data from 54 countries. *Cross-Cultural Research, 52*(3), 309–333. https://doi.org/10.1177/1069397117727488.

Purser, R. E., & Milillo, J. (2015). Mindfulness revisited. *Journal of Management Inquiry, 24*(1), 3–24. https://doi.org/10.1177/1056492614532315.

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Retrived 18 February, 2020, https://www.r-project.org/

Radoń, S. (2014). Validation of the Polish adaptation of the five facet mindfulness questionnaire. *Roczniki Psychologiczne/Annals of Psychology, 17*(4), 737–760.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111. https://doi.org/10.2307/271063.

Ramos, A., Rosado, A., Serpa, S., Cangas, A., Gallego, J., & Ramos, L. (2018). Validity evidence of the portuguese version of the five facet mindfulness questionnaire. *Revista de Psicología Del Deporte/Journal of Sport Psychology, 27*(2), 87–98.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Rosseel, Y. (2012). {lavaan}: an {R} package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Schmidt, C., & Vinet, E. V. (2015). Atención plena: validación del five facet mindfulness questionnaire (FFMQ) en estudiantes universitarios chilenos. *Terapia Psicológica, 33*(2), 93–102. https://doi.org/10.4067/S0718-48082015000200004.

Schmidt, C., Reyes, G., Barrientos, M., Langer, Á. I., & Sackur, J. (2019). Meditation focused on self-observation of the body impairs metacognitive efficiency. *Consciousness and Cognition, 70*, 116–125. https://doi.org/10.1016/J.CONCOG.2019.03.001.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136.

Solem, S., Thunes, S. S., Hjemdal, O., Hagen, R., & Wells, A. (2015). A metacognitive perspective on mindfulness: an empirical investigation. *BMC Psychology, 3*(1), 24. https://doi.org/10.1186/s40359-015-0081-4.

Steiger, J. H. (2016). Notes on the Steiger–Lind (1980) handout. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(6), 777–781. https://doi.org/10.1080/10705511.2016.1217487.

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology, 38*, 227–242. https://doi.org/10.1177/0022022106297301.

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*(9), 951–965. https://doi.org/10.1002/jclp.21996.

Triandis, H. C. (1995). *Individualism and collectivism*. Boulder: Westview Press.

Tucker, L. (1951). *A method for synthesis of factor analysis studies*. Washington, DC: Department of the Army.

Uz, I. (2015). The index of cultural tightness and looseness among 68 countries. *Journal of Cross-Cultural Psychology, 46*(3), 319–335. https://doi.org/10.1177/0022022114563611.

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*(5), 516–521. https://doi.org/10.1016/J.PAID.2009.05.005.

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*(2), 198–204. https://doi.org/10.1177/1073191112438743.

Van De Vijver, F. J. R., & Leung, K. (2011). Equivalence and bias: a review of concepts, models, and data analytic procedures. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology (pp. 17–45)*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511779381.003.

Verhaeghen, P. (2018). The mindfulness manifold: exploring how self-preoccupation, self-compassion, and self-transcendence translate mindfulness into positive psychological outcomes. *Mindfulness, 10*(1), 1–15. https://doi.org/10.1007/s12671-018-0959-3.

Verhaeghen, P. (2019). The examined life is wise living: the relationship between mindfulness, wisdom, and the moral foundations. *Journal of Adult Development*. https://doi.org/10.1007/s10804-019-09343-y.

Verhaeghen, P., & Aikman, S. N. (2020). How the mindfulness manifold relates to the five moral foundations, prejudice, and awareness of privilege. *Mindfulness, 11* 241–254. https://doi.org/10.1007/s12671-019-01243-2.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*(2), 307–333. https://doi.org/10.2307/1912557.

Wilson, J. (2014). *Mindful America: the mutual transformation of Buddhist meditation and American culture*. Oxford: Oxford University Press.

Wong, S. Y., Zhang, D., Li, C. C., Yip, B. H., Chan, D. C., Ling, Y., Lo, C. S., Woo, D. M., Sun, Y., Ma, H., Mak, W. W., Gao, T., Lee, T. M., & Wing, Y. (2017). Comparing the effects of mindfulness-based cognitive therapy and sleep psycho-education with exercise on chronic insomnia: a randomised controlled trial. *Psychotherapy and Psychosomatics, 86*(4), 241–253. https://doi.org/10.1159/000470847.

Affiliations

Johannes A. Karl[1] · Silvia Mariela Méndez Prado[2] · Asmir Gračanin[3] · Paul Verhaeghen[4] · Alexandre Ramos[5] · Satchit Prasun Mandal[6] · Johannes Michalak[7] · Chun-Qing Zhang[8] · Carlos Schmidt[9] · Ulrich S. Tran[10] · Elena Druica[11] · Stian Solem[12] · Andreea Astani[13] · Xinghua Liu[14] · Juan V. Luciano[15] · Mladenka Tkalčić[3] · Josefine L. Lilja[16] · Ingrid Dundas[17] · Samuel Y. S. Wong[18] · Ronald Fischer[1]

[1] Victoria University of Wellington, Wellington, New Zealand

[2] Escuela Superior Politécnica del Litoral, ESPOL, Faculty of Social Sciences and Humanities, ESPOL Polytechnic University, Guayaquil, Ecuador

[3] Faculty of Humanities and Social Sciences, University of Rijeka, Rijeka, Croatia

[4] Georgia Institute of Technology, Atlanta, GA, USA

[5] Lisbon, Portugal

[6] Defence Research and Development Organization (DRDO), Ministry of Defence, New Delhi, India

[7] Department of Psychology and Psychotherapy, Witten/Herdecke University, Witten, Germany

[8] Department of Sport and Physical Education, Hong Kong Baptist University, Kowloon Tong, Hong Kong

[9] Instituto de Estudios Psicológicos, Universidad Austral de Chile, Valdivia, Chile

[10] Department of Basic Psychological Research and Research Methods, School of Psychology, University of Vienna, Vienna, Austria

[11] Centre for Applied Behavioural Economics, University of Bucharest, Bucharest, Romania

[12] Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

[13] The County Center for Resource and Educational Assistance Bacău, Bacău, Romania

[14] Department of Psychology, Beijing Key Laboratory of Learning and Cognition, Capital Normal University, Beijing, China

[15] Teaching, Research & Innovation Unit, Parc Sanitari Sant Joan de Déu, St. Boi de Llobregat, Spain

[16] Department of Psychology, University of Gothenburg, Gothenburg, Sweden

[17] Department of Clinical Psychology, University of Bergen, Bergen, Norway

[18] School of Public Health and Primary Care, Chinese University of Hong Kong, Shatin, Hongkong