



# Applying Generalizability Theory to Differentiate Between Trait and State in the Five Facet Mindfulness Questionnaire (FFMQ)

Quoc C. Truong<sup>1</sup> · Christian U. Krägeloh<sup>2</sup> · Richard J. Siegert<sup>2</sup> · Jason Landon<sup>2</sup> · Oleg N. Medvedev<sup>1</sup>

Published online: 3 February 2020  
© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

**Objectives** Accurate evaluation of mindfulness-based training requires understanding of the differences between state and trait changes, and the Generalizability Theory (G-Theory) is the most appropriate method to differentiate these aspects in a measure. The Five Facet Mindfulness Questionnaire (FFMQ) is widely used measure of dispositional mindfulness, but its ability to accurately capture stable aspects of mindfulness has not been rigorously investigated using appropriate methodology.

**Method** G-Theory was applied to differentiate between trait and state aspects of mindfulness and to examine temporal reliability of the FFMQ in a sample of 83 participants who completed the scale at three occasions separated by 2-week intervals.

**Results** The total 39-item FFMQ and its short version FFMQ-18 have demonstrated good reliability in measuring trait mindfulness with G coefficients of 0.89 and 0.75, respectively, while individual facet subscales of the FFMQ appeared less reliable in measuring either trait or state. Subsequent analysis attempted to combine the FFMQ items that were least stable over time into a state mindfulness subscale. However, this did not result in acceptable psychometric properties for such a state subscale.

**Conclusions** The findings of this study indicate that reliable measurement of stable aspects of mindfulness can be achieved by using the full FFMQ scale or its short version FFMQ-18 with scores generalizable across sample population and occasions. The scores obtained on individual facet subscales of the FFMQ predominantly measuring trait mindfulness, but their reliability is affected by measurement error due to interaction between person, item, and occasion.

**Keywords** Mindfulness · State and trait · Measurement · Five Facet Mindfulness Questionnaire · Generalizability Theory

There has been a growing body of mindfulness research during the past 30-year period, with the methods and apparatus used in mindfulness studies steadily developing (Krägeloh et al. 2019). Mindfulness has been used in the development of a structured program to treat psychological symptoms such as stress, anxiety, and chronic pain (Kabat-Zinn 1982). Early studies showed evidence of the effectiveness of mindfulness treatment based on the changes of specific hypothesized

outcomes, such as melatonin levels (Massion et al. 1995), or increasing the effect of phototherapy and photochemotherapy in patients with the skin condition psoriasis (Kabat-Zinn et al. 1998). Later studies applying mindfulness-based interventions (MBIs) such as mindfulness-based cognitive therapy (MBCT; Segal et al. 2002) and mindfulness-based stress reduction (MBSR; Kabat-Zinn 1990) relied on self-report measures that were designed to evaluate the goals of those interventions such as burnout, life satisfaction (Shapiro et al. 2005), and depression (Ma and Teasdale, 2004). However, these earlier studies could not demonstrate expected changes in mindfulness levels to support their validity, which required development of reliable and valid instruments to assess the construct.

When evidence demonstrated the positive effects of MBIs in therapeutic settings (Bohlmeijer et al. 2010; Chang et al. 2004; Chiesa and Serretti 2009; Ledesma and Kumano 2009), research started focusing more on the application of mindfulness practice in many different

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s12671-020-01324-7>) contains supplementary material, which is available to authorized users.

---

✉ Oleg N. Medvedev  
oleg.medvedev@waikato.ac.nz

<sup>1</sup> School of Psychology, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand

<sup>2</sup> Department of Psychology, Auckland University of Technology, Auckland, New Zealand

contexts. Thus, the application of mindfulness in the workplace (Hyland et al. 2015), educational contexts (Bush 2011; Hwang et al. 2019), and sporting (Birrer et al. 2012) involved measurement of both mindfulness and related outcomes. Although alternative mindfulness assessments such as experience sampling (Frewen et al. 2014) or counting of breath (Levinson et al. 2014) have been proposed, self-report measures of mindfulness remain by far the most widely used method to assess mindfulness in research studies (Krägeloh et al. 2019). The importance of self-report mindfulness measures may be explained by subjective nature of human experience of the world, self and their interaction and a problem to derive such experience from more objective (e.g., neurophysiological) measures (Libet 2004).

The Five Facet Mindfulness Questionnaire (FFMQ; Baer et al. 2006) is a widely used psychometric measure of mindfulness including five subscales: *Act with Awareness*, *Describe*, *Nonjudge*, *Nonreact*, and *Observe*. To date, according to Google Scholar, the original FFMQ article has been cited over 5700 times since it was published. The growing popularity of the FFMQ may be explained by its ability to enhance exploration of specific mindfulness aspects and the growing body of validation studies supporting its robustness (Brown et al. 2015; Coffey et al. 2010; MacDonald and Baxter 2017; Medvedev et al. 2017b). A number of short versions of the FFMQ have been developed using the classical test theory (CTT) approach (e.g., Baer et al. 2012; Gu et al. 2016; Bohlmeijer et al. 2011), which were unable to address the limitations of ordinal scales such as limited precision and compatibility with parametric statistics (Allen and Yen 1979; Stucki et al. 1996). To address these problems, Medvedev et al. (2018) conducted a study to examine and compare the existing short versions of the FFMQ using Rasch analysis and proposed an 18-item FFMQ version (FFMQ-18).

Mindfulness can be defined as either state or a trait (Medvedev et al. 2017a). Growing evidence has shown that mindfulness practice causes both state and trait changes, and inability to differentiate clearly between the two may confound assessment results of MBIs (Tang et al. 2015). Trait or dispositional mindfulness is described as a relatively stable characteristic of an individual and reflects an ability to remain mindful across different situations and contexts (Baer et al. 2006; Davis et al. 2009). State mindfulness refers to a characteristic feature displayed in a given situation or time (Bishop et al. 2006; Lau et al. 2006; Tanay and Bernstein, 2013). While the FFMQ is widely considered as a measure of dispositional (trait) mindfulness, its ability to differentiate between dispositional and dynamic (state-like) aspects of mindfulness has not been carefully investigated using appropriate

methodology. Recently, Generalizability Theory (G-Theory) was proposed as the most adequate method to distinguish between state and trait aspects in a measure and to evaluate various sources of error variance and to establish generalizability of assessment scores as well as reliability of the instrument (Medvedev et al. 2017a; Paterson et al. 2017).

G-Theory was developed by Cronbach et al. (1963) and provides more advanced statistical method compared with classical test theory (CTT) methods for evaluating the reliability of psychometric assessments, such as rating scales and performance tests. G-Theory is able to evaluate specific sources of measurement error and generalizability of assessment scores to all possible circumstances using data obtained from a specific testing situation (Cronbach et al. 1963). Thus, G-Theory considers and estimates unique sources of error variance affecting the main variable of interest (e.g., a mindfulness score), while CTT considers error variance as a single factor and postulates that any measurement consists of true variance and error variance (Allen and Yen 1979). However, in complex natural environments, there are multiple sources of error that potentially influence the accuracy of measurement. For instance, Generalizability analysis will consider interactions between person and different factors including methodological (e.g., scale items) and situational (e.g., time of the day) that might each independently (or via interactions) contribute to the error of measurement. In summary, while CTT considers only one aspect of reliability (e.g., test-retest, inter-rater, internal consistency) at a time, G-Theory closely examines all these influences on reliability (including their interactions) simultaneously thus improving the methodology and precision of a psychometric assessment.

The traditional CTT approach to the state/trait distinction examines test-retest reliability coefficients to investigate temporal reliability of an instrument, which tends to be lower for a state measure (e.g., < 0.60) and higher for a measure of trait (e.g., > 0.70) (Ramanaiah et al. 1983; Spielberger et al. 1970; Spielberger 1999). Therefore, this method is based entirely on the total score correlations at two different time points (i.e., time 1 and time 2) and does not consider variability at individual item level and interactions between person, item, and occasion. Robust estimation of reliability requires consideration of the contributions made by item effects, scale effects, person effects, and occasion effects to the changes in the overall assessment score. Similarly, the intraclass correlation coefficient (ICC) that can be used to estimate temporal reliability has limited accuracy because it does not account for variability of individual items (Bloch and Norman 2012; Medvedev et al. 2017a).

G-Theory is a suitable approach to examine the distinction between trait and state components in an instrument

and comprehensively evaluate multiple sources of error variance (Medvedev et al. 2017a; Shavelson et al. 1989). A state is a dynamic aspect that results when a person interacts with an occasion, which is the unique adaptation of an organism to the momentary environment (Spielberger et al. 1970). Reliable distinction between dynamic and stable patterns of a construct or condition is important in both clinical and research contexts. For example, the accuracy of assessment could be affected by evaluating characteristics of a person while avoiding temporary changes (e.g., mood) and might lead to inappropriate conclusions. There should be a clear distinction between state and trait aspects of the presentation of a person in any psychometric measure, which requires identification and consideration of the relevant sources of error variance using appropriate psychometric techniques such as G-Theory (Bloch and Norman 2012; Paterson et al. 2017).

G-Theory partitions the overall variance into different parts related to particular sources and examines their impacts on the overall reliability (Cronbach et al. 1963). The proportions of specific parts can be used to quantify the contribution of a person variance reflecting a trait, and an interaction between person and occasion reflecting a state to the measurement (Medvedev et al. 2017a). By computing the ratios of state variance or trait variance to the sum of state and trait variance, we can reliably distinguish between state and trait components in a measure (Medvedev et al. 2017a; Paterson et al. 2017). Therefore, the current study was to apply G-Theory to examine the reliability of the FFMQ and its short 18-item version over time, distinguish between state and trait components of mindfulness items and subscales, as well as to identify sources of error that may affect the measurement. This research utilized a repeated-measures design with participants assessed at three occasions separated by equal two-week intervals. Application of G-Theory involved two parts, a Generalizability study (G-study) and Decision study (D-study). The G-study examined the overall generalizability and evaluated sources of error variance of the original FFMQ and its short version FFMQ-18 as well as its subscale scores. G-study computed a generalizability coefficient (*G* coefficient) for each scale under investigation, which is the overall measure of reliability representing the ratio of true person variance to the total variance of the data (Cardinet et al. 2011). The D-study was subsequently conducted to evaluate psychometric properties of individual items and their combinations to optimize reliability of the measurement and distinction between state and trait (Shavelson et al. 1989; Medvedev et al. 2017a). Data from D-study can be used to identify items that are reflecting state or trait aspects of mindfulness.

## Method

### Participants

The current sample included 83 university students who partook in the study on a voluntary basis and did not receive any payment or academic credit for their participation. The sample size satisfied requirements for reliability studies of this type of research (Shoukri et al. 2004). The sample included 22 males (26.5%) and 61 females (73.5%). From the total sample, ten participants (12%) engaged in regular meditation practice. The age of participants ranged from 18 to 47 years, with a mean of 21.34 (SD = 5.83). Ethnic groups were represented by 57% Caucasian, 11% Māori, 10% Pasifika, 6% Asian, and 17% others.

### Procedures

Participants completed the FFMQ items in class before the lecture or during a break and were instructed to return the completed forms to the researcher, submit it to a locked collection box at their faculty, or use a self-addressed pre-paid envelope to post their completed forms to the researcher university address. Each participant was required to complete the same questionnaire at three occasions with equal 2-week intervals. Respondents also provided demographic information such as sex, age, and ethnic group and to ensure anonymity were asked to include a personal code with three letters and three numbers to match the forms completed by the same participant at three occasions. This research was not expected to involve any risk, discomfort, or harm, and participants were informed about the nature of the study. The study was approved by the authors' university ethics committee.

### Measures

The FFMQ (Baer et al. 2006) consists of 39 items that assess aspects of mindfulness grouped into five subscales: *Act with Awareness*, *Describe*, *Nonjudge*, *Nonreact*, and *Observe*. Each individual item uses a 5-point Likert scale with options ranging from 1 = "Never or very rarely true" to 5 = "Very often or always true". There are 19 items that require reverse coding before conducting data analysis. After reverse coding, the total score and individual subscale scores are calculated by adding responses to the relevant items together (see Appendix A).

### Data Analyses

IBM SPSS Statistics 25 software was used to compute means, standard deviation (SD), Cronbach's alpha, test-retest coefficients, and ICC for the FFMQ, FFMQ-18, and individual subscales of the both FFMQ versions. Missing data comprised

0.04%, which were negligible and were replaced using mean imputation (Huisman 2000).

Generalizability analyses were conducted using EduG 6.1-e software (Swiss Society for Research in Education Working Group 2006) by following the guidelines described by Medvedev et al. (2017a). Both G-study and D-study used a random effect design: person (P) by item (I) by occasion (O), expressed as  $P \times I \times O$ , where the P and O facets are infinite and the facet I is fixed because the same set of items were used across all assessments using the FFMQ. In a G-study, all error variances are counted as 100% after controlling for person variance (P), which reflects true differences between persons. Person was the object of measurement (differentiation facet) and not a source of error, while I and O were instrumentation facets (Cardinet et al. 2011). The effects for all facets were presented by observed scores  $X$  which were calculated for the G-study (Shavelson et al. 1989):

$X = \mu + X_p + X_i + X_o + X_{pi} + X_{po} + X_{io} + X_{residual}$ ; where  $\mu$  is grand mean of  $X$

$$X_p = \mu_p - \mu \text{ (person effect)}$$

$$X_i = \mu_i - \mu \text{ (item effect)}$$

$$X_o = \mu_o - \mu \text{ (occasion effect)}$$

$$X_{pi} = \mu_{pi} - \mu_p - \mu_i + \mu \text{ (person} \times \text{item effect)}$$

$$X_{po} = \mu_{po} - \mu_p - \mu_o + \mu \text{ (person} \times \text{occasion effect)}$$

$$X_{io} = \mu_{io} - \mu_i - \mu_o + \mu \text{ (item} \times \text{occasion effect)}$$

$$X_{residual} = X_{pio} - \mu_{pi} - \mu_{po} - \mu_{io} + \mu_p + \mu_i + \mu_o - \mu$$

Each of the effects has estimated variance components, which were possible sources of error that might impact measurement and were calculated as follows:

$$\text{Person variance component: } \sigma_p^2 = (MS_p - MS_{pi} - MS_{po} + MS_{pio})/n_i n_o$$

$$\text{Item variance component: } \sigma_i^2 = (MS_i - MS_{pi} - MS_{io} + MS_{pio})/n_p n_o$$

$$\text{Occasion variance component: } \sigma_o^2 = (MS_o - MS_{io} - MS_{po} + MS_{pio})/n_i n_p$$

$$\text{Person} \times \text{item variance component: } \sigma_{pi}^2 = (MS_{pi} - MS_{pio})/n_o$$

$$\text{Person} \times \text{occasion variance component: } \sigma_{po}^2 = (MS_{po} - MS_{pio})/n_i$$

$$\text{Item} \times \text{occasion variance component: } \sigma_{io}^2 = (MS_{io} - MS_{pio})/n_p$$

Residual / person  $\times$  item  $\times$  occasion variance component:  $\sigma_{pio}^2 = MS_{pio}$ ; where MS stands for the mean of effect square and  $n$  represents facet sample size

Generalizability analysis estimates reliability using relative  $G$  coefficient ( $G_r$ ) and absolute  $G$  coefficient ( $G_a$ ) for the object of measurement (person). The relative model of measurement involves interpretation of test scores in a norm-referenced manner in which the score of a person is compared against the scores of others (Suen & Lei 2007; Vispoel et al. 2018).  $G_r$  accounts for a relative error variance ( $\sigma_\delta^2 = \frac{\sigma_i^2}{n_i} + \frac{\sigma_o^2}{n_o} + \frac{\sigma_{pio}^2}{n_i n_o}$ ; where  $n_i$  = number of items,  $n_o$  = number of occasions), which is

directly related to the object of measurement that may influence a relative measurement (e.g., person  $\times$  occasion and person  $\times$  item interactions) and includes divisions by desired sample sizes (Shavelson et al. 1989; Shavelson & Webb 1991):

$$G_r = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\delta^2}$$

The absolute model of measurement is based on the test scores, which are interpreted in a criterion-referenced manner where the score of a person is compared against some agreed-upon absolute standard.  $G_a$  is equivalent to the phi ( $\Phi$ ) coefficient, which is obtained after applying Whimbey's correction. It accounts for an absolute error variance ( $\sigma_\Delta^2 = \frac{\sigma_o^2}{n_o} + \frac{\sigma_i^2}{n_i} + \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{po}^2}{n_o} + \frac{\sigma_{io}^2}{n_i n_o} + \frac{\sigma_{pio}^2}{n_i n_o}$ ) that includes item and occasion interaction which may influence an absolute measure indirectly (Cardinet et al. 2010; Shavelson & Webb 1991):

$$G_a \approx \Phi = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2}$$

Both  $G_r$  and  $G_a$  are estimating reliability of a trait measure if the object of measurement is a person.  $G_r$  of 0.80 or higher is reflecting good reliability of assessment score (Cardinet et al. 2010), and while similar criteria are generally applied for  $G_a$ , coefficients above 0.70 were considered as reliable in some studies (Arterberry et al. 2014).

A state component index (SCI) and trait component index (TCI) were obtained, which reflect the proportion of variance attributed to a dynamic (state) and an enduring (trait) component in a measure. The formulae used were developed by Medvedev et al. (2017a):

$$SCI = \frac{\sigma_{po}^2}{\sigma_{po}^2 + \sigma_p^2}; TCI = \frac{\sigma_p^2}{\sigma_{po}^2 + \sigma_p^2}$$

SCI and TCI of 0.50 mean that an equal amount of variance is attributed to state and trait, and SCI above 0.60 (TCI < 0.40) would indicate that the majority of variance is reflecting a state. Conversely, TCI of 0.60 or higher (SCI < 0.40) would signify the majority of variance is reflecting a trait. These coefficients can be interpreted in a similar way to other reliability coefficients, where a higher score reflects a higher proportion of variance attributed to a state (SCI) or a trait (TCI) (Medvedev et al. 2017a).

In the D-study, variance components were obtained for each individual item and SCI values were calculated applying the formula described above. Therefore, items that show high SCI (i.e.,  $\geq 0.80$ ) are very sensitive to changes over time and can be considered as state items and items with lower SCI (i.e., < 0.30) as reflecting trait mindfulness (Medvedev et al. 2017a).

## Results

Descriptive statistics for the 39-item FFMQ, its subscales, and FFMQ-18 at three occasions are presented in Table 1. The internal consistency Cronbach’s alpha of the total FFMQ over three occasions ranged between 0.89 and 0.92. The test-retest reliability scores for Occasion 2 and Occasion 3 (with reference to Occasion 1) were 0.92 and 0.83, respectively, and were reflected by ICC of 0.83. These reliability values were overall higher than that of the FFMQ-18 and the individual subscales of the FFMQ. The mean scores of both FFMQ versions and individual subscales were not significantly different across occasions, as evidenced by paired *t* tests (all *p* values below 0.05). The subscales of *Nonjudge* and *Describe* obtained the highest Cronbach’s alpha and ICC values compared with other subscales. Overall, all assessed FFMQ scales and subscales showed acceptable internal consistency and temporal reliability expected for a trait measure. An exception was

the *Nonreact* subscale, which displayed the lowest Cronbach’s alpha value 0.69 at Occasion 1 and the lowest test-retest value at Occasion 3 (0.64).

## G-Study

Table 2 presents the variance components attributed to person (P), item (I), and occasion (O), and their interactions (P×I, P×O, I×O, P×I×O) together with generalizability coefficients and state and trait component indices for the FFMQ, its five subscales, and the FFMQ-18. The best reliability and generalizability of scores across persons and occasions was found for the total FFMQ with both relative and absolute *G* coefficients (*G<sub>r</sub>* and *G<sub>a</sub>*) of 0.89 and the main source of error variance due to P×O interaction that accounted for 98.2% of the total error. Slightly lower but still acceptable *G<sub>r</sub>* and *G<sub>a</sub>* values of 0.76 and 0.75, respectively, were observed for the FFMQ-18, with measurement error mainly explained by P×O and P×I×O interactions, which took up 79% of

**Table 1** Means, standard deviation (SD), Cronbach’s alpha, test-retest coefficients, and intraclass correlation coefficient (ICC) for the FFMQ total, its short version FFMQ-18 together with five facet subscales (*n* = 83 × 3 occasions)

Scale/assessment	Occasion 1	Occasion 2	Occasion 3	ICC(95% CI)
<b>Observe</b>				
Mean (SD)	25.54 (4.95)	25.40 (5.71)	26.17 (5.89)	
Cronbach’s alpha	0.74	0.81	0.83	
Test-retest ( <i>r</i> ) <sup>a</sup>	–	0.74	0.74	0.77(0.69–0.84)
<b>Describe</b>				
Mean (SD)	26.27 (6.16)	26.33 (6.61)	26.62 (7.05)	
Cronbach’s alpha	0.89	0.92	0.92	
Test-retest ( <i>r</i> ) <sup>a</sup>	–	0.90	0.83	0.86(0.81–0.90)
<b>Act with awareness</b>				
Mean (SD)	26.02 (4.77)	25.11 (6.17)	25.27(6.15)	
Cronbach’s alpha	0.81	0.90	0.90	
Test-retest ( <i>r</i> ) <sup>a</sup>	–	0.87	0.77	0.81(0.74–0.87)
<b>Nonjudge</b>				
Mean (SD)	26.34 (6.69)	26.74 (7.15)	27.10 (7.54)	
Cronbach’s alpha	0.90	0.94	0.95	
Test-retest ( <i>r</i> ) <sup>a</sup>	–	0.81	0.87	0.85(0.79–0.90)
<b>Nonreact</b>				
Mean (SD)	20.14 (3.83)	19.74 (4.37)	20.64 (4.07)	
Cronbach’s alpha	0.69	0.82	0.79	
Test-retest ( <i>r</i> ) <sup>a</sup>	–	0.75	0.64	0.71(0.61–0.79)
<b>FFMQ-18</b>				
Mean (SD)	57.68 (8.46)	57.37 (9.51)	57.71 (9.90)	
Cronbach’s alpha	0.79	0.84	0.86	
Test-retest ( <i>r</i> ) <sup>a</sup>	–	0.87	0.80	0.82(0.74–0.87)
<b>FFMQ total</b>				
Mean (SD)	124.25 (17.67)	123.38 (19.67)	125.04 (20.30)	
Cronbach’s alpha	0.89	0.91	0.92	
Test-retest ( <i>r</i> ) <sup>a</sup>	–	0.92	0.83	0.83(0.80–0.91)

Note: Mean differences are not significant compared with occasion 1 (Bonferroni corrected)

<sup>a</sup> Test-retest bivariate correlations between occasion 1 and subsequent occasions 2 and 3

**Table 2** G-study estimates for the FFMQ and FFMQ-18 and five subscales of the FFMQ including coefficient  $G$  relative ( $G_r$ ), coefficient  $G$  absolute ( $G_a$ ), trait component index (TCI), state component index (SCI), variance components (in %), and the person (P)  $\times$  occasion (O)  $\times$  item (I) design including interactions ( $n = 83$ )

Facets	FFMQ total		FFMQ-18		Observe		Describe		Act*		Nonjudge		Nonreact	
	$\sigma^2$	%	$\sigma^2$	%	$\sigma^2$	%	$\sigma^2$	%	$\sigma^2$	%	$\sigma^2$	%	$\sigma^2$	%
P	0.052		0.057		0.024		0.037		0.041		0.035		0.030	
I	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
O	0.000	1.8	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0	0.000	3.8
P $\times$ I	0.000	0.0	0.003	18.4	0.016	31.2	0.019	34.0	0.010	17.8	0.008	17.1	0.022	27.8
P $\times$ O	0.006	98.2	0.006	29.8	0.003	5.9	0.008	14.3	0.012	21.8	0.008	16.5	0.017	22.0
I $\times$ O	0.000	0.0	0	2.6	0.003	5.3	0.002	3.9	0.004	6.5	0.001	2.7	0.003	3.5
P $\times$ I $\times$ O	0.000	0.0	0.009	49.2	0.029	57.7	0.027	47.8	0.030	53.9	0.031	63.7	0.033	42.9
$G_r$	0.89		0.76		0.33		0.40		0.44		0.42		0.29	
$G_a$	0.89		0.75		0.32		0.39		0.43		0.42		0.28	
TCI	0.90		0.90		0.89		0.82		0.77		0.81		0.64	
SCI	0.10		0.10		0.11		0.18		0.23		0.19		0.36	

Numbers in italics signify acceptable reliability/generalizability coefficients

\*Act with awareness

error variance in combining. The TCI values reflecting the ability of an instrument to reliably assess a trait were calculated for both the FFMQ and FFMQ-18 (both TCI = 0.90). TCI values together with reliability estimates indicate that both the FFMQ and FFMQ-18 are consistent with expectations of a valid trait measure. In contrast,  $G_r$  and  $G_a$  for all individual subscales of the FFMQ were below 0.45 meaning that all subscales were not meeting expectations for a reliable trait measure (Shavelson et al. 1989). The SCI reflecting the ability of a measure to reliably assess state changes were below expectations for a valid state measure for all individual FFMQ subscales (all SCI < 0.40). Even though TCI value for all five FFMQ subscales were high, ranging from 0.64 (*Nonreact*) to 0.89 (*Observe*), all subscales were affected by measurement error due to interaction between person, item, and occasion. This resulted in low reliability of all subscales in measuring trait (all  $G_r$  < 0.50) meaning that the FFMQ subscales cannot be considered as measuring either state or trait mindfulness reliably.

### D-Study

Individual item analysis was conducted to obtain variance components for individual items by excluding all other items. The estimates for variance of person, occasion, and person-occasion interaction together with computed SCI are included in Table 3. There were nine items (i.e., 1, 2, 4, 12, 15, 18, 28, 30, and 38) which presented with high SCI ( $\geq 0.80$ ) reflecting high sensitivity for state changes over time. On the other end, there are nine items with low SCI ( $\leq 0.50$ ) that are least sensitive to state changes and reflecting predominantly trait mindfulness. All other items had SCI between these benchmarks ( $0.50 < SCI < 0.80$ ) and cannot be clearly classified as reflecting either state or trait.

Furthermore, a series of generalizability analyses were conducted by combining the most dynamic items with the highest SCI because we expected that this will result in a reliable state measure. Table 4 shows D-study results including reliability estimates and variance components attributed to person, item, and occasion and their interactions for these analyses. The first analysis was conducted with the five most dynamic items from each subscale including 1, 4, 12, 30, and 38 (Table 4, (a)). In the analyses b (Table 4), the first five items with the highest SCI selected from the total scale (1, 12, 15, 30, and 38) were combined, and subsequent analyses added the next most dynamic item from the remaining items (4, 18, and 28). The results showed that person-item-occasion interaction was the main source of error variance across all these analyses and ranged from 76.50 to 91.40% of the total error variance. As expected,  $G_r$  and  $G_a$  for all analyses of most dynamic items were below the acceptable generalizability for a trait measure (0.70). However, all SCI values for these analyses were lower than 0.19, which is far below expectations for a state measure (i.e., SCI

**Table 3** Variance components of person (P), occasion (O) and P×O interaction together with state component index (SCI) for each individual item of the FFMQ ( $n = 83 \times 3$ )

Items/facets	P	O	P×O	SCI
<b>Observe</b>				
15 I pay attention to sensations	0.04	0.00	0.32	0.88
31 I notice visual elements in art or nature	0.21	0.02	0.32	0.60
20 I pay attention to sounds	0.34	0.01	0.21	0.38
26 I notice the smells and aromas of things	0.26	0.02	0.31	0.55
6 I stay alert to the sensations of water	0.23	0.01	0.22	0.49
1 I notice the sensations of my body moving	0.04	0.03	0.34	0.89
11 I notice how foods and drinks affect thoughts	0.18	0.03	0.27	0.60
36 I notice how emotions affect thoughts and behaviour	0.08	0.06	0.25	0.75
<b>Act with awareness</b>				
38 <i>doing things without paying attention</i> <sup>R</sup>	0.01	0.00	0.30	0.95
13 I am easily distracted <sup>R</sup>	0.31	0.01	0.25	0.45
5 my mind wanders off and I'm easily distracted <sup>R</sup>	0.12	0.03	0.38	0.76
8 I do not pay attention to what I'm doing <sup>R</sup>	0.12	0.08	0.32	0.72
34 I do jobs or tasks automatically <sup>R</sup>	0.23	0.07	0.21	0.48
18 I find it difficult to stay focused <sup>R</sup>	0.06	0.02	0.38	0.86
28 I rush through activities without being attentive <sup>R</sup>	0.06	0.00	0.26	0.83
23 I am "running on automatic" <sup>R</sup>	0.16	0.05	0.26	0.62
<b>Nonjudge</b>				
25 I should not be thinking the way I'm thinking <sup>R</sup>	0.09	0.00	0.28	0.77
35 I judge myself as good or bad <sup>R</sup>	0.16	0.01	0.33	0.67
17 I make judgments about my thoughts <sup>R</sup>	0.16	0.00	0.34	0.67
30 I think my emotions are bad or inappropriate <sup>R</sup>	0.01	0.00	0.42	0.98
14 I believe my thoughts are abnormal or bad <sup>R</sup>	0.12	0.02	0.29	0.70
10 I should not be feeling the way I'm feeling <sup>R</sup>	0.12	0.03	0.25	0.68
39 I disapprove of myself <sup>R</sup>	0.13	0.00	0.29	0.69
3 I criticize myself for inappropriate emotions <sup>R</sup>	0.13	0.03	0.25	0.66
<b>Describe</b>				
37 I can usually describe how I feel at the moment	0.24	0.03	0.29	0.55
2 I'm good at finding words to describe my feelings	0.07	0.00	0.31	0.81
12 It's hard for me to find the words to describe <sup>R</sup>	0.04	0.02	0.32	0.89
16 I have trouble thinking of the right words <sup>R</sup>	0.23	0.03	0.21	0.48
7 I can easily put my thoughts into words <sup>R</sup>	0.36	0.00	0.19	0.35
27 when upset, I can find a way to put it into words	0.24	0.03	0.30	0.56
32 tendency is to put experiences into words	0.37	0.02	0.19	0.34
22 I cannot find the right words to describe sensation <sup>R</sup>	0.24	0.01	0.34	0.58
<b>Nonreact</b>				
33 I just notice distressing things and let them go	0.17	0.01	0.27	0.61
29 notice distressing things without reacting	0.12	0.01	0.27	0.69
24 I feel calm soon after distressing things	0.28	0.02	0.25	0.47
9 I watch my feelings without getting lost in them	0.22	0.02	0.35	0.62
19 I am aware of distressing thought or image	0.36	0.01	0.19	0.35
21 I can pause without immediately reacting	0.17	0.10	0.24	0.59
4 I perceive my emotions without reacting to them	0.11	0.00	0.44	0.80

should be above 0.60 to be considered as a state measure). These findings mean that none of the tested item combinations can be used reliably for the assessment of state mindfulness. Further analyses were conducted to test

whether removing items with higher SCI from each subscale will improve its reliability in measuring trait mindfulness. The items with the highest SCI were removed first one at a time and  $G$  coefficients of a relevant

**Table 4** D-study reliability estimates and variance components for the person (P) × occasion (O) × item (I) design including interactions for combine FFMQ items with the highest state component index (SCI)

Facets	5 state items(a)		5 state items(b)		6 state items		7 state items		8 state items	
	$\sigma^2$	%	$\sigma^2$	%	$\sigma^2$	%	$\sigma^2$	%	$\sigma^2$	%
P	0.11		0.09		0.03		0.05		0.03	
I	0.00	0.00	0.00	0.00	0.00	1.60	0.00	0.80	0.00	4.10
O	0.01	10.10	0.00	0.00	0.00	7.50	0.00	6.40	0.00	5.40
PI	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.10	0.00	1.80
PO	0.00	0.00	0.00	0.00	0.01	11.80	0.01	11.20	0.00	9.40
IO	0.00	1.30	0.01	8.60	0.00	2.00	0.00	2.00	0.00	1.80
PIO	0.07	88.60	0.06	91.40	0.05	76.90	0.04	79.50	0.03	77.50
GM	3.25		4.14		3.20		3.20		3.13	
SE	0.10		0.09		0.09		0.08		0.08	
Gr	0.61		0.59		0.37		0.48		0.47	
Ga	0.58		0.57		0.35		0.45		0.44	
TCI	1.00		1.00		0.82		0.88		0.89	
SCI	0.00		0.00		0.18		0.12		0.11	

GM grand mean, SE standard error of the grand mean

subscale were examined. However, no improvement of reliability was achieved for any of the FFMQ facets (all  $G_r < 0.60$ ).

## Discussion

The aim of this study was to distinguish between state and trait components in the FFMQ and to examine temporal reliability and generalizability of this scale using G-Theory. The results show that the total 39-item FFMQ and the FFMQ-18 are reliable in measuring trait mindfulness with  $G$  coefficients of 0.89 and 0.75, respectively, meaning that their scores are generalizable across persons and occasions. However, all five individual subscales of the FFMQ were found to measure trait mindfulness with TCI above 0.60 (SCI below 0.40) but they appear less reliable ( $G$  coefficients below 0.45) compared with the total FFMQ and FFMQ-18. Our results indicated that individual subscale scores were affected by measurement error due to interactions between person, item, and occasion, which presented the highest percentage of the error variance ranging from 43 to 64% across subscales. Individual subscales were also affected by interaction error between person and item that was specifically evident in the subscales *Describe* (34%), *Observe* (31.2%), and *Nonreact* (27.8%). In contrast, the FFMQ total scores contained a state component of person and occasion interaction that constitute 98% of the total error variance, but its influence on the overall reliability of measurement was negligible with  $G \geq 0.80$  (Shavelson et al. 1989).

A D-study was conducted in an attempt to develop a subscale to measure mindfulness as a state by combining the

FFMQ items identified as the most dynamic over time, which did not result in a sensitive state measured as reflected by low SCI. It is possible that dynamic changes in specific aspects of mindfulness are not occurring simultaneously and cancel each other out if different state items are combined. For example, item 38 (“doing things without paying attention”) and item 30 (“I think my emotions are bad or inappropriate”) had SCI at 0.95 (TCI = 0.05) and 0.98 (TCI = 0.02), respectively, which indicates they are measure a state aspects of mindfulness to the large extent. However, combining these items may counter balance state changes on each aspect over time because they are less likely to occur at the same time. This notion is supported by our results in Table 4 where we attempted to combine state items resulting in lower SCI. These findings are consistent with psychometric studies that demonstrated reduction of measurement error due to individual items by combining them into super-items or parcels (Medvedev et al. 2018; Taylor et al. 2017).

We note that each of the FFMQ subscales except for *Nonjudge* included both state and trait items. Although, all *Nonjudge* subscale items were sensitive to change overtime but the overall subscale sensitivity was low (SCI = 0.19; TCI = 0.81) meaning that this subscale is not reflecting state changes. This could be explained by the fact that different aspects of non-judgmental attitude captured by individual items (e.g., self, emotions, thoughts) may not co-occur together in time. Therefore, combining *Nonjudge* items together may reduce the overall subscale sensitivity to change because state related variances may cancel each other out (Medvedev et al. 2018; Taylor et al. 2017). However, these findings indicate that various aspects of non-judgmental attitude are very



dynamic and should be the primary focus of any MBIs because they are more amendable and were consistently found as a strong predictor of psychological symptoms (Baer et al. 2008; Medvedev et al. 2018).

In the *Observe* subscale, there were only three items (“I pay attention to sensations”, “I notice the sensations of my body moving,” and “I notice how emotions affect thoughts and behaviour”) that clearly indicated measuring state due to their high SCI and low TCI (0.89, 0.88, and 0.75, respectively; TCI of 0.11, 0.12, and 0.25, respectively). If considering to develop mindful observing, then focusing on emotions, sensations, and thoughts in the first place may be helpful as these are the most amendable features. The results also show that “I pay attention to sounds”, “I notice the smells and aromas of things,” and “I stay alert to the sensations of water” obtained lower SCI, which are more stable trait-like aspects of a person.

The *Describe* subscale shows psychometric patterns comparable to those of the *Observe* subscale. Only two items (“I’m good at finding words to describe my feelings” and “It’s hard for me to find the words to describe”) in the *Describe* subscale clearly displayed high sensitivity to change (state) with SCI of 0.81 and 0.89 (TCI of 0.19 and 0.11), respectively. The remaining items in this facet reflected predominantly enduring patterns. Although *Describe* had a higher number of trait-like items than items reflecting a state, this facet can still not be regarded as a reliable trait-like mindfulness measure according to our results ( $G_r = 0.40$ ). This may be explained by the fact that individual items measuring the ability to describe mindfulness related to unobservable behaviors such as feelings, sensations, and thoughts change over time, which is reflected in the high measurement error due to interactions between person, item, and occasion.

In the *Nonreact* subscale, there were four items with  $SCI > 0.60$  that indicated high sensitivity to change, with the most sensitive item “I perceive my emotions without reacting to them” ( $SCI = 0.80$ ;  $TCI = 0.20$ ). The remaining three items in this subscale can be psychometrically quantified as measuring a person’s trait. Although the *Nonreact* subscale included items sensitive to change over time, the overall SCI was low ( $0.36$ ;  $TCI = 0.64$ ), meaning that this subscale did not reflect dynamic aspects of mindfulness reliably when these items were combined together. Similar to the other subscales of the FFMQ, *Nonreact* was affected by measurement error due to interactions between person, item, and occasion. This indicates that people may respond to the same item differently at different occasions because individual thoughts and feelings varying over time.

There was an obvious imbalance between items reflecting state and trait mindfulness in *Act with Awareness* facet. There were only two items, “I am easily distracted” ( $SCI = 0.45$ ;  $TCI = 0.55$ ) and “I do jobs or tasks automatically” ( $SCI = 0.48$ ;  $TCI = 0.52$ ), that were less sensitive to changes over occasions. The remaining six out of eight items of this subscale reflected state aspects of mindfulness, with three items

showing high SCIs ranging from 0.83 up to 0.95 (TCIs ranging from 0.05 to 0.17). However, combining these items did not result in a sensitive state measure.

## Limitations and Future Research

Some limitations need to be acknowledged. The current study was conducted with participants who were all university students, which has a degree of homogeneity and large population of females, and the results should be replicated in more diverse samples. The gender imbalance may influence the results and it would be beneficial for future studies to replicate this analysis with a more balanced sample and analyze different genders separately. The FFMQ-18 was analyzed using data from the full scale which is a potential limitation because responding to items presented in a different order may influence the results. Although the FFMQ contains 19 reverse scored items designed to reduce response bias, they may potentially affect reliability of the scale meaning that obtained  $G$  coefficients could be higher if there would be no reverse scored items.

In the current study, we found that there were 25 items (i.e., 1, 2, 3, 4, 5, 8, 9, 11, 12, 14, 15, 17, 18, 23, 25, 28, 29, 30, 31, 33, 35, 36, 38, and 39) with high  $SCI (\geq 0.60)$  reflecting high sensitivity for state changes over time. On the other hand, the remaining fourteen items had SCI between the benchmarks ( $0.30 < SCI < 0.60$ ) and cannot be clearly classified as reflecting either state or trait because they are measuring both aspects. It means that there are no items with low  $SCI (\leq 0.30)$  that are least sensitive to state changes and are reflecting predominantly trait mindfulness. These findings should be replicated in future research using different samples to confirm replicability of this result.

In conclusion, the findings of this study indicate that reliable measurement of trait mindfulness can be achieved by using the full FFMQ scale or its short version FFMQ-18 with scores generalizable across sample population and occasions. The scores obtained on individual facet subscales of the FFMQ predominantly measuring trait mindfulness, but their reliability is affected by measurement error due to interaction between person, item, and occasion. Robust psychometric properties of the FFMQ full scale and the FFMQ-18 permit assessment of trait mindfulness reflecting long-lasting effects of MBIs and evaluation of their long-term effectiveness. State items identified in this study are reflecting dynamic components of mindfulness that are the most amendable and should be the primary target of MBIs.

**Author Contributions** QCT: designed and conducted the study, analyzed the data, and wrote the paper. CUK: collaborated with developing the study and writing the manuscript. RJS: collaborated with developing the study, collecting the data and editing the manuscript. JL: collaborated with collecting the data and writing the manuscript. ONM: collaborated with designing and conducting the study, analyzing the data, and writing the paper.

**Funding Information** The data used in this study were from the doctoral work of the last author funded by the Vice-Chancellor's Scholarship of Auckland University of Technology.

## Compliance with Ethical Standards

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

**Informed Consent** All participants involved in this study provided their informed consent.

**Ethics Statement** The study complied with the guidelines of the Auckland University of Technology ethics committee, which is based on internationally accepted ethical standards.

## References

- Allen, M. J., & Yen, W. M. (1979). *Introduction on to measurement theory*. Monterey, CA: Brooks/Cole.
- Arterberry, B. J., Martens, M. P., Cadigan, J. M., & Rohrer, D. (2014). Application of generalizability theory to the big five inventory. *Personality and Individual Differences, 69*, 98–103.
- Baer, R. A., Carmody, J., & Hunsinger, M. (2012). Weekly change in mindfulness and perceived stress in a mindfulness-based stress reduction program. *Journal of Clinical Psychology, 68*(7), 755–765. <https://doi.org/10.1002/jclp.21865>.
- 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., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., & Walsh, E. (2008). Construct validity of the Five Facet Mindfulness Questionnaire in meditating and nonmeditating samples. *Assessment, 15*(3), 329–342.
- Birrer, D., Röthlin, P., & Morgan, G. (2012). Mindfulness to enhance athletic performance: Theoretical considerations and possible impact mechanisms. *Mindfulness, 3*(3), 235–246. <https://doi.org/10.1007/s12671-012-0161-y>.
- Bishop, S. R., Lau, M. A., Shapiro, S., Carlson, L. E., Anderson, N. D., Carmody, J., & Devins, G. (2006). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice, 11*, 230–241.
- Bloch, R., & Norman, G. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE guide no. 68. *Medical Teacher, 34*, 960–992.
- Bohlmeijer, E., Prenger, R., Taal, E., & Cuijpers, P. (2010). The effects of mindfulness-based stress reduction therapy on mental health of adults with a chronic medical disease: A meta-analysis. *Journal of Psychosomatic Research, 68*(6), 539–544. <https://doi.org/10.1016/j.jpsychores.2009.10.005>.
- 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*(3), 308–320. <https://doi.org/10.1177/1073191111408231>.
- Brown, D. B., Bravo, A. J., Roos, C. R., & Pearson, M. R. (2015). Five facets of mindfulness and psychological health: Evaluating a psychological model of the mechanisms of mindfulness. *Mindfulness, 6*(5), 1021–1032. <https://doi.org/10.1007/s12671-014-0349-4>.
- Bush, M. (2011). Mindfulness in higher education. *Contemporary Buddhism, 12*(1), 183–197. <https://doi.org/10.1080/010183-197>.
- Cardinet, J., Pini, G., & Johnson, S. (2011). *Applying generalizability theory using EduG*. London: Routledge Academic.
- Chang, V. Y., Palesh, O., Caldwell, R., Glasgow, N., Abramson, M., Lusk, F., et al. (2004). The effects of a mindfulness-based stress reduction program on stress, mindfulness self-efficacy, and positive states of mind. *Stress and Health, 20*(3), 141–147. <https://doi.org/10.1002/smi.1011>.
- Chiesa, A., & Serretti, A. (2009). Mindfulness-based stress reduction for stress management in healthy people: A review and meta-analysis. *Journal of Alternative and Complementary Medicine, 15*(5), 593–600. <https://doi.org/10.1089/acm.2008.0495>.
- Coffey, K. A., Hartman, M., & Fredrickson, B. L. (2010). Deconstructing mindfulness and constructing mental health: Understanding mindfulness and its mechanisms of action. *Mindfulness, 1*(4), 235–253. <https://doi.org/10.1007/s12671-010-0033-2>.
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory†. *British Journal of Statistical Psychology, 16*(2), 137–163. <https://doi.org/10.1111/j.2044-8317.1963.tb00206.x>.
- Davis, K. M., Lau, M. A., & Cairns, D. R. (2009). Development and preliminary validation of a trait version of the Toronto Mindfulness Scale. *Journal of Cognitive Psychotherapy, 23*(3), 185.
- Frewen, P. A., Unholzer, F., Logie-Hagan, K. R.-J., & MacKinley, J. D. (2014). Meditation breath attention scores (MBAS): Test-retest reliability and sensitivity to repeated practice. *Mindfulness, 5*(2), 161–169. <https://doi.org/10.1007/s12671-012-0161-y>.
- Gu, J., Strauss, C., Crane, C., Barnhofer, T., Karl, A., Cavanagh, K., & Kuyken, W. (2016). Examining the factor structure of the 39-item and 15-item versions of the Five Facet Mindfulness Questionnaire before and after mindfulness-based cognitive therapy for people with recurrent depression. *Psychological Assessment, 28*(7), 791–802. <https://doi.org/10.1037/pas0000263>.
- Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality and Quantity, 34*(4), 331–351.
- Hwang, Y., Goldstein, H., Medvedev, O. N., Singh, N. N., Noh, J., & Hand, K. (2019). Mindfulness-based intervention for Educators: Effects of a School-Based Cluster Randomized Controlled Study. *Mindfulness, 10*(7), 1417–1436. <https://doi.org/10.1007/s12671-019-01147-1>.
- Hyland, P. K., Lee, R. A., & Mills, M. J. (2015). Mindfulness at work: A new approach to improving individual and organizational performance. *Industrial and Organizational Psychology, 8*(4), 576–602. <https://doi.org/10.1017/iop.2015.41>.
- Kabat-Zinn, J. (1982). An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry, 4*(1), 33–47. [https://doi.org/10.1016/0163-8343\(82\)90026-3](https://doi.org/10.1016/0163-8343(82)90026-3).
- Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness*. New York, NY: Delacourt.
- Kabat-Zinn, J., Wheeler, E., Light, T., Skillings, A., Scharf, M. J., Cropley, T. G., Hosmer, D., & Bernhard, J. D. (1998). Influence of a mindfulness meditation-based stress reduction intervention on rates of skin clearing in patients with moderate to severe psoriasis undergoing phototherapy (UVB) and photochemotherapy (PUVA). *Psychosomatic Medicine, 60*(5), 625–632. <https://doi.org/10.1097/00006842-199809000-00020>.
- Krägeloh, C. U., Henning, M. A., Medvedev, O., Feng, X. J., Moir, F., Billington, R., & Siegert, R. J. (2019). *Mindfulness-based intervention research characteristics approaches and developments*. London and New York: Routledge.
- Lau, M. A., Bishop, S. R., Segal, Z. V., Buis, T., Anderson, N. D., Carlson, L., et al. (2006). The Toronto Mindfulness Scale: Development and validation. *Journal of Clinical Psychology, 62*(12), 1445–1467.

- Ledesma, D., & Kumano, H. (2009). Mindfulness-based stress reduction and cancer: A meta-analysis. *Psycho-Oncology*, 18(6), 571–579. <https://doi.org/10.1002/pon.1400>.
- Levinson, D. B., Stoll, E. L., Kindy, S. D., Merry, H. L., & Davidson, R. J. (2014). A mind you can count on: Validating breath counting as a behavioral measure of mindfulness. *Frontiers in Psychology*, 5, 1202. <https://doi.org/10.3389/fpsyg.2014.01202>.
- Libet, B. (2004). *Mind time: The temporal factor in consciousness*. Cambridge, MA: Harvard University Press.
- Ma, S. H., & Teasdale, J. D. (2004). Mindfulness-based cognitive therapy for depression: Replication and exploration of differential relapse prevention effects. *Journal of Consulting and Clinical Psychology*, 72(1), 31–40. <https://doi.org/10.1037/0022-006X.72.1.31>.
- MacDonald, H. Z., & Baxter, E. E. (2017). Mediators of the relationship between dispositional mindfulness and psychological well-being in female college students. *Mindfulness*, 8(2), 398–407. <https://doi.org/10.1007/s12671-016-0611-z>.
- Massion, A. O., Teas, J., Herbert, J. R., Wetheimer, M. D., & Kabat-Zinn, J. (1995). Meditation, melatonin and breast/prostate cancer: Hypothesis and preliminary data. *Medical Hypotheses*, 44(1), 39–46. [https://doi.org/10.1016/0306-9877\(95\)90299-6](https://doi.org/10.1016/0306-9877(95)90299-6).
- Medvedev, O. N., Krägeloh, C. U., Narayanan, A., & Siegert, R. J. (2017a). Measuring mindfulness: Applying Generalizability Theory to distinguish between state and trait. *Mindfulness*, 8(4), 1036–1046. <https://doi.org/10.1007/s12671-017-0679-0>.
- Medvedev, O. N., Norden, P. A., Krägeloh, C. U., & Siegert, R. J. (2018). Investigating unique contributions of dispositional mindfulness facets to depression, anxiety, and stress in general and student populations. *Mindfulness*, 9(6), 1757–1767. <https://doi.org/10.1007/s12671-018-0917-0>.
- Medvedev, O. N., Siegert, R. J., Kersten, P., & Krägeloh, C. U. (2017b). Improving the precision of the Five Facet Mindfulness Questionnaire using a Rasch approach. *Mindfulness*, 8(4), 995–1008. <https://doi.org/10.1007/s12671-016-0676-8>.
- Paterson, J., Medvedev, O. N., Sumich, A., Tautolo, E., Krägeloh, C. U., Sisk, R., et al. (2017). Distinguishing transient versus stable aspects of depression in New Zealand Pacific Island children using Generalizability Theory. *Journal of Affective Disorders*, 227, 698–704. <https://doi.org/10.1016/j.jad.2017.11.075>.
- Ramanaiah, N. V., Franzen, M., & Schill, T. (1983). A psychometric study of the State-Trait Anxiety inventory. *Personality Assessment*, 47, 531–535.
- Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. New York: Guilford Press.
- Shapiro, S. L., Astin, J. A., Bishop, S. R., & Cordova, M. (2005). Mindfulness-based stress reduction for health care professionals: Results from a randomized trial. *International Journal of Stress Management*, 12(2), 164–176. <https://doi.org/10.1037/1072-5245.12.2.164>.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer* (Vol. 1). Thousand Oaks: Sage Publications, Inc.
- Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist*, 44, 599–612.
- Shoukri, M. M., Asyali, M. H., & Donner, A. (2004). Sample size requirements for the design of reliability study: Review and new results. *Statistical Methods in Medical Research*, 13, 251–271.
- Spielberger, C. D. (1999). *Manual for the state-trait anger expression inventory-2*. Odessa, FL: Psychological Assessment Resources.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Test manual for the state trait anxiety inventory*. Palo Alto: Consulting Psychologists Press.
- Stucki, G., Daltroy, L., Katz, J. N., Johannesson, M., & Liang, M. H. (1996). Interpretation of change scores in ordinal clinical scales and health status measures: The whole may not equal the sum of the parts. *Journal of Clinical Epidemiology*, 49(7), 711–717. [https://doi.org/10.1016/0895-4356\(96\)00016-9](https://doi.org/10.1016/0895-4356(96)00016-9).
- Suen, H. K., & Lei, P. W. (2007). Classical versus generalizability theory of measurement. *Educational Measurement*, 4, 1–13.
- Swiss Society for Research in Education Working Group. (2006). *EDUG User Guide*. Neuchatel: IRDP.
- Tanay, G., & Bernstein, A. (2013). State Mindfulness Scale (SMS): development and initial validation. *Psychological Assessment*, 25(4), 1286–1299.
- Tang, Y., Hölzel, B. K., & Posner, M. I. (2015). Traits and states in mindfulness meditation. *Nature Reviews Neuroscience*, 17(1), 59–59. <https://doi.org/10.1038/nrn.2015.7>.
- Taylor, T. A., Medvedev, O. N., Owens, R. G., & Siegert, R. J. (2017). Development and validation of the State Contentment Measure. *Personality and Individual Differences*, 119, 152–159. <https://doi.org/10.1016/j.paid.2017.07.010>.
- Vispoel, W. P., Morris, C. A., & Kilinc, M. (2018). Applications of generalizability theory and their relations to classical test theory and structural equation modeling. *Psychological Methods*, 23(1), 1–26. <https://doi.org/10.1037/met0000107>.

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