



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

Applying Generalisability Theory to Differentiate between Trait and State in the Five Facet Mindfulness Questionnaire (FFMQ)

A thesis
submitted in fulfilment
of the requirements for the degree
of
Master of Social Science in Psychology
at
The University of Waikato
by
QUOC CUONG TRUONG



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Year of submission

2020

Abstract

Mindfulness has been applied as an effective therapy in clinical settings to treat psychological symptoms (e.g., depression, stress, anxiety) and also in many other contexts (e.g., the workplace, educational contexts, and sports). Accurate evaluation of mindfulness-based training or mindfulness research requires distinguishing between state and trait changes. The failure to distinguish trait from state in a mindfulness measurement may confuse the assessment results of neurophysiological and psychological mindfulness studies, as well as mindfulness-based treatments, because mindfulness can be understood as either a state or a trait. The Five Facet Mindfulness Questionnaire (FFMQ) is the most widely used multidimensional assessment tool of dispositional mindfulness, comprised of five subscales: “act with awareness”, “describe”, “nonjudge”, “nonreact”, and “observe”. Its short version, consisting of 18 items (FFMQ-18), was proposed by examining and comparing the existing short versions of the FFMQ. However, the ability of the FFMQ to accurately capture stable aspects of mindfulness has not been studied thoroughly, using robust methodology. The generalisability theory (G-theory) is the most suitable and robust statistical approach to differentiate and measure these aspects.

In this study, G-theory was applied to distinguish between dynamic (state) and enduring (trait) aspects of mindfulness and to evaluate the reliability of the FFMQ and the FFMQ-18 over time. The study used data from 83 participants, who completed the test on three occasions separated by 2-week intervals. The full 39-item FFMQ and the shorter FFMQ-18 version proved to be very reliable in measuring trait mindfulness, with G coefficients of 0.89 and 0.75, respectively, while individual facet subscales of the FFMQ appeared to be less reliable in measuring trait mindfulness. Subsequent analyses attempted to combine the FFMQ items that were least stable over time to form a measure of state mindfulness. However, these did not result in acceptable psychometric properties for such a state subscale. The findings of this study indicate that a reliable assessment of stable aspects of mindfulness can be achieved by using the full FFMQ scale or its short FFMQ-18 version, with scores that can be generalised across the sample population and assessment occasions. The scores obtained on individual facet subscales of the FFMQ predominantly measure trait mindfulness but their reliability is affected by measurement error, resulting from interaction between persons, items and occasions.

Keywords: Mindfulness, State and Trait, Assessment Tool, Five Facet Mindfulness Questionnaire, Generalisability Theory

Acknowledgements

I would like to give special thanks to Dr. Oleg Medvedev for his supervision of this project, and for providing data from his doctoral study not used or published previously.

I am also grateful to all those who participated in this study.

I express my heartfelt thanks to my wife Loan, who tirelessly supported me in completing this project. I also appreciate her work in caring for me and my family while I was working on this thesis. I thank my sons Bach and Khoa for their empathy and tolerance as well as for being a motivation for me to complete this work. Many thanks to my parents, who always supported me in my education and encouraged me to advance to this level.

Co-Authored Works

The study presented in my thesis is co-authored work with Christian U. Krägeloh, Richard J. Siegert, Jason Landon and Oleg N. Medvedev. This study was recently published as a journal article in *Mindfulness* journal (Impact Factor =3+) (see Appendix A1). I am the first author and have the major contribution to this study. However, this study was only successfully published with the advice and support of my supervisor- Doctor Oleg Medvedev and other collaborators. I have a major contribution to this project that includes designing and conducting the study, completing data analysis under guidance of the supervisor, writing the manuscript for submission to *Mindfulness* journal and responding to reviews/addressing their comments during peer-review process.

Table of Contents

Abstract	ii
Acknowledgements	iv
Co-Authored Works	v
Table of Contents	vi
List of Tables.....	viii
Chapter 1 Mindfulness, Health and Well-being	
Defining Mindfulness.....	1
Mindfulness-based Interventions	4
Benefits of Mindfulness for Health and Well-being	8
Conclusion.....	17
Chapter 2 Measuring Mindfulness	
Overview of Mindfulness Assessment Tools.....	19
Assessing Mindfulness in Neurophysiological Research.....	20
Psychometric Mindfulness Measures and Their Properties	26
The Five Facet Mindfulness Questionnaire.....	33
State and Trait Distinction in the Measurement of Mindfulness.....	36
Conclusion.....	38
Chapter 3 Measurement Theories: G-theory and CTT	
Introduction	39
Classical Test Theory	39
Generalisability Theory.....	41
Application of G-theory to the State versus Trait Distinction.....	46

Summary and Aim of the Present Work	53
Chapter 4 Generalisability Study Method and Results	
The Purpose of the Study	54
Participants	54
Procedure.....	55
Measuring.....	55
Data Analysis	55
Results	56
Chapter 5: Discussion	
Main Findings of G-Study.....	65
Contrast of G-Study and CTT Findings	67
D-Study Findings	68
Five Facets of Mindfulness	69
Implications of G-study Findings.....	72
Implications of D-study Findings.....	73
Limitations	74
Directions for Further Research	76
Conclusion.....	76
References	78
Appendix A1	101
Appendix A2	112
Appendix B	113
Appendix C1-51	117

List of Tables

Table	Title	Page
1	Psychometric properties of common mindfulness assessment tools and the number of citations of each scale in Google Scholar and Web of Science (22 July, 2019).	27
2	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 x 3 occasions).	58
3	G-study estimates for the FFMQ and FFMQ-18 and five subscales of FFMQ standard errors (SE), coefficient G relative (Gr), coefficient G absolute (Ga), Trait Component Index (TCI), State Component Index (SCI), Grand Mean (GM), and variance components (σ^2) in percent, for the Person (P) \times Occasion (O) \times Item (I) design, including interactions (n = 83).	60
4	Variance components of Person (P), Occasion (O) and P x O interaction together with the State Component Index (SCI) for each individual item of the FFMQ (n = 83x3).	63
5	D-study reliability estimates and variance components for the Person (P) \times Occasion (O) \times Item (I) design, including interactions for combining the FFMQ items with the highest State Component Index (SCI). One item selected per subscale (a), and items selected from the total scale (b).	64

Chapter 1 Mindfulness, Health and Well-being

Defining Mindfulness

There has been an enormous increase in mindfulness research over the last three decades, since mindfulness has proved to be a critical component in clinical interventions to treat a variety of mental health problems (Chiesa & Serretti, 2010; Goldin & Gross, 2010; Zoogman, Goldberg, Hoyt, & Miller, 2014; Thornton, Williamson, & Cooke, 2017) as well as an effective way to enhance well-being (Keng, Smoski, & Robin, 2011; Thornton et al.). According to Williams, Leumann and Cappeller (2004), mindfulness originally derives from traditional Buddhism. The term “mindfulness” originates from the Sanskrit word “स्मृति” (Smṛti) meaning “that which is remembered”. Mindfulness can be understood as focusing the attention on the experience of the present moment (Nyanaponika, 1973; Siegel, Germer, & Olendzki, 2009; Shapiro & Carlson, 2009; Black, 2011).

Brown and Ryan (2003) define mindfulness as “attention to and awareness of whatever is occurring in the present” (p. 824). Segal, Williams, and Teasdale (2013) understand mindfulness as “the awareness that emerges through paying attention on purpose in the present moment and non-judgmentally to things as they are” (p. 132). Kabat-Zinn (1994) proposed the most commonly cited definition of mindfulness – “paying attention in a particular way; on purpose, in the present moment, and nonjudgmentally” (p. 4). This definition has been used to complement many approaches in health psychology as well as in other relevant contexts (Bishop et al., 2006; Hayes, Strosahl, & Wilson, 1999; Mace, 2008; Siegel et al., 2009). Even though additional factors, such as the ability to “describe experiences”, “acting with awareness” and “non-reacting to inner experience”, were included in some assessments of mindfulness, these were derived from the originally proposed definition (Baer, Smith, & Allen, 2004; Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006). For example, the concept of “paying attention in a particular way” (Kabat-Zinn, 1994) may be interpreted as including kindness, acceptance, friendliness, curiosity and allowing (Segal et al., 2013). Suggesting an alternative understanding, Lau et al. (2006) described mindfulness as a state-like quality which has two components:

“(a) the intentional self-regulation of attention to facilitate greater awareness of bodily sensations, thoughts, and emotions; and (b) a specific quality of attention characterized by endeavouring to connect with each object in one’s awareness (e.g., each bodily sensation, thought, or emotion) with curiosity, acceptance, and openness to experience. Such a state involves an active process of relating openly with one’s current experience by allowing current thoughts, feelings, and sensations.” (p. 1447)

The ability to focus one’s attention on the present stands as the most essential component of mindfulness. Attention is the basic component of consciousness and is usually defined as focusing on one or many stimuli which are specific, sufficiently intense, and ignoring other stimuli (Brown, Ryan, & Creswell, 2007). In mindfulness, attention is focused on the experience of the current moment, while disregarding distracting events from the past or future, even with such events come from cognitive processes (Bishop et al., 2004). The famous experiment carried out by Simons and Chabris (1999) on focused attention can clarify this point. In their study, participants were unlikely to recall whether a person wearing a gorilla suit crossed the court during a basket game ball but successfully counted the number of ball passes in the game they were watching. This is an example of mindlessness and generalising from it helps to understand how a mindful person perceives. In this situation, if the scene remains the same and participants are asked to observe what is taking place in the present moment without focusing on any particular stimuli, they will give their attention to all events in their surroundings and are more likely to recall a person dressed as a gorilla. By contrast, a mindful person will extend their awareness to the full field of perception, without focusing on anything specific.

It is undisputable that ordinary perception is affected through the way sensory information is experienced, involving the interaction of emotions, cognition and behaviour (Whittle, 2013; Cohen, 2017). Specifically, emotional and behavioural responses are often influenced or conditioned by earlier experiences and can easily embed a perception into a mental schema (Beck & Alford, 2009). Although the benefits of ordinary perception cannot be denied, such perception tends to be associated with the processes of automatic labelling

and conceptualising, as well as judging (Bargh & Chartrand, 1999). This perception is also influenced by an individual's mental schemas, beliefs and ideas (Beck & Alford; Leary, 2004; Leary, Adams, & Tate, 2006).

By contrast, attention in mindfulness is a receptive attitude associated with focusing on actual "perceptual experiences in the present moment" (Kabat-Zinn, 1994). Mindfulness requires the ability to focus on experiences from external and internal sources, without employing the current cognitive schemas typically involved in generating an individual's conceptual world and entities (Olendzki, 2005). This means that contact with reality is instant (i.e., in the present moment) and not contaminated or influenced by individual habits and conceptual cognition, thus allowing individuals to reduce perceptual bias (Brown et al., 2007). Current mindfulness techniques have much in common, despite differences in practices and training methods (Germer, Siegel, & Fulton, 2005; Kabat-Zinn, 2003; Marlatt & Kristeller, 1999; Olendzki).

Mindfulness meditation based on the definitions and theoretical view described above require a practitioner to pay attention to all external and internal stimuli, while maintaining a non-judgmental attitude, without the distractions of "aversion, attraction and grasping" (Kabat-Zinn, 2003; Olendzki, 2005). Basically, this means that mindfulness helps an individual to recognise habits, which tend to be learned reactions to daily events both emotionally and physiologically, then step away from them. Practising mindfulness allows individuals to be entirely present in their life and work and reduces distractions associated with daydreaming or thinking about events from the past or future, thus improving their quality of life. Daydreaming is understood as a persistent, recurrent fantasizing activity, which may cause dysfunction and affect adaptive cognition (Somer, Somer, & Jopp, 2016).

Mindfulness is one of two general approaches used to distinguish meditation practices methodologically (Bhikkhu, 1997; Ivanovski & Malhi, 2007). The other is concentration-based meditation, which requires practitioners to pay attention voluntarily focussing on a single sensory or mental stimulus, such as a sound, the sensation of breathing, a vision or a thought (e.g., lovingkindness) (Baer, 2003; Ivanovski & Malhi). For

instance, practitioners practising Shamatha meditation is performed by paying attention to a single point, such as the physical sensation of breathing (Marlatt & Kristeller, 1999; Wallace, 1999). This form of meditation is occasionally known as the mindfulness of breathing.

Hence, a number of meditation practices cannot be identified as neither mindfulness nor concentration-based meditation because one approach may have some aspects of the other. However, since mindfulness extends “awareness from a single point to the full perceptual field”, mindfulness practitioners can be conceptually distinguished from those who practise concentration-based approaches to meditation (Baer, 2003; Ivanovski & Malhi, 2007). For example, the Dalai Lama, Baron, and Gaffiney (2004) suggest a mindfulness exercise where practitioners remain relaxed, sitting erect and watching the space in front of them. They are also required not to focus on anything specific. The single object of focus here is the space one metre away from the eyes. This method starts with concentration on empty space, which is not concentration *per se* because there is no stimulus to focus on. It helps to expand awareness to all objects in one’s perceptual field without focusing on anything specific.

Mindfulness-based Interventions

Growing evidence has demonstrated the positive effect of the application of mindfulness interventions. Research methods and applicable apparatus have steadily developed over the past 30-year period (Krägeloh et al., 2019). Mindfulness has been used to develop a structured programme to treat chronic pain and psychological symptoms, such as stress and anxiety (Kabat-Zinn, 1982). Early studies evaluated the effectiveness of mindfulness-based therapeutic methods through the changes observable in specific hypothesized outcomes (Massion, Teas, Herbert, Wetheimer, & Kabat-Zinn, 1995; Kabat-Zinn et al., 1998). Later studies applied several types of mindfulness practice, resulting in mindfulness-based interventions (MBIs).

Developed by Kabat-Zinn (1982), Mindfulness-Based Stress Reduction (MBSR) is the first mindfulness-based therapy to be widely practised. This is an 8-week programme, including one 2-hour session weekly and one meditation retreat lasting for a full day

(Kabat-Zinn, 1982, 1990). Participants in this programme are typically given instructions on meditation and practice 6 days a week, with a minimum duration of 45 minutes a day. They are also encouraged to practise mindfulness in their daily activities, such as standing, walking, sitting, and eating, during as well as after completion of the programme (Kabat-Zinn, 1982; Will et al., 2015). There are three main formal techniques: mindfulness meditation (e.g., traditional sitting meditation), body scanning (mindful body perception), and simple yoga postures (e.g., Hatha yoga postures). MBSR is based on a feature common to all mindfulness-based treatments, which focuses the attention of meditators on the “now” moment (Germer et al., 2005). The observation of all emotions, sensations, and thoughts during meditation is accomplished with a non-judgmental attitude toward their content. Participants who practise this MBI consequently realise that their thoughts, sensations and emotions appear and disappear naturally (Linehan, 1993b).

Mindfulness-Based Cognitive Therapy (MBCT; Segal, Williams, & Teasdale, 2002) is another MBI which uses mindfulness meditation as an integrated component in its treatment (Germer et al., 2005). The MBCT programme follows a similar format to that of the MBSR (Felder, Dimidjian, & Segal, 2012; Krägeloh et al., 2019). According to Segal et al. (2002), participants are taught the core exercise of 3-minute breathing space meditation, which includes three essential components. The first is “awareness”, which refers to directing one’s attention to the present moment by following a traditional meditation body posture and asking about what meditators are thinking and feeling here and now. The next is “gathering”, which is redirecting the attention to the physical sensation of breathing (i.e., breathing in and out) as the subject continues to breathe naturally. The last component is “expanding”, which refers to extending awareness of breathing itself to awareness of bodily posture, facial expression, and a sense of the whole body as if it was breathing. Unlike traditional cognitive therapies aimed at altering participants’ cognition, MBCT helps participants to explore their cognitions from a state of mindfulness (Segal et al). After completing the MBCT programme, participants are able to process their cognition differently. Their thoughts are not facts and participants can allow their thoughts to come and go, without paying attention to their content (Germer et al., 2005).

Dialectical Behavioural Therapy (DBT) is a cognitive behavioural treatment with mindfulness at its core because every DBT skill begins with mindfulness meditation (Hayes, Follette, & Linehan, 2004). The concept of mindfulness in DBT is non-judgmental observation, an approach that originates in Zen Buddhism and Western contemplative traditions that promote unconditional acceptance of suffering in life (Hayes et al., 2004). The intervention begins by developing skill in observing emotions, thoughts, and external stimuli by describing them in detail. DBT affirms “acting with awareness” as a skill which can be cultivated through practice to develop the ability to pay attention to activities. Another crucial skill in this MBI process is non-judgemental acceptance. To enhance this skill, participants are encouraged to accept non-judgementally whatever is happening in their daily life (i.e., reality) as well as tolerate unwanted thoughts or feelings (Linehan, 1993a, 1993b).

Acceptance and Commitment Therapy (ACT) is an MBI which combines behavioural principles based on Relational Frame Theory with acceptance, mindfulness, and a reconsideration of values (Hayes, Strostahl, & Wilson, 1999). ACT was proposed with the aim of instructing participants to accept uncontrollable life suffering, while committing to activities embodying the individual’s primary values and making for a meaningful life (Hayes et al., 1999; Hayes, Luoma, Bond, Masuda, & Lillis, 2006). Participants are encouraged to accept the external and internal experiences of the present moment without passing judgement, and at the same time may attempt to modulate or alter a specific behaviour by using reinforcement (Gaudiano & Herbert, 2006). ACT helps participants to minimize the impact of negative cognitive experiences by teaching mindfulness skills, including letting unwanted thoughts and feelings diffuse or fall away, accepting unwanted sensations and drives as well as unpleasant emotions, and focusing attention on the present moment with alertness and openness.

Like ACT, numerous therapeutic approaches draw on mindfulness as an integral component in their methods. One such method is behavioural activation (BA) treatment, which primarily uses behavioural techniques with the application of mindfulness. In BA, mindfulness is applied to address pathological mental conditions caused by dysfunctional rumination, such as depression. BA treatment is conducted without its usual aim to alter the

cognitive content (thoughts) of patients. Patients receive instruction to practise meditation and to be mindful, and are encouraged to note their own ruminations, then switch their attention immediately towards stimuli in the external environment (Jacobson, Martell, & Dimidjian, 2001). In treating major depression, the combination of behavioural activation treatment with mindfulness has been shown to be as effective as medication and traditional cognitive therapy (Dimidjian et al., 2006).

Mindfulness-integrated Cognitive Behaviour Therapy (MiCBT) is another therapeutic approach that employs mindfulness for treatment purposes, to help clients improve their feelings, and change inappropriate behaviours. In this treatment approach, CBT aims to change challenging behaviours by altering clients' unrealistic cognitive experiences. This mindfulness practice aims to increase the subject's ability to gain control over the processes perpetuating such adverse experiences. Specifically, MiCBT attempts to change the client's thinking patterns rather than just the content of their thoughts (MiCBT Institute, 2011).

MiCBT is a four-stage approach designed to help clients reach a level of stability in their emotions and improve their ability to pay attention. The first stage is the personal, in which clients will be trained in mindfulness skills, focusing on internalising attention, aiming for an inner regulation of emotions and thoughts, as well as experiences. The second is the exposure stage, which applies exposure techniques based on behavioural principles to develop self-confidence and reduce reactivity, while maintaining the ability to deal with the experiences of daily life. The third stage is the interpersonal, which focuses on the development of interpersonal understanding and communication skills and develops the ability to face tense situations with others and gain the ability not to react to the reactivity of others. Finally, the empathic stage aims to increase the capacity to empathise and be compassionate towards others, based on immediate experience (Cayoun, 2011).

Generally, mindfulness plays a central role and constitutes the core of the therapeutic processes of MBSR, MBCT and MiCBT. However, cognitive-behavioural components are also added to MBCT and MiCBT treatments (Segal et al., 2002, 2013). By contrast, approaches such as DBT, ACT, and BA, employ mindfulness in a comparatively limited

way as a sub-component in conjunction with their other therapeutic tools to heighten perceptual and sensory awareness in typical circumstances in non-meditative settings (Hayes et al., 1999; Linehan, 1993a)

Benefits of Mindfulness for Health and Well-Being

Early studies showed evidence of the effectiveness of mindfulness treatment, based on changes in specific hypothesised outcomes, such as increased melatonin levels among breast and prostate cancer patients (Massion et al., 1995), or the increased effect of photochemotherapy and phototherapy in patients with psoriasis (Kabat-Zinn et al., 1998). Contemporary research provided evidence of the efficacy of MBIs in reducing the severity of symptoms and developing coping skills in individuals affected by psychological and mental health conditions, such as stress, anxiety, depression, emotional instability, borderline personality disorder, substance abuse and suicidal/self-harm behaviour (Teasdale et al., 2000, Chiesa & Serretti, 2009, 2010; Dimidjan et al., 2006; Hayes et al., 2006; Hofmann, Sawyer, Witt, & Oh, 2010; Ivanovski & Malhi, 2007). Research has also demonstrated the effects of MBIs in improving psychological conditions in non-clinical therapeutic settings (Chang et al., 2004; Chiesa & Serretti, 2009; Ledesma & Kumano, 2009; Bohlmeijer, ten Klooster, Fledderus, Veehof, & Baer, 2010).

For example, the meta-analysis study conducted by Chiesa and Serretti (2009) investigated the evidence for the effect of MBSR in reducing stress in healthy participants. This research included articles published from 1979 to September 2008. The findings showed that MBSR was more efficacious than the control conditions in promoting a reduction in stress and an increase in spiritual values. A specific effect was also found with MBSR in comparison to a structurally equivalent meditation programme. MBSR treatment was shown to be just as able to reduce stress as standard relaxation training. Furthermore, MBSR was able to enhance empathy and self-compassion as well as reduce ruminative thinking and symptoms of chronic anxiety (Chiesa & Serretti).

Ma and Teasdale (2004) conducted another study to investigate the effects of MBCT in preventing relapse in depressed patients who had recovered. The results indicated that MBCT lowered the relapse rate by 36% in participants who had previously undergone

three or more episodes of depression. This led to the conclusion that MBCT can be an effective, efficient way to protect patients recovering from major depression, with three or more previous episodes from relapse or recurrence. Moreover, Lush et al. (2009) showed that MBSR was an effective treatment in reducing pain in patients with fibromyalgia who were suffering from chronic pain. The test involved assessing skin conductance levels before and after the treatment and during mindfulness practice.

ACT treatment showed positive effects for several clinical conditions, such as chronic pain, anxiety, depression, obsessive compulsive disorder, eating disorders, burn-out, post-traumatic stress disorder, psychosis, and opioid use disorder (Bach & Hayes, 2002; Bond & Bunce, 2000; Branstetter, Wilson, Hildebrandt, & Mutch, 2004; Dahl, Wilson, & Nilsson, 2004; Twohig, Hayes, & Masuda, 2006). In a randomized controlled trial study, Bach and Hayes (2002) examined the effect of a brief ACT treatment on 80 inpatient participants suffering from positive psychotic symptoms. Assigned at random, participants received the usual treatment, some with and some without ACT. The results indicated that participants who were treated without ACT responded with reported significantly fewer but also less believable symptoms than those treated with ACT. This means that patients who received ACT together with their usual treatment had a better outcome. Furthermore, ACT participants were rehospitalised at half the rate of participants without ACT over a 4-month follow-up period.

Bond and Bunce (2000) included 90 voluntary participants from a media organisation in their study. Volunteers were randomly allocated to three groups. The ACT group aimed to improve participants' ability to cope with strain related to work, whereas an Innovation Promotion Programme (IPP) group helped participants first to identify, then to make changes in the causes of occupational strain. The final group was the control group. The findings showed that both interventions led to improvements both work-related variables and in mental health. While outcome variable changes were mediated by efforts to reshape stress factors in the IPP group, outcome changes in the ACT group were mediated through accepting undesirable thoughts and feelings.

A study conducted by Branstetter et al. (2004) compared ACT treatment and treatment as usual (TAU) in treating emotional distress among 47 women diagnosed with Stage III or IV ovarian cancer. Patients were assigned to one of the treatments at random. Results showed that patients' mood and quality of life improved in both treatment groups, but the improvements in the ACT group showed a significant advance over those in the TAU group. Additionally, Twohig et al. (2006) applied ACT in treatment for four participants with Obsessive-Compulsive Disorder (OCD). The results indicated that at the end of ACT, scores using standardized measures of OCD decreased significantly in all participants. When followed up 3 months later, these results persisted, and all participants reported a high acceptance of ACT treatment.

Another therapeutic technique employing mindfulness as a sub-component is BA therapy, which was also found to be effective in treating patients with major depression. Employing a randomized placebo-controlled design, Dimidjan et al. (2006) carried out a study to compare the efficacy of BA with cognitive therapy and antidepressant medication based on a sample of 241 adult participants with major depressive disorder. The authors found that BA was as effective as antidepressant medication in treating severely depressed patients, and both were significantly better than cognitive therapy.

In addition, Dahl et al. (2004) studied the effects of a brief ACT intervention with public health sector workers suffering chronic stress or pain, who were likely to draw frequently on sick leave. Nineteen participants were assigned randomly to two groups: the first received medical treatment as usual (MTAU) and the second group received MTAU with ACT (ACT group). At two different points, immediately after the treatment then at the 6-month follow-up, those in the ACT group had claimed a smaller number of sick days and had used less medical treatment than those in the MTAU group, although there were no significant differences between the two groups in levels of stress, pain, or quality of life.

Furthermore, studies have shown that the practice of mindfulness can indeed enhance empathy. In these studies, mindfulness scores were associated with empathy scores through self-reporting instruments (Beitel, Ferrer, & Cecero, 2005; Dekeyser, Raes, Leijssen, Leysen, & Dewulf, 2008; Greason & Cashwell, 2009; Chiesa & Seretti, 2010;

Cameron & Fredrickson, 2015). Moreover, research also indicated that mindfulness practices tend to activate regions of the brain relating to empathy. Chiesa and Serretti, for example, found that mindfulness meditation activated specific brain regions linked to empathy, such as the anterior cingulate cortex, the prefrontal cortex, and the anterior insula. Other studies indicated that practising mindfulness affected areas of a core network activating empathy (Fan, Duncan, de Greck, & Northoff, 2011; Masten, Morelli, & Eisenberger, 2011).

In addition, mindfulness skills can serve as a protective factor against anti-social problems. Wilson et al. (2017), for example, conducted a study with a sample of 112 young people (ages 14 to 24), who had received inpatient detoxification and residential treatment for opioid use disorders. They found that mindfulness skills, assessed by self-report questionnaires, were significantly associated with the time of progression to injected opioid use among young people. The majority of youth with opioid use disorders showed a lack of mindfulness skills. The relationship between opioid use and mindfulness is influenced by emotion regulation capacity (Wilson et al., 2017). A meta-analysis by Karyadi, Vanderveen, and Cyders (2014) suggested that trait mindfulness may help protect against problematic alcohol and tobacco use among youth. Dakwar, Mariani, and Levin's (2011) study with adult patients found that mindfulness deficits are common in the substance-using population.

Mindfulness interventions demonstrated a consistent association with well-being in varied populations (Keng et al., 2011). Some interventions are particularly tailored for pregnant women, who are among the most noteworthy groups. Studies showed that pregnant women derive benefit from mindfulness interventions. For instance, there has been a number of successful psycho-educational programmes supporting prenatal health and well-being that incorporate prenatal mindfulness interventions based on MBSR and MBCT (Byrne, Hauck, Fisher, Bayes, & Schultze, 2014; Vieten & Astin, 2008). Moreover, Matvienko-Sikar, Lee, Murphy, and Murphy (2016) conducted a thorough review of eight studies to evaluate the effects of mindfulness interventions on prenatal well-being. The results indicated that mindfulness interventions improve prenatal well-being with potential

benefits reducing levels of anxiety, depression, and negative affect during pregnancy as well as gains in self-compassion and perceived self-efficacy in childbirth.

Likewise, research has shown that youth affected by learning disabilities are more vulnerable to depression and anxiety than typical young people without these disabilities (Fisher, Allen, & Kose, 1996). Accordingly, this population would benefit from significant support if mindfulness practices could be provided. Singh, Wahler, Adkins, and Myers (2003) developed “soles of the feet meditation”, a meditation practice for people with learning disabilities. Donnelly, James, and Walz (2015) investigated the use of this practice and found that it could reduce aggressive behaviour in their participants. Chapman et al. (2013) carried out a systematic review to explore the effectiveness of mindfulness among people with intellectual disabilities. The authors analysed eleven studies, many of them employing “soles of the feet meditation”. The review indicated that there was a decrease in aggression levels and staff team application of physical restraint, together with an increase in job satisfaction among staff, in parental well-being and satisfaction, and an improvement in parent-child interaction. Similarly, Hwang and Kearney (2013) reviewed 12 studies in their systematic overview and concluded that practising mindfulness helped to reduce challenging behaviour and psychological problems in subjects with learning disabilities.

Yildiran and Holt (2015) concluded that adults with learning disabilities who practised mindfulness could improve mood, concentration, and relationships with others. However, questions were raised about how mindfulness practices could be implemented in this population. More recently, Thornton et al. (2017) described a study with a group of five young participants with learning disabilities, and their carers. The authors developed an evidence-based mindfulness practice which was appropriate for the participants and aimed to examine whether youth with learning disabilities could practise mindfulness. The carers reported that the young people continued to practise mindfulness in their daily lives after the group exercise was completed. These findings led to the conclusion that young people with learning disabilities may find mindfulness effective and accessible. However, due to differences in degrees of learning disability, participants might acquire mindfulness skills at different levels. Thus, if the study were conducted using a method capable of distinguishing between participants' levels of competence in mindfulness, the results might

reveal not only the effectiveness of mindfulness practice but also changes in mindfulness skills before and after mindfulness treatment. Advanced treatment could then be offered for those making progress.

In recent years, there has been a dramatic increase in the number of mindfulness research publications (Black, 2014). When evidence demonstrated the positive effects of MBIs in therapeutic settings (Chang et al., 2004; Chiesa & Serretti, 2009; Ledesma & Kumano, 2009; Bohlmeijer et al. 2010), research focused more on the application of mindfulness practice in many different contexts. A growing body of validation studies reported that mindfulness is positively associated with well-being and provides various psychological and physical, benefits, and even in performance (Baer et al., 2006; Brown & Ryan, 2003; Carlson & Brown, 2005; Christopher & Gilbert, 2010; Frewen, Evans, Maraj, Dozois, & Partridge, 2008). More mindfulness programmes have been offered to employees in many organizations and corporations, such as Google (Kelly, 2012), General Mills (Gelles, 2012), and Aetna (Wolever et al., 2012). Such mindfulness programmes showcase their therapeutic effect in enhancing employee well-being and effectiveness (Hyland, Lee & Mills, 2015). Studies have found that mindfulness may develop task commitment and enjoyment, as well as memory, and also improve resilience, task performance and even social relationships among employees (Levy, Wobbrock, Kaszniak, & Ostergren, 2012; Glomb, Duffy, Bono, & Yang, 2012).

Studies on the MBSR have demonstrated that such MBIs successfully reduce stress levels in health professionals (Fortney, Luchterhand, Zakletskaia, Zgierska, & Rakel, 2013; Krasner et al., 2009; Praissman, 2008; Shapiro, Astin, Bishop, & Cordova, 2005). A study conducted by Irving, Dobkin, and Park (2009) found evidence for MBSR as an intervention promoting wellness among healthcare providers. MBIs helped healthcare providers reduce depersonalization in patients and their families (Goodman & Schorling, 2012), and increased their empathy towards their colleagues (Krasner et al., 2009; McCracken & Yang, 2008; Shapiro, Schwartz, & Bonner, 1998). Beach et al. (2013) found that mindfulness was linked with the patient-centred care model among 45 health practitioners, meaning that healthcare providers offer more patient-centred communication and satisfaction.

Shapiro et al. (2005) also suggested that MBIs have a positive effect on burnout and improve life satisfaction among healthcare professionals. Other studies have indicated that MBIs improve mood, wellbeing and psychological functioning in mental health trainees (Shapiro, Brown, & Biegel, 2007), nurses (Cohen-Katz, Wiley, Capuano, Schaufeli, & Shapiro, 2005; Foureur, Besley, Burton, Yu, & Crisp, 2013), and primary care physicians (Krasner et al. 2009). Recently, Janssen, Heerkens, Kuijer, Heijden, and Engels (2018) carried out a systematic review of 24 articles aiming to evaluate the effects of MBIs (e.g., MBSR and MBCT) on the mental health of employees. The findings showed that practising MBIs could help participants reduce their levels of psychological distress, emotional exhaustion, stress, anxiety, depression and work stress. The results also indicated that there were improvements in mindfulness skills, sleep quality, as well as in measures of relaxation, personal accomplishment and self-compassion among employees receiving MBIs. These findings lead to the suggestion that MBIs may help to improve psychological functioning among healthcare providers.

Moreover, Jacobs et al. (2017) examined whether mindfulness-based skills training can improve sleep quality and reduce problems of stress and burnout in a diverse paraprofessional workforce sample. This study included 26 paraprofessionals who completed assessments at three stages: before and after training, and 4 weeks after the completion of training. As the results showed, participants reported that their sleep quality improved and stress levels and emotional exhaustion declined significantly after mindfulness training. Participants also found mindfulness practices acceptable (Jacobs et al). Another study conducted by Byron, Ziedonis, McGrath, Frazier, and Fulwiler (2015) reported that employees who received mindfulness training showed improvement in their assessment of sleep quality and in self-reported measures of perceived stress when compared to control groups. Byron et al. also reported that job burnout scores decreased among employees who were in a mindfulness training group. However, even though these studies confirmed the effectiveness of mindfulness meditation programmes, they did not provide a clear distinction between dispositional mindfulness (trait) and dynamic changes of mindfulness (state) over time.

In addition, mindfulness was applied in educational contexts with positive outcomes (Bush 2011; Hyland et al., 2015; Hwang et al., 2019). For example, according to Hyland et al., a number of schools in the USA, such as the Stern School of Business at New York University, Boalt Hall School of Law at the University of California, Berkeley, Drucker Graduate School of Management, and Harvard Business School, have implemented mindfulness training programmes. For younger students, a study carried out by Schonert-Reichl, Oberle, and Lawlor (2015) found that mindfulness-based programmes improved emotional control, empathy, optimism, perspective-taking, self-image in school, and peer acceptance, at the same time reducing symptoms of depression, stress levels, and peer-rated aggression among fourth and fifth grade children.

Broderick and Metz (2009) applied a mindfulness-based programme adapted from the MBSR model to senior high school girls and found that participants in the programme showed improvement in self-acceptance, self-reported calm and relaxation in comparison with the control group. Authors also found that negative affect increased significantly in participants in mindfulness-based programmes in comparison to other groups. Flook et al. (2010) carried out a randomized 8-week controlled trial of school-based mindfulness activities for second and third graders and found that there were significant improvements in executive function in the sample. Schonert-Reichl and Lawlor (2010) piloted a mindfulness programme for fourth to seventh graders which consisted of 10 lessons in their classroom. The findings from this programme reported that students who received the intervention improved in self-reported enthusiasm, positive affect, and less disruptive behaviour compared to the control group. Also, teacher ratings indicated better attention and social-emotional competence in students who received the intervention compared with those in the control group. This study could not specify the effect of mindfulness deriving from the exercise of mindfulness skills, however, due to the lack of a means of assessing mindfulness levels in participants.

Furthermore, the application of mindfulness in sports has shown some positive results (Birrer, Röthlin, & Morgan, 2012). For example, Kabat-Zinn, Beall, and Rippe (1985) developed and provided a specific application of mindfulness for Olympic rowers. The researchers reported that by comparison with their previous physical ability and level

of experience, their subjects performed above their coach's expectations. Several who had won Olympic medals explained that the practice of mindfulness had assisted them to reach their full potential. By the end of the 2000s, two specific MBIs for sport had been developed, known as the Mindfulness–Acceptance–Commitment approach (MAC; Gardner & Moore, 2007) and the Mindful Sports Performance Enhancement approach (MSPE; Kaufman, Glass, & Arnkoff, 2009).

Moreover, research on MBIs in sports suggests that there are psycho-physiological health effects that improve performance in athletes or sports players (Birrer et al., 2012). Due to the evidence that practising mindfulness leads to reductions in stress and anxiety, Hewett, Ransdell, Gao, Petlichkoff, and Lukas (2011) suggested that increasing mindfulness skills would help athletes develop an awareness of their potential stressors and decrease perceived stress, thus lowering their resting heart rates. This efficient heart function can lead to improved physical performance and greater endurance. Additionally, mindfulness practices alter athletes' sympathetic nervous systems, which may also help them to control their anxiety levels (Hewett et al., 2011) and improve their concentration (Bernier, Thienot, Codron, & Fournier, 2009; Gardner & Moore, 2004), thus boosting their performance in sport.

Recently, using a representative sample of 34,525 people in the US, Strowger, Kiken, and Ramcharran (2018) investigated whether mindfulness meditation was associated with activity and with meeting recommendations for aerobic physical activity. The findings showed that American adults who had practised mindfulness meditation over the past year were more active and more likely to meet physical activity recommendations. Also, the association between mindfulness meditation and indices of physical activity was stronger than that for mantra meditation and guided imagery. However, this study of mindfulness meditators focused only on the previous year of practising mindfulness meditation and did not differentiate between individual meditation experiences among participants (e.g., some might have stopped practising mindfulness for a few months before the study was conducted, or the number of years practising meditation might be different). If this study had used mindfulness assessment tools to assess participants' mindfulness levels, such a weakness could have been avoided.

Due to individual differences, practitioners may have differing abilities to acquire mindfulness skills through practice. If the results of a mindfulness study are based only on changes in related variables (e.g., anxiety score, stress score, depression score), they would provide no evidence that these changes occurred because of mindfulness interventions and were not due to a third variable (e.g., a therapist effect). Proper evaluation therefore requires a reliable and valid mindfulness assessment score that accurately reflects the mindfulness ability of an individual (trait mindfulness) or a person's mindfulness levels (state mindfulness). The reliable, valid measurement of individual mindfulness levels will ensure that the outcomes of mindfulness studies also are reliable and valid if they are reflected by direct changes in mindfulness. In other words, accurate mindfulness scores ensure that the correlation between a mindfulness intervention and outcomes is the only relation between them. If there is no variable representative for mindfulness, the effect should be questioned as to whether it (the effect) derives from mindfulness or from something else. An accurate measurement of mindfulness along with other outcomes is important to ensure the reliability and validity of the reported results/effects of MBIs. However, this essential point was neglected in the majority of studies, in which researchers evaluated all health-related outcomes (e.g., stress, depression, anxiety) (Medvedev et al., 2017a).

Conclusion

Mindfulness involves directing the attention in the present moment to each event experienced in mind and body with a non-reactive, non-judgmental, and accepting attitude. Mindfulness has been practised for more than 2,000 years and has yielded many positive results associated with both physical and mental health. Contemporary psychological research provides evidence of the benefits of mindfulness on health and wellbeing through the application of MBIs. The increasing body of evidence continues to indicate that there is a positive effect from utilising MBIs in clinical settings as well as in non-clinical settings, such as workplaces and sports. One of the most important concerns is the accurate measurement of mindfulness, i.e., how mindfulness can be measured accurately and rigorously by using reliable, valid assessment tools. Conducting an accurate measurement of mindfulness and of other variables simultaneously would ensure that the reported

outcomes of MBIs were both reliable and valid. However, this requirement has been largely disregarded because researchers have focused mainly on outcomes related to health concerns, such as stress, depression, anxiety.

Interest in mindfulness-based approaches is growing and calls for more attention to be given to rigorous research, accompanied by reliable and valid assessment in this area.

Chapter 2 Measuring Mindfulness

Overview of Mindfulness Assessment Tools

Prior to the development of psychometric assessment tools used in mindfulness research, mindfulness could not be measured directly but only through the evaluation of the effectiveness of MBIs. For example, mindfulness research investigated the effectiveness of intervention based on physiological measures, such as the levels of melatonin in cancer patients (Massion et al., 1995), or skin conductance levels in patients with psoriasis (Kabat-Zinn et al., 1998). In neurophysiological studies, gauging the effectiveness of mindfulness was based on changes in specific hypothesized outcomes, using electroencephalograms (EEG), including event-related potential (ERP) and neuroimaging techniques (i.e., MRI – Magnetic Resonance Imaging, fMRI – Functional Magnetic Resonance Imaging) (Cahn & Polich, 2006; Chiesa & Serretti, 2010; Coelho, Canter, & Ernst, 2007). Experience sampling (Frewen, Unholzer, Logie-Hagan, & MacKinley, 2014) or counting breaths (Levinson, Stoll, Kindy, Merry, & Davidson, 2014) were alternative standards to assess the effectiveness of mindfulness interventions, but such techniques were rarely used in research, perhaps because their reliability and validity were not well established.

The main problem with the methods used in neurophysiological research was that they assessed mindfulness indirectly. It was difficult to use such assessment methods to reliably monitor dynamic (state) or enduring (trait) changes when conducting mindfulness interventions, so that the upshot was mixed results for MBIs. Consequently, mindfulness studies which did not control for mindfulness levels might raise concerns about their validity. Self-reporting mindfulness assessment tools were proposed to address these problems. Many contemporary studies on applied MBIs rely on self-reporting measures to evaluate the effectiveness of the interventions (e.g., Shapiro et al., 2005; Ma & Teasdale, 2004). The importance of self-reporting mindfulness assessment tools may be explained by the subjective nature of human experience of the world, the self and their interaction. However, assessing such experience using more objective (e.g., neurophysiological) measures (Libet, 2004) it was problematic.

Assessing Mindfulness in Neurophysiological Research

Neurophysiological studies on mindfulness employed several methods but the EEG was the main one used to gather data by recording electrical activities or wave patterns of the brain. Studies found that brain wave patterns correlated with an individual's psychological state (e.g., wakefulness). For instance, gamma waves (30-100 Hz) were shown to relate to many sensory (e.g., pain) and cognitive functions (Fitzgibbon, Pope, Mackenzie, Clark, & Willoughby, 2004). Beta wave activity (12-30 Hz) was related to activity, busyness or alertness, and active concentration (i.e., working) (Baumeister, Barthel, Geiss, & Weiss, 2008). Alpha waves (8-12 Hz) were associated with a state of relaxation and reflection. Theta waves (4-8 Hz) were linked to the transitional state between sleeping and waking (e.g., drowsiness, meditation) while delta waves (< 4 Hz) indicated that an individual was in deep sleep (Gazzaniga, Ivry, Mangun, & Steven, 2009).

Kasamatsu and Hirai (1966) carried out an EEG study to compare the brain wave patterns of participants who practised traditional Buddhist mindfulness, called Zen meditation, at three points in time— before, during and after engaging in meditation. The results indicated that the changes in EEG activities (i.e., a decline in alpha frequency, an increase in frontal alpha activity, and a theta burst) were associated with high levels of meditation experience, in comparison to “before meditation” and control conditions. In other words, the authors learned which brain waves reflected the practice of mindfulness. Similar outcomes were replicated in the studies conducted by Murata, Koshino, and Omori (1994), and Takahashia et al. (2004).

Another finding, reported by Kasamatsu and Hirai (1966), indicated that participants engaged in meditation showed an absence of alpha-blocking habituation in response to the repetition of a clicking sound, but those in the control group did not. Alpha-blocking, which is recognised as a reduction of alpha power on exposure to a stimulus, is indicated by the EEG. The absence of alpha habituation in meditators may indicate that they are paying attention to the present moment while in the relaxed state of meditation (Kasamatsu & Hirai). However, this discrepancy between meditators and those in control

groups was not detected in the study conducted by Becker and Shapiro (1981), possibly the result of differing experience among meditators in the samples.

Another study using EEGs is the research by Dunn, Hartigan, and Mikulas's (1999), which compared brain activity in 10 participants practising mindfulness meditation, focused meditation, and engaged in regular relaxation conditions. The results showed that mean amplitudes (i.e., the amplitudes of frontal theta, frontal and posterior delta, beta in all cortical areas, and central and posterior alpha) were significantly higher when subjects were engaged in mindfulness meditation than in the other two activities. This led to the conclusion that mindfulness meditation is unique and completely different from focused meditation and relaxation (Dunn et al., 1999).

However, the other meditation techniques were performed differently, which is a limitation of the study. Whereas participants had their eyes open during mindfulness practice, those engaged in both relaxation and focused meditation sat with their eyes closed. The EEG-based evidence showed that the average of alpha, beta, delta and theta activities increased with the eyes open in comparison with the stance with the eyes closed (Barry, Clark, Johnstone, Magee, & Rushby, 2007).

Another limitation of the study by Dunn et al. (1999) was that the same group of participants received the same training, assuming that they might switch from one to the other. According to Chiesa and Serretti (2010), an experience of mindfulness would affect other experimental conditions in a meditative study or would influence the efficiency of MBIs. In addition, the failure of some replicated studies might be explained by the fact that such studies did not measure mindfulness levels, nor did they differentiate between state or trait changes. Trait changes were more likely to be detected through EEG metrics under experimental conditions (Stelmack, 2004).

Davidson et al. (2003) examined changes in cortical activity in two groups: MBSR and a control group. In their study of 41 participants, 25 received MBSR and 16 were in the control group. The variables were measured using EEGs and psychometric measures of anxiety and affect. The results showed that in participants receiving MBSR, there was significantly higher left-side anterior activation, associated with a positive affective style,

than in those in the control group. However, these positive changes were not found in the study applying MBCT (Keune, Bostanov, Hautzinger, & Kotchubey, 2011). Furthermore, Davidson et al. (2003) found that the level of anxiety and negative affect decreased significantly, and antibody cell numbers rose in the mindfulness group in comparison to the control group.

Another EEG study by Berkovich-Ohana, Glicksohn, and Goldstein (2011) indicated that mindfulness meditation is correlated with a reduction of gamma power in frontal areas and an increase of posterior gamma activity. This reduction is linked to a decrease in self-referential processing, whereas an increase is related to heightened sensory attention. However, Berkovich-Ohana et al. reported that a difference in gamma band activity occurred in the mindfulness group in the closed-eyes stance, regardless of participants' level of experience, a result which limits the generalisability of their findings.

Studies used neuroimaging techniques (i.e., MRI and fMRI) yielded evidence for the effect of mindfulness meditation on the activities of the brain but failed to measure mindfulness levels. One fMRI study was conducted by Pagnoni, Cekic, and Guo (2008) with 12 participants who had practised Zen meditation every day for more than 3 years, and 12 participants, who had never practised meditation, as control subjects. The results showed that compared to participants in the control group, experienced Zen meditators were characterised by a reduced duration of neural response related to conceptual automatic perception. The researchers suggested that the meditative experience of an individual with training may facilitate the ability to voluntarily regulate the flow of spontaneous mental processes.

Moreover, Pagnoni and Cekic (2007) used MRI in their research to examine how meditation practice may affect the decline of grey matter volume and cognitive performance related to age. The study reported that there was no meaningful correlation between grey matter volume and attentional performance, and age among participants who were Zen meditators. By contrast, participants who were non-meditators showed a negative correlation of both measures with age. These findings suggest that Zen meditation practice

may be a protective factor helping to prevent the decline of grey matter volume and the deterioration of cognition related to age.

Consistent with the above findings, Hölzel et al. (2011) reported an increase in grey matter density in regions of the brain after MBSR training, including the cingulate cortex, hippocampus, temporo-parietal junction and cerebellum. These areas are linked with the processes of learning, memory, self-related cognition and emotional regulation. An fMRI investigation by Ives-Deliperi, Solms, and Meintjes (2010) found that during mindfulness meditation, the signal was significantly altered in brain regions (specifically the left ventral anterior cingulate cortex, bilateral anterior insula, bilateral precuneus, right medial prefrontal cortex, and right posterior cingulate cortex) which are associated with emotional regulation and self-perception, resulting in altered experiences of the self. However, this study was affected by methodological limitations, due to its lack of a control group as well as a baseline condition for comparison (Ives-Deliperi et al., 2010). Also, the results of these mindfulness studies using neuroimaging techniques did not reveal differences in neurophysiological changes due to differing abilities among meditators, nor did they differentiate between dynamic or static changes caused by mindfulness interventions.

Besides research conducted employing EEG and neuroimaging techniques, a limited number of studies use acceptable methodology to investigate the effects of mindfulness on autonomic functions (Cahn & Polich, 2006; Chiesa & Serretti, 2010). Barnes, Davis, Murzynowski, and Treiber (2004) evaluated the influence of a mindfulness meditation programme on heart rate and blood pressure in healthy youth with normal blood pressure. The authors found that both before and after the test, significant differences in blood pressure and heart rate were detected between the meditative group and the control group.

Takahashi et al. (2004) measured heart rate variability for Zen meditators and control subjects, and found that Zen meditator participants showed an increase in parasympathetic activity during meditation and a reduction in sympathetic activity in contrast with the control group. Carlson, Speca, Farris, and Patel (2007) found a positive effect from MBSR on the immune system, endocrine levels and blood pressure of breast

and prostate cancer outpatients. This evidence suggests that practising mindfulness may play an essential role in facilitating autonomic functions. However, further research in this field needs to apply more rigorous data collection methods (Carlson et al.). For example, data should reflect the level of mindfulness practices in participants and a measure of the stability of the changes in order to verify the effects, whether a state or a trait.

A number of other neurophysiological studies have been omitted from this chapter due to their methodological limitations, because they use small samples and/or simplified or unspecific mindfulness-related techniques. For instance, a study conducted by Barnhofer, Chittka, Nightingale, Visser, and Crane (2010) used a small sample with 17 participants and simplified mindfulness practices (i.e., depending only on breathing exercises and wishing others well). Reviews of neurophysiological studies reveal their most common limitations, which include the absence of measurements of mindfulness levels, invalid comparisons between research groups (e.g., because of the absence of baseline conditions or control groups, EEG with closed eyes versus open eyes) (Barry et al., 2007), as well as non-matching demographic variables between groups which might affect EEGs and neuroimaging data (Erwin, Mawhinney-Hee, Gur, & Gur, 1989).

Moreover, neurophysiological studies have illustrated that mindfulness meditation has an influence on changes in human biological phenomena and correlates with a reduction in psychological symptoms. They have also indicated that different mindfulness practices may have differing effects on neurophysiological functions over time. However, there have been somewhat contradictory results in neurophysiological studies (e.g., Kasamatsu & Hirai, 1966 versus Becker & Shapiro, 1981; Davidson et al., 2003 versus Keune et al., 2011), possibly because participants were often recruited based on their experience with meditation (e.g., number of years of practice). This criterion may be inaccurate and may not reflect participants' competence in mindfulness or mindfulness levels because these people may be employing different approaches to mindfulness meditation, such as Zen, MBSR and MBCT. Therefore, it may be suggested that psychometric trait-mindfulness assessment tools should be used to select participants appropriately.

Also, individual differences, such as the capacity to acquire mindfulness skills and the varying frequency of practice sessions, may influence an individual's mindfulness levels (Dalai Lama et al., 2004; Kabat-Zinn, 2000). However, the majority of neurophysiological studies on mindfulness do not measure individual mindfulness levels and that may seriously undermine their results (Park, Reilly-Spong, & Gross, 2013; Chiesa & Serretti, 2010; Medvedev et al., 2017; 2017a).

Moreover, mindfulness has been conceptualised as both a state and a trait, so that the validity of mindfulness studies can only be established if reliable, valid measures are used to control for mindfulness levels, distinguishing clearly between dispositional (trait) and dynamic (state) mindfulness (Medvedev et al., 2017a). Apparently, neurophysiological mindfulness research has been unable to distinguish trait from state mindfulness (Park et al., 2013; Chiesa & Serretti, 2010; Medvedev et al., 2017; 2017a). Therefore, neurophysiological studies on mindfulness have been unable to reliably demonstrate expected changes in mindfulness levels to support their validity, since verification requires the development of reliable, valid instruments to assess the construct. This finding confirms the importance of reliable, valid psychometric instruments to assess mindfulness.

A number of psychometric mindfulness instruments have been proposed but only a few of them have been rigorously examined for their validity and reliability, employing appropriate methods. The use of a psychometric instrument whose validity and reliability have not been established can seriously affect the validity of clinical interventions. An investigation and modification of measurement tools with appropriate methodology, such as the generalisability theory method, can enhance precision in measurement (Medvedev et al., 2017). Enhancing the psychometric properties of commonly-used assessment tools of mindfulness and ascertaining their ability to differentiate between state and trait mindfulness would be advantageous for evaluating both transient and long-lasting psychological changes that can be connected with a mindfulness-based intervention, a particular type of mindfulness meditation practice, and any related therapeutic result.

Psychometric Mindfulness Measures and Their Properties

Table 1 provides a list of self-report mindfulness assessment tools, together with the number of times these assessments have been cited and their basic psychometric properties. The first developed self-report mindfulness assessment tool is the Freiburg Mindfulness Inventory (FMI; Buchheld, Grossman, & Walach, 2001; Walach, Buchheld, Büttenmüller, Kleinknecht, & Schmidt, 2006). However, this mindfulness assessment tool was specifically designed to evaluate the progress of Vipassana meditators, suggesting that it may not be suitable for more general mindfulness assessment (Belzer et al., 2013). Since then, several other scales were developed, with the Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003) being the most cited measure. After the MAAS, the Five Facets Mindfulness Questionnaire (FFMQ; Baer et al., 2006) is widely used and cited, and the “act with awareness” subscale of the FFMQ includes the most robust MAAS items. This means that the MAAS is inherently included in the FFMQ.

Table 1. Psychometric properties of common mindfulness assessment tools and the number of citations of each scale in Google Scholar and Web of Science (22 July, 2019).

Scales	Reference (original)	Subscales /items	Cronbach's alpha	Test-retest reliability	Validity construct	Google scholar	Web of science
MAAS	Brown & Ryan (2003)	1/15	0.78 – 0.92	ICC*= .81;	FMI, KIMS, MMS, CAMS-R, SMQ: $r = 0.14 - 0.51$;	9431	3410
FFMQ	Baer et al. (2006)	5 /39	0.67 - 0.93	not reported	Based on FMI, KIMS, MMS, SMQ, and MAAS items.	5008	2086
KIMS	Baer et al. (2004)	4/39	0.72 - 0.97	$r = 0.81 - 0.86$, Observe $r = 0.65$	FMI, SMQ, CAMS $r = 0.51 - 0.67$	2344	834
SMS	Tanay & Bernstein (2013)	2/23	0.95	$r = 0.64 - 0.65$	TMS, FFMQ: $r = 0.31 - 0.47$	2099	889
FMI	Buchheld et al. (2001) Walach et al. (2006)	1/30 1/14	0.80 - 0.94 0.86	not reported not reported	KIMS, MAAS, SMQ, and CAM-R: $r = 0.31 - 0.60$	651 1032	446
CAMSR	Feldman et al. (2007)	1/12	0.61 - 0.81	not reported	KIMS, MAAS, FMI and SMQ: $r = 0.51 - 0.67$;	1016	381
TMS	Lau et al. (2006)	2/13	$\alpha = .85 - .91$	not reported	KIMS, MAAS, FFMQ, CAMS-R, SMQ, FMI $r = .10 - .74$	1114	397
PHLMS	Cardaciotto et al. (2008)	2/20	0.75 - 0.91	not reported	KIMS, MAAS $r = 0.38 - 0.61$;	759	271
CAMM	Greco et al. (2011)	1/10	0.81	not reported	Based on KIMS items	386	156
SMQ	Chadwick et al. (2008)	1/16	0.82 - 0.89	not reported	KIMS, MAAS, FMI: $r = 0.38 - 0.61$;	374	136
MMS	Haigh et al. (2011)	4/21	0.95	$r = 0.64 - 0.65$	TMS, FFMQ: $r = 0.31 - 0.47$	90	35
CHIME	Bergomi et al. (2014)	8/37	0.70 - 0.90	0.70 - 0.90	FFMQ: $r = 0.85$	36	

Notes. *Intraclass correlation coefficient; MAAS: Mindful Attention Awareness Scale; FFMQ: Five Facets Mindfulness Questionnaire; KIMS: Kentucky Inventory of Mindfulness Skills; TMS: Toronto Mindfulness Scale; FMI: Freiburg Mindfulness Inventory; CAMS-R: Cognitive and Affective Mindfulness Scale-Revised; PHLMS: Philadelphia Mindfulness Scale; CAMM: Child and Adolescent Mindfulness Measure; SMQ: Southampton Mindfulness Questionnaire; MMS: Mindfulness/Mindlessness Scale; SMS: State Mindfulness Scale; CHIME: Comprehensive Inventory of Mindfulness Experiences.

The TMS was specifically designed to measure state mindfulness while the other scales (MAAS, KIMS, FFMQ CAMS-R, SMQ, FMI, and PHLMS) were designed to assess trait mindfulness (Bergomi et al., 2013). The Child and Adolescent Mindfulness Measure (CAMM), which was adapted from the KIMS, was a mindfulness instrument used for child and adolescent populations. Of all the mindfulness scales, the MAAS, the FFMQ, and the KIMS are the most popular, and are the most commonly used by researchers. These are also the scales most evaluated and cited (Table 1). The FFMQ is the most widely used dispositional mindfulness assessment tool and, using factor analysis, measures five different aspects of mindfulness by incorporating items from other measurement approaches available at the time (Baer et al., 2006). Accordingly, the FFMQ and its development will be described in the following section in greater detail. The section will also describe other available measures that provided the basis for the FFMQ's development.

Firstly, the MAAS is the most cited among all mindfulness instruments, and the number of studies which supported its assessment of psychometric properties is larger than that of any other tool (Park et al., 2013). Many studies have also supported the one-dimensional character of the MAAS (e.g., Brown & Ryan, 2003; Carlson & Brown, 2005; Christopher, Charoensuk, Gilbert, Neary, & Pearce, 2009; MacKillop & Anderson, 2007). The MAAS is a 15-item self-reporting measure that employs a 6-point Likert-scale response format, ranging from 1 = "almost always" to 6 = "almost never", to assess a subject's attention and awareness of current or immediate experience. This scale is a unidimensional mindfulness assessment tool suitable for the variety of purposes in measuring mindfulness. The scores assessed through the MAAS vary among individuals and can be used to measure the result of practice (Brown & Ryan, 2003).

Cordon and Finney (2008) applied the MAAS in their study and found that participants who were non-meditators had significantly lower MAAS scores than those who engaged in practicing meditation regularly. Other research found that MAAS scores were positively correlated with scores measuring openness, well-being and positive affect, and negatively correlated with measures of anxiety, stress, rumination and neuroticism. Such findings support the validity of the MAAS (Baer et al., 2006; Brown & Ryan, 2003; Carlson & Brown, 2005; Christopher & Gilbert, 2010; Frewen et al., 2008). Studies on

MAAS reliability indicate that it achieves a good degree of internal reliability and acceptable external reliability (test-retest) (Table 1).

Furthermore, studies aiming to test the MAAS convergent validity have shown positive correlations, ranging from weak to moderate, between the MAAS and other measures of mindfulness, including CAMS-R, FMI, MMS, KIMS, and SMQ (Baer et al., 2006; Brown & Ryan, 2003; Christopher & Gilbert, 2010). The validity of the MAAS construct was also supported by positive correlations between the MAAS scores and several variables such as well-being measures, openness and positive affect , and by negative correlations between the MAAS and measures of rumination, neuroticism, anxiety and stress (Baer et al.; Brown & Ryan; Carlson & Brown, 2005; Christopher & Gilbert; Frewen et al., 2008). However, although the MAAS has been affirmed for its construct validity and reliability with large studies and for being the most cited unidimensional measure, it may be called in question for its unidimensional conceptualisation of dispositional mindfulness because the latter is commonly regarded as a multifactorial concept (Siebelink et al., 2019).

In contrast, the KIMS is a multi-dimensional assessment tool for assessing mindfulness skills in the mindfulness component integrated with DBT (Baer et al., 2004; Dimidjian & Linehan, 2003). The KIMS comprises 39 self-report items designed to assess four essential skills of mindfulness according to four subscales. These subscales are labelled “observe”, “describe”, “accept non-judgementally”, and “act with awareness” (Linehan, 1993a; Segal et al., 2002). This scale applies a 5-point Likert-scale approach, with responses spanning from “1= never” to “5= almost always”. The subscale termed “accept non-judgementally” measures an individual’s self-criticism and judgemental behaviour as these are described in the common definitions of mindfulness (Baer et al). The “observe” subscale gauges an individual’s ability to pay attention to experiences in the present moment. The “act with awareness” subscale refers to assessing the subject’s ability to be fully aware in the course of any activities in which a person is engaged (e.g., eating, walking, and driving). “Describe” indicates a subscale measuring an individual's ability to describe their experiences, both internal and external.

As can be seen, all KIMS subscales, with the exception of the “describe” subscale, appear to be consistent with the definitions of mindfulness given in Chapter 1. The novel “describe” skill is not included in most definitions of mindfulness from a psychological point of view or even in traditional mindfulness concepts (Dalai Lama et al., 2004; Gunaratana, 2002). According to Baer (2003), the ability to describe individual experiences may be useful in a therapeutic process but it seems to be unrelated to the concepts of mindfulness both psychologically and traditionally. The reason for its inclusion in the KIMS may possibly be based on Nickerson’s (1978) old idea, that a subject’s ability to focus on the present moment is reduced by linguistic processing and therefore descriptive skill is necessarily included with other subscales to fully measure mindfulness levels.

The KIMS scale has shown good external reliability and acceptable internal consistency, as have all its subscales ($r = 0.81$ to 0.86), except for the “observe” subscale ($r = 0.65$). Evidence from the application of exploratory factor analysis (Nunnally & Bernstein, 1994) has provided backing for the KIMS’ four-factor model (Baer et al., 2004). Confirmatory factor analysis has also provided support for the four-dimensional KIMS model but has provoked questions about the validity of its score (Baer et al., 2004; Baum et al., 2010). The validity of the “accept without judgement” and “act with awareness” subscales has received good supporting evidence but the remaining subscales of “describe” and “observe” have not (Baer et al.; Christopher & Gilbert, 2010; Frewen et al., 2008). The KIMS has also shown its limitations by not achieving low correlations among subscales, raising concerns about the validity of its content (e.g., the “describe” subscale). For the “observe” subscale (Table 1), the test-retest reliability score of 0.65 was lower than the acceptable cut-off point for measuring a trait, meaning that such a measure seemed indicative of a state (Barker, Wadsworth, & Wilson, 1976; Spielberger, 1999).

Proposed due to disagreement concerning the number of mindfulness dimensions, the CHIME scale consists of 37 items and was developed in the German language by Bergomi et al. (2014). The English version is not available to date. Bergomi, Tschacher, and Kupper (2013) conducted an analysis of currently available mindfulness measures and identified eight mindfulness aspects. Thus, the CHIME was constructed to cover all mindfulness aspects. The subscales of the CHIME are “awareness of external experiences”,

“awareness of internal experiences”, “accepting a nonjudgmental attitude”, “acting with awareness”, “openness to experience”, “nonreactive decentering”, “insightful understanding” and “awareness of thought relativity” (Bergomi et al., 2013). The full CHIME scale with all its subscales achieved satisfactory internal consistency, with α ranging from 0.70 to 0.90, and acceptable reliability through test-retest with r ranging from 0.70 to 0.90 (Bergomi et al.).

The report of Bergomi et al. (2014) also confirmed the structure of the eight-facet CHIME scale by the strong correlations ($r = 0.85$) between CHIME and FFMQ in their total scores of a sample size of 202 participants. There were also strong correlations between the individual FFMQ subscale scores and the conceptually similar CHIME subscale scores, a result which confirmed CHIME’s construct validity. The authors also reported that the total CHIME score correlated moderately with measures of wellbeing, depression, and anxiety (Bergomi et al.). Even though CHIME demonstrated its acceptable psychometric properties through classical test theory approaches (e.g., test-retest reliability, Cronbach's alpha), further research is needed to apply modern approaches (e.g., G-theory) to investigate its temporal reliability and its ability to discriminate between state and trait mindfulness.

The Toronto Mindfulness Scale (TMS) (Lau et al., 2006) differs from the scales mentioned above owing to its design focusing on the assessment of state mindfulness. The TMS was developed on the basis of the proposal by Lau et al. that mindfulness should be understood as a state-like quality. The TMS comprises two subscales composed through EFA and supported by CFA (Lau et al.). Called “curiosity” and “decentering”, these subscales have shown good internal consistency with Cronbach’s alpha, ranging between 0.86 to 0.91 and 0.85 to 0.87, respectively (Park et al., 2013). The modest correlation ($r = 0.42$) supports the TMS’s two-dimensional structure (Lau et al.).

Several studies have been carried out to compare the various TMS subscales and also these subscales with other mindfulness assessment tools. The results found that the TMS “decentering” subscale showed higher correlations (r ranges from 0.20 and 0.74) than the “curiosity” subscale (r ranges from 0.10 to 0.54) (Davis, Lau, & Cairns, 2009). Others

indicated that the “decentering” subscale correlated highly with the MAAS (Brown & Rayan, 2003), the Freiburg Mindfulness Inventory (Walach et al., 2006), the Kentucky Inventory of Mindfulness Skills (Baer et al., 2004), the Cognitive and Affective Mindfulness Scale-Revised (Feldman, Hayes, Kumar, Greeson, & Laurenceau 2007), the Southampton Mindfulness Questionnaire (Chadwick et al., 2008), and the Five Facets Mindfulness Questionnaire (Baer et al., 2006).

Furthermore, both the “curiosity” and “decentering” subscales correlated positively with other measures, such as the Tellegen Absorption Scale (Tellegen, 1982), the Reflection subscale of the Rumination-Reflection Questionnaire (Trapnell & Campbell, 1999), the surroundings subscale of the Situational Self-Awareness Scale (Govern & Marsch, 2001; Lau et al., 2006), and the Psychological Mindedness Scale (Conte et al. 1990). On both TMS subscales, participants who were meditators tended to achieve higher scores than those who were not experienced meditators (Davis et al., 2009). The scores on the “decentering” subscale tended to reflect competence in mindfulness among meditators (Davis et al.) and changing psychological symptoms (Lau et al., 2006).

Both subscales of the TMS showed an increase in their scores after participants had engaged in mindfulness practice, a result which provided support for TMS construct validity. Most studies supporting TMS construct validity and distinguishing between state and trait scales used classical test theory (CTT) methods. Notably, Medvedev et al. (2017a) employed a more reliable statistical technique to carry out a study to investigate whether the TMS measured state or trait mindfulness. Assessing mindfulness using the TMS, the authors applied generalisability theory (G-theory) to analyse the data of the repeated-measure design with a sample of 55 participants. The results showed that the variance attributed to a state component of mindfulness was significantly greater than that associated with trait mindfulness, a result that was consistent with the core purpose of the TMS measure (Medvedev et al., 2017a). However, compared with using other scales, assessors require training to use the TMS, due to its specialised vocabulary (Wray, 2004).

The Five Facet Mindfulness Questionnaire

The FFMQ (Baer et al., 2006; 2008) is made up of 39 items that assess aspects of mindfulness, grouped into five facets or subscales: “observe”, “describe”, “act with awareness”, “nonjudge”, and “nonreact” (see Appendix B). According to Baer et al. (2006; 2008), the first facet is “observe”, which is defined as the ability to notice or focus on experiences both internal and external, such as a smell, sight, or sound, or any sensation, cognition or emotion. The facet “describe” refers to identifying internal experiences using words. “Acting with awareness” is defined as focusing on the activities of the moment and can be distinguished from automatic behaviour and focusing attention elsewhere. The subscale “nonjudge” refers to adopting non-judgemental attitudes toward feelings and thoughts. The “nonreact” subscale indicates the ability to let feelings and thoughts come and go without reacting to them. Each individual item employs a 5-point Likert scale, with options from 1 = “never or very rarely true” to 5 = “very often or always true”. There are 19 reversal coded items which perhaps reduce or prevent response bias (Tourangeau, Rips, & Rasinski, 2012) and increase acquiescence (Krosnick & Presser, 2010).

The FMI was the first mindfulness psychometric instrument proposed to quantitatively assess individual mindfulness (Buchheld et al., 2001) but it can only be used with people learning mindfulness meditation (Wray, 2004). The FFMQ was designed for application with a variety of users, because it was based on samples from participants with varying meditation experience (i.e., no experience, a little experience, medium experience, considerable experience, and a lot of experience) (Baer et al., 2006). Results from the studies using such samples indicated positive significant correlations between the FFMQ scales and meditation experience (Baer et al.). Many research studies have demonstrated that therapeutic changes take place through MBIs, such as MBCT and MBSR, by using the FFMQ to assess mindfulness at the pre- and post-treatment stages (e.g., Bränström, Kvillemo, Brandberg, & Moskowitz, 2010; Carmody & Baer, 2008; Nyklíček & Kuijpers, 2008; Vøllestad, Sivertsen, & Nielsen, 2011; McManus, Surawy, Muse, Vazquez-Montes, & Williams, 2012). In other words, FFMQ scores reflected participant mindfulness levels, so that the FFMQ can be used for diverse participants (e.g. meditators and non-meditators),

and for evaluating the effect of mindfulness treatment, or in research looking for the correlation between mindfulness and other variables or concepts.

The FFMQ is the most commonly used multidimensional psychometric instrument to assess dispositional mindfulness. To date, according to Google Scholar, the original FFMQ article has been cited 5008 times since it was published (Table 1). 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, Bravo, Roos, & Pearson, 2015; Coffey, Hartman, & Fredrickson, 2010; MacDonald & Baxter, 2017; Medvedev et al, 2017). In addition, its popularity may be attributable to the accessibility of the vocabulary used. The FFMQ can easily be used without any training necessary while the TMS and FMI require assessors to be trained to use them because of their specialised vocabulary. Moreover, unlike the CHIME, which was originally in German, the FFMQ was written in English from the start, so is more likely to receive attention from experts and be translated into other languages.

The FFMQ was also shown to be the best mindfulness assessment tool for the following reasons. Baer et al. (2004) first developed the KIMS but only 2 years later it was developed to form the FFMQ. The reason for developing the scale further may be that the “describe” subscale of the KIMS did not obtain low correlations with other subscales. The FFMQ achieved weak correlations between subscales, from -0.07 to 0.34, which supported the FFMQ’s multidimensionality (Baer et al., 2006; 2008). Christopher, Neuser, Michael, and Baitmangalkar (2012) confirmed the FFMQ five-factor model by using a sample of both meditating and non-meditating participants. Moreover, the “observe” subscale added to the FFMQ enabled it to capture more aspects of mindfulness (Baer et al., 2006).

Another important reason for considering the FFMQ to be the best tool is that it was originally constructed by combining 112 items from five different mindfulness scales: the FMI, the MAAS, the KIMS, the Cognitive Affective Mindfulness Scale (Hayes et al., 2004), and the Mindfulness Questionnaire (Chadwick, Hember, Mead, Lilley, & Dagnan 2005). These five mindfulness scales were the most common instruments at the time. Baer et al. (2006) selected five factors from the 112-item collection by applying principal axis

factor analysis with oblique rotation. The subscale categories “observe”, “describe”, “act with awareness”, and “nonjudge” were given the same labels as those used for the KIMS subscales. The additional component was termed “nonreact to inner experience” or, in shortened form, “nonreact” (Baer et al.). In this analysis, the extracting procedure first excluded items with factor loadings below 0.40, leaving 64 items representing the five facets of mindfulness. The process continued by excluding items showing the lowest scores on their factor loadings and various cross-loadings. Finally, confirmatory factor analysis showed that although both the five- and four-factor subscales had similarly acceptable fit indices in the student sample, the five-factor model proved to be a better fit in the meditator sample. Thus, the designers ended up with the final five FFMQ subscales (Baer et al.).

The five FFMQ subscales achieved good internal consistency with Cronbach’s alphas, ranging from 0.67 to 0.93 (Park et al., 2013). The validity of the FFMQ construct was supported by positive correlations between the scores of the total FFMQ and its individual subscales and the measures of well-being, emotional intelligence, openness and self-compassion; and by negative correlations between the scores for neuroticism, alexithymia, anxiety, depression, and dissociation, and the FFMQ scores (Baer et al., 2006; Cash & Whittingham, 2010; Fisak & von Lehe, 2012). Baer et al. also noted that three FFMQ subscales (“nonjudge”, “act with awareness”, and “nonreact”) were valid predictors of psychological symptoms.

Furthermore, although the FFMQ is cited less frequently than the MAAS in both Google scholar and Web of Science, the FFMQ construct is more comprehensive than that of the MAAS in terms of dimensionality and items included because the FFMQ was designed to feature the most useful items from the MAAS in its “act with awareness” facet. The only available instrument which has more subscales than the FFMQ is the CHIME. The eight-factor CHIME includes 37 items while the five-factor FFMQ has 39. Normally, questionnaires with more subscales tend to feature a higher total number of items (Krägeloh et al., 2019). For this reason, the comprehensiveness of the CHIME rather than the FFMQ would be more likely to be questioned.

Finally, an additional reason for the pre-eminence of the FFMQ is the number of systematic short versions that have been developed. The questionnaire has received significant attention through the development of shortened versions suitable for quick assessments or to avoid imposing the burden of the full FFMQ on participants. A number of shortened versions of the FFMQ have been developed using the CTT approach (e.g., Baer, Carmody, & Hunsinger, 2012; Gu et al., 2016; Bohlmeijer et al., 2011). However, such shortened versions have been unable to address the limitations of ordinal scales, such as limited precision and compatibility with parametric statistics (Allen & Yen, 1979; Stucki, Daltroy, Katz, Johannesson, & Liang, 1996). To address these problems, Medvedev et al. (2018) carried out a study to examine and compare the current short versions of the FFMQ using Rasch analysis, and proposed an 18-item FFMQ version (FFMQ-18). All short versions of the FFMQ contribute to the comprehensiveness of the FFMQ compared with other mindfulness scales available.

State and Trait Distinction in the Measurement of Mindfulness

All the above reasons show that the FFMQ is the best psychometric instrument to date for measuring mindfulness. However, the temporal reliability of the FFMQ and its subscales was not rigorously tested with an appropriate statistical method. The ability of the FFMQ to differentiate trait from state was not examined carefully using appropriate methodology. Table 1 only reports test-retest reliability for the MAAS, the KIM, and the CHIME. Moreover, test-retest reliability is currently the only psychometric criterion employed to distinguish between state and trait aspects of an instrument (Arterberry, Martens, Cadigan, & Rohrer, 2014).

Nonetheless, mindfulness can be applied in a variety of contexts and growing evidence shows that mindfulness practice causes both state and trait changes (Tang, Hölzel, & Posner, 2015). As indicated above, more instruments were designed to measure trait than state mindfulness (Bergomi et al., 2013). As can be seen in Table 1, the test-retest reliability scores above 0.70 of the MAAS, CHIME and KIMS (except for the latter's "observe" subscale) demonstrate that all reliably measure trait (Ramanaiah, Franzen, & Schill, 1983; Spielberger, Gorsuch, & Lushene, 1970; Spielberger, 1999). In contrast, the

SMS test-retest score is below 0.70, indicating that the SMS is a state measure of mindfulness (Tanay & Bernstein, 2013).

A clear differentiation of trait and state mindfulness is needed because it helps to monitor the changes in both these aspects over time in order to provide the most suitable treatment at the appropriate time as well as to improve the methodology of mindfulness research (Paterson et al., 2017). For example, if a neurophysiological study cannot accurately separate trait from state mindfulness or vice versa, the selection of participants based on their mindfulness experience will be compromised because the mindfulness levels of practitioners cannot be measured clearly and reliably. Moreover, if trait mindfulness changes can be identified after an MBI, such an intervention will demonstrate its efficacy because any MBI tends to target long-lasting or trait changes, which may decrease the likelihood of relapse (Medvedev et al., 2017a). Where a state is involved, a change in mindfulness can reveal that an immediate condition, such as an occasion or short-term experience, caused the change, which will suggest a particular factor causing the effect.

A trait is described as an individual's more or less stable characteristic while a state is a characteristic feature exhibited in a given situation or point in time (Hamaker, Nesselroade, & Molenaar, 2007; Spielberger et al., 1970). For instance, a student who mostly experiences no stress (trait) may begin to feeling worried and stressed when an exam is imminent (state). Basically, a state is reflected by the interaction between a particular person and a specific occasion and its reflection in the person is the unique way that individual adapts to the present moment and environment (i.e., here and now) (Buss, 1989; Epstein, 1984). All aspects, including trait and state and the interactions between them, must be considered in order to understand the dynamic and enduring ways an individual functions (Buss, 1989; Epstein, 1984).

However, the separation of trait from state, which is based on the test-retest reliability of one correlation between test scores at two different times, points or occasions, tends to be less clear when measuring a state in comparison to a trait (Medvedev et al., 2017a). Thus, identifying and comparing the contribution of variance components of person (trait) and occasion (state) and their interaction (person x occasion) in a specific

measure will produce a more accurate estimate, whether a scale measures a state or a trait. According to Bloch and Norman (2012), it can be argued that G-theory is the most suitable method for making a clear separation of trait from state. The study by Medvedev et al. is a good example of applying G-theory to investigate the distinctive ability of differentiating between these aspects (state versus trait) in a mindfulness assessment tool (TMS).. Nevertheless, the number of studies using this method in the field of mindfulness is still limited.

Conclusion

Mindfulness has been measured in different ways in neurophysiological and psychological studies. Growing evidence from mindfulness research suggests that the practice of mindfulness leads to neurophysiological, cognitive and psychological changes. Such changes require accurate measurement of mindfulness and its related variables. However, neurophysiological studies cannot measure individual mindfulness levels, nor distinguish between state and trait mindfulness. Consequently, mindfulness assessment instruments or self-reporting measures of mindfulness are the most commonly used methods for assessing mindfulness in research studies. The MAAS is the best-validated mindfulness measure with satisfactory psychometric properties but it does not assess the most essential component of mindfulness, a “non-judgemental attitude”.

The CHIME is a very promising tool to measure multidimensional mindfulness because it combines all currently relevant mindfulness facets from available measurements, but its English version is unavailable to date and more research is also necessary to confirm the construct validity of the CHIME. The FFMQ is the most cited multidimensional assessment tool and includes all the most relevant MAAS items in its subscale of “acting with awareness”. Furthermore, the FFMQ was composed in order to address the limitations of other scales. These reasons make the FFMQ the best tool for mindfulness assessment. However, the temporal reliability and ability of the FFMQ to separate trait from state were not rigorously examined using appropriate methodology.

Chapter 3 Measurement Theories: G-theory and CTT

Introduction

Measurement theories, also known as statistical theories of measurement scores, have been developed for over 100 years. Spearman (1904) proposed the “true score” model, which stated that an observed score obtained in a test is the sum of a true score (i.e., an error-free score) and an error score. Subsequently, many developments have taken place in this field over the last several decades. Gulliksen (1950) distilled all developments and suggested the classical theory of measurement or CTT which included a system of statistical theorems and equations. Subsequently, Cronbach, Rajaratnam and Gleser (1963) established the G-theory, which does not require large sample sizes and is more robust against the violation of parametric assumptions (Suen & Lei, 2007). CTT and G-theory share certain foundational commonalities and some scholars consider both of them to be “true score models” (Vispoel, Morris, & Kilinc, 2018). However, there are many differences between the two theories. The purpose of this chapter is to discuss CTT and G-theory as statistical models and approaches, as well as their ability to separate trait from state in a psychometric assessment, particularly in mindfulness assessment tools.

Classical Test Theory

CTT was the most important theory in the last century and aimed to investigate and improve the reliability of psychometric instruments. The foundations of CTT were established by well-known psychologists and psychometricians, such as C. Spearman, J. Cronbach, R. B. Cattell, L. Guttman, L. L. Thurstone, and J. Loewinger (Lord & Novick, 1968; Cohen & Swerdlik, 2010). CTT postulates were based on the idea that an observed score (X) on a test is the sum of a true score (T) and an error score (E), values expressed by the following formula:

$$X = T + E \quad (1)$$

Assuming that both true scores (T) and error scores (E) are independent or unrelated to each other, the variance of observed scores (X) can be represented by the equation:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \quad (2)$$

The reliability coefficient is the most important concept in CTT and is the degree of true variance within the overall observed variance, as follows:

$$\text{reliability coefficient} = P_{xx} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2} \quad (3)$$

The reliability coefficient is estimated indirectly using different methods based on various assumptions. The parallel test assumptions are the most common approach used to estimate the reliability coefficient (Suen & Lei, 2007). According to these assumptions, the same test made at two different times will meet a restrictive set of statistical conditions. The correlation coefficient value between the observed scores of these two tests is then the reliability coefficient value. There are practical strategies based on these assumptions, such as the equivalent forms method, the test-retest method, and the internal consistency method (Suen & Lei).

Another approach is to estimate the reliability coefficient on the basis of “a set of somewhat less restrictive assumptions, called the essentially τ -equivalent assumptions” (Suen & Lei, 2007, p.3). One of the most popular methods based on these assumptions is to estimate the internal reliability coefficient by means of Cronbach’s alpha (Cronbach's α):

$$\text{Cronbach's } \alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_X^2} \right) \quad (4)$$

where k is the number of items in the scale, and σ_i^2 is the variance of the i^{th} item in the scale.

The standard error of measurement can be deduced after the reliability coefficient is computed, as below:

$$\text{Standard error of measurement} = \sigma_X \sqrt{1 - P_{xx}} \quad (5)$$

The standard error of measurement is a measure of the average instability of observed scores used instead of true scores. This value can be employed in forming confidence intervals for an observed score, indicating the range within which true scores might

decrease. The larger the reliability coefficient, the more stability is expected of the observed score if a retest takes place, and vice versa (Vispoel et al., 2018; Suen & Lei).

Several benefits can be obtained through the application of CTT in developing and evaluating a scale. First, CTT uses relatively simple mathematical procedures to analyse collective data and straightforward model parameter estimations (Bichi, 2016; Hambleton and Jones, 2005). Secondly, it is clear that traditional testing procedures easily meet the assumptions of the CTT method (Kline, 2013). In addition, the main advantage of CTT is the statistical software available (e.g., IBM's SPSS), which is popular and easy to use. However, even though CTT appears to be a simple, useful theory to apply in evaluating the psychometric properties of a measurement, the simplicity of this model hides a number of disadvantages (Suen & Lei, 2007).

CTT considers error variance as a single factor and postulates that any measurement consists of true variance and error variance (Allen & Yen, 1979) even though, in fact, there are multiple sources of error that potentially influence the accuracy of measurement and may affect the observed scores (Bloch & Norman, 2012). Many factors, such as personal error, item error, occasion (e.g., environmental influences) error, and their interactions may contribute to measurement error. Thus, if the evaluation of measurement scales relies solely on CTT results, conclusions about such scales used to assess people will be inaccurate (Bloch & Norman, 2012).

Generalisability Theory (G-theory)

Cronbach et al. (1963) developed G-theory, which is a more advanced statistical method in comparison to traditional CTT methods for evaluating the reliability of psychometric measurements (e.g., rating scales, performance tests). G-theory has the ability to evaluate specific sources of measurement error, and is also capable of estimating the generalisability of assessment scores to all possible circumstances, using data collected from a particular testing situation (Cronbach et al.). In other words, G-theory considers and estimates all possible sources of error variance affecting the true score of interest (e.g., a score from a mindfulness instrument) because various sources of error are more likely to

occur in complex natural environments and potentially influence the accuracy of measurement (Bloch & Norman, 2002).

For example, generalisability analysis (G-analysis) will consider interactions between persons and different factors. Such alternative factors may be methodological (e.g., instrument items) or situational (e.g., occasions) factors that may independently (or via interactions) contribute to the measurement error (Brenan 1992; 2001). In summary, G-theory demonstrates a more advanced statistical methodology than CTT because G-theory closely examines all possible influences on reliability (including their interactions) simultaneously, thus improving the methodology and precision of a psychometric assessment, whereas CTT considers only one aspect of reliability (e.g., test-retest or internal consistency) at a time.

Conceptually, G-theory is an extension of classical test theory but with many facets (Suen & Lei, 2007). In CTT, an observed test score X comprises the true score T of a person (participant) on the test and a measurement error E , and the value of E is essentially the deviation of X from T , or when E has an effect on X . In G-theory, a single score, such as an item score, subscale score, or total scale score, is understood as a sample from a multitude of legitimate observations for “persons”, the objects of measurement (Vispoel et al., 2018; Suen & Lei; Brenan, 1992; 2001). Any possible source of measurement error defined in G-theory is called a facet. Such facets are analogous to factors in traditional ANOVA. A measurement which has only one potential source of error, for example, an item, refers to single-facet design because an item is the single facet of interest. Thus, the observed score of a person for a particular item can be expressed as follows (Vispoel et al., 2018; Suen & Lei):

$$X = \mu + \mu_p + \mu_i + \mu_{pi} \quad (6)$$

The universe score μ or grand mean is a constant of an experimental condition that represents the mean of the true item response values across all persons (participants) on all items. The deviation of a person’s average item response from the universe score ($\mu_p = \mu_p - \mu$) represents the person effect. The other deviations from this universe score are

$(\mu_i \sim = \mu_i - \mu)$ representing the item effect, and $(\mu_{pi} \sim = X - \mu_p - \mu_i + \mu)$ is the residual random error or the effect of the person's X item interaction (Vispoel et al., 2018; Suen & Lei, 2007).

Then the variance of observed scores X can be represented by:

$$\sigma_X^2 = \sigma_p^2 + \sigma_i^2 + \sigma_{pi}^2 \quad (7)$$

Unlike CTT, which attempts to estimate the reliability coefficients indirectly under parallel test conditions or essentially τ -equivalent assumptions, G-theory uses a direct approach to estimate reliability coefficients using the common analysis of variance (Suen & Lei). Thus, appropriate estimates of reliability coefficients are composed on the basis of variance components computed for facets with their interactions serving as building blocks (Vispoel et al.; Suen & Lei)

G-theory estimates the relative contribution of potential sources to the overall error of a measurement, also known as “noise” (Brennan, 2001). Each contribution that is identified can be referred to as an intra-class correlation coefficient (ICC). ICC is similar to other reliability coefficients, ranging from 0 to 1 (Bloch & Norman, 2012). For instance, an ICC may reflect the ability of the mindfulness measurement to differentiate among participants through the amount of variance between its scores, and can be expressed as in Equation (3). Here, the value of ICC depends on an instrument's discriminative ability among participants and the degree of noise resulting from other contributors or factors. As a coefficient of reliability, ICC indicates the ratio between the amount of true score variance (i.e. the main variable) and the total amount of observed variance (Brennan, Bloch & Norman). ICC was first employed in CTT as a common indicator for the quality of a measurement and is represented by a formula (10) drawing on the notion of “signal-to-noise ratio” (SNR) (Fisher, 1925). SNR is the square of the effect size (9) and the effect size is a ratio between consistent change in the variable X (ΔX) and the total variance (σ), as follows (Bloch & Norman):

$$ES = \frac{\Delta X}{\sigma} \quad (8)$$

$$SNR = \frac{\Delta X^2}{\sigma^2} \quad (9)$$

$$ICC = \frac{SNR}{1+SNR} \quad (10)$$

Interpreting these equations, if the amount of variance in a measured true score (signal) compared to errors (noise) is sufficiently large, the likelihood of detecting the changes becomes greater. This means that an ICC close to 1 indicates a significant difference between signal and a comparatively negligible level of noise, meaning that it is all signal and no noise. On the other hand, an ICC close to 0 indicates that it is mainly noise or that a large amount of error exists in the data. ICC refers to a general G-coefficient (G_p) in G-theory and is expressed in a way similar to CTT format. This is the ratio of the observed (true) variance of a person (the object of measurement) across facets (σ_p^2) and the total variance of universe scores, consisting of the variance of the observed (true) score and the total variance of error (σ_{error}^2) (Brennan, 1992; 2001; Shavelson, Webb, & Rowley, 1989):

$$G_p = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{\text{error}}^2} \quad (11)$$

Here, p represents a “person” effect since “person” is typically a facet of differentiation, which is equivalent to an object of measurement in psychometric instruments. A G-coefficient not only indicates the main variable (e.g., state mindfulness, trait mindfulness) but can also represent factors influencing error variance in a research design assessing variability due to these contributions (Bloch & Norman, 2012). Such a G-coefficient presents the ability to generalise influences attributable to particular factors across possible contexts and situations.

G-theory distinguishes between absolute and relative G-coefficients for the object of measurement (person). The relative model of measurement, referred to as test scores, is interpreted in a norm-referenced manner, in which the score of a person is compared to the scores of others (Vispoel et al., 2018; Suen & Lei, 2007). Using single-facet (person by item) as an example, the relative G-coefficient (G_r) can be computed as below:

$$G_r = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\delta^2} \quad (12); \text{ here } \sigma_\delta^2 = \sigma_p^2 + \frac{\sigma_{pi}^2}{n_i} \text{ is the relative error variance, } n_i$$

= number of items

Alternatively, the absolute model of measurement is based on the test scores, which are interpreted in a criterion-referenced manner in which the score of a person is compared with an absolute standard generally agreed upon. The reliability coefficient in this model is called the absolute G-coefficient (G_a) and can be equated with the phi (Φ) coefficient obtained after applying Whimbey's correction (Whimbey, Vaughan, & Tatsuoka, 1967). Whimbey's correction is the expression of $(K-1)/K$, where K is the size of the facet universe (i.e., facet item) in G-design. G_a accounts for an absolute error variance that includes facets (e.g., item variance), which may indirectly "influence an absolute measure" (Cardinet, Pini, & Johnson, 2010):

$$G_a \simeq \Phi = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2} \quad (13); \text{ here } \sigma_\Delta^2 = \sigma_p^2 + \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_i^2}{n_i} \text{ is the absolute error variance}$$

A more complicated measurement, in which there are two potential sources of error, is called two-facet design. For instance, occasions and items are facets of interest. The equations below present estimates of G-theory analysis for this design:

$$X = \mu + \mu_{\tilde{p}} + \mu_{\tilde{i}} + \mu_{\tilde{o}} + \mu_{\tilde{pi}} + \mu_{\tilde{po}} + \mu_{\tilde{io}} + \mu_{\tilde{pio}} \quad (14)$$

where X is a person's observed score on a particular item across occasions, $\mu_{\tilde{o}}$ represents the occasion effect, $\mu_{\tilde{po}}$ represents the effect of person x occasion interaction, $\mu_{\tilde{io}}$ represents the effect of item x occasion interaction, and $\mu_{\tilde{pio}}$ represents the effect of person x item x occasion interaction.

The variance of observed scores X can be represented by:

$$\sigma_X^2 = \sigma_p^2 + \sigma_i^2 + \sigma_o^2 + \sigma_{pi}^2 + \sigma_{po}^2 + \sigma_{io}^2 + \sigma_{pio}^2 \quad (15)$$

$$G_r = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\delta^2} \quad (16); \text{ here } \sigma_\delta^2 = \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{po}^2}{n_o} + \frac{\sigma_{pio}^2}{n_i n_o} \text{ is the relative error variance, } n_o$$

= number of occasions

$$G_a \simeq \Phi = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2} \quad (17);$$

here $\sigma_\Delta^2 = \frac{\sigma_o^2}{n_o} + \frac{\sigma_i^2}{n_i} + \frac{\sigma_{pi}^2}{n_i} + \frac{\sigma_{po}^2}{n_o} + \frac{\sigma_{io}^2}{n_i n_o} + \frac{\sigma_{pio}^2}{n_i n_o}$ is the absolute error variance.

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

By comparing the equations of single- and two- facet G-theory design, it is clear that the greater the quantity of possible sources of measurement error that are considered, the more accurately the reliability coefficients can be estimated. However, an increase in potential sources of measurement error leads to more complex formulations for reliability coefficients and standard error calculations. For this reason, G-theory, which is the most powerful method for estimating measurement reliability, is still underused and should be applied more widely to establish the true reliability of measurement in research.

Application of G-theory to the State Versus Trait Distinction

A 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 an assessment could be affected by evaluating a person's characteristics while not accounting for temporary changes (e.g., mood). This might lead to inappropriate conclusions and both the reliability and validity of a psychometric instrument may be compromised due to confusing a trait with a state (Medvedev et al., 2017a). In other words, the inability to distinguish reliably between state and trait aspects in psychological measurement may lead to an inaccurate evaluation of an intervention's effectiveness over time. There should be a clear distinction between the state and trait characteristics of a

person in any psychometric measure. This requires identification and consideration of the relevant sources of error variance using the most appropriate statistical method, such as G-theory (Paterson et al., 2017; Bloch & Norman 2012).

In the field of mindfulness, conducting any MBI aims at enduring or trait changes because if a treatment leads only to state changes, relapse is inevitable. This is due to the fact that a state can change immediately following an assessment or after a short-term experience, whereas a trait is established over a longer term. Therefore, the accurate measurement of both trait and state aspects of mindfulness in both mindfulness research and treatments is necessary (Cahn & Polich 2006; Chiesa & Serretti 2010).

The traditional CTT approach to the state/trait distinction examines test-retest reliability coefficients to investigate the reliability of an instrument over time, which tends to be lower for a measure of state (e.g., <0.60), and higher for a trait measure (e.g., >0.70) (Ramanaiah, Franzen, & Schill, 1983; Spielberger et al., 1970; Spielberger, 1999). This method bases itself completely on total score correlations at two different times (i.e., Time 1 and Time 2) and does not consider variability at the individual item level or interactions between person, item, and occasion. A robust estimation of reliability requires consideration of the contributions made by the effects of scale, item, occasion, and person and their interactions with changes in the overall assessment score.

Similarly, the intraclass correlation coefficient (ICC) that can be used to determine reliability over time has limited accuracy because it does not account for the variability of individual items and interaction effects (Medvedev et al., 2017a; Bloch & Norman 2012). Take the scores of two items in the FFMQ “nonjudge” subscale, for example. If a participant scores 5 on item 35 (*I judge myself as good or bad*) and 1 on item 14 (*I believe my thoughts are abnormal or bad*) in the first assessment, and then 2 and 4 on items 35 and 14, respectively, in the second assessment after 2 weeks, the total score remains unchanged (i.e., 6). This implies full agreement between tests after 2 weeks, which may not reflect clinically important changes in specific aspects of mindfulness that predict psychological symptoms (e.g., a non-judgemental attitude).

In this simple example, therefore, the total scores at two points in time do not give an accurate estimate of reliability over time, nor a clear differentiation between subscales measuring state and trait components of mindfulness. Moreover, this estimation of reliability does not involve error variance due to potential sources such as subscale, item, occasion or their interactions with person (the object of measurement) (Medvedev et al., 2017a; Bloch & Norman, 2012).

The State and Trait Anxiety Inventory (STAI; Spielberger et al., 1970) is an example of a scale measuring both state and trait, validated using predominantly CTT approaches which can clarify the idea sketched above. The STAI is a self-report instrument comprising two subscales: the state subscale and the trait subscale. Each subscale contains 20 items to assess either state or trait anxiety. A general perception of the environment as dangerous was put forward as a means to measure anxiety as a trait, while the measurement of the experience of anxiety at a given time aims to measure only state anxiety.

Further examples of scales measuring state and trait which use a similar format to the STAI include the State-Trait Anger Expression Inventory-2 (Spielberger, 1999) and the Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988). All these example measures were proposed to provide instructions to help a participant clarify temporal aspects of responses. That is, a participant has to indicate “how they are feeling here and now”, in the subscale designed to measure a state, and how they feel “generally” in the trait subscale. In addition, the two subscales were correlated with each other with r ranging from 0.70 and 0.80. These values confirm the expectations of the relationship between a trait and a state in the STAI (Ramanaiah et al., 1983; Spielberger, 1999).

However, studies on the validation of the STAI reported overall high test-retest reliability coefficients, ranging from 0.78 to 0.94 for the trait subscale but for the state subscale, test-retest correlations ranged from 0.16 to 0.96 (Barnes et al., 2002; Spielberger, 1999). It was clear that test-retest reliability scores were unreliable because in the study by Barnes et al., the test-retest maximum scores for the STAI trait subscale were lower (0.94) compared to the maximum scores on the state subscale (0.96). Therefore, the application of CTT for distinguishing between two components in these psychometric assessments clearly

showed its limitations, overlooking important effects due to subscale, occasion, and item, and their interactions with the object of measurement (Bloch & Norman, 2012). In the evaluation of state and trait measurements, scores which are hypothesized to measure a trait are not expected to show much variation across occasions. Rather, the person x occasion interaction is naturally understood to measure a state (Epstein, 1984; Chaplin, John, & Goldberg, 1988).

Moreover, the structural equation modelling (SEM) approach, which is based on CTT, is currently the main method to explore and assess the variability of state and trait (Hamaker et al. 2007; Geiser et al. 2015; Kenny and Zautra 2001). SEM is a multivariate statistical method of analysis, and employs the combined technique of multiple regression analysis and factor analysis to examine the structural relationship between latent constructs and measured variables (Kaplan, 2008; Kline, 2011). A number of analytical models using SEM have investigated the differences in trait and state variability (Hamaker et al., 2007; Kenny & Zautra, 1995; Steyer, Ferring, & Schmitt, 1992). None of the SEM methods, however, was proposed to explicate all the relevant sources of variance (e.g., persons, individual items, and interactions between them) assisting in accurately distinguishing state and trait (Medvedev et al., 2017a). Consequently, such SEM approaches reveal limitations in their application in validating measures of temporal (state) and enduring (trait) components. The differences in variability require dedicated research into the factors that can influence state and trait aspects, including item, occasion, and person and their interactions, as well as the dynamicity (state) or stability (trait) of each individual item. In other words, if components can be quantitatively identified, changes in state and trait can be foreseen through an awareness of the changes in person and occasion, and this is a “true generalisability” (Medvedev et al., 2017a; Suen & Lei, 2007).

G-theory can be employed to evaluate the degree of variance indicated by the object of measurement (person) and facets (i.e., item and occasion), and also their interactions with each other (Brennan, 2001; Bloch & Norman 2012). While the variance which represents the effect of person x occasion interaction is directly reflected in the ‘stateness’ of a latent construct, the variance due to the person effect alone indicates a trait (Buss, 1989; Chaplin et al., 1988; Medvedev et al., 2017a). G-theory can be employed to conduct

a series of analyses for many aspects of a scale (i.e., the total scale, subscales and each individual item). In other words, it is possible to distinguish items which are representative of a state from items that are not sensitive to change across occasions. Nevertheless, no precise benchmarks of the ratios between state and trait components are currently established for state and trait measures (Medvedev et al., 2017).

A generalisability study, or G-study, which applies G-theory, produces an estimation of variance in the object of measurement (e.g., persons) and in facets (e.g., items, occasions). The variance components are computed on the basis of the observed score obtaining from the universe and all possible influencing sources. A state measurement (e.g., the total scale, each subscale and individual items) is expected to obtain a high variance component attributed to the person x occasion interaction and reflect a low level of generalisability across occasions (e.g., G-coefficient <0.70), in contrast with a reliable trait measurement, which should have a G-coefficient of 0.80 or more (Arterberry et al., 2014; Cardinet et al., 2009). Generally, traits can be distinguished from states by estimating the effect of the interaction with situational factors (e.g., occasions) based on relevant variance components (Medvedev et al., 2017a; Paterson et al., 2017). Moreover, a more precise distinction between the state and trait aspects of a measure can be made through a decision study (D-study). D-study involves experimenting with measurement designs by changing facet levels in an attempt to reduce measurement error and enhance reliability (Shavelson et al., 1989; Brennan, 2001). Data from D-study reveal the distinctive aspect of each individual item (e.g., items measuring state in a trait instrument). This makes it possible to modify an instrument or optimize the measurement.

G-theory has been previously used to estimate the reliability of psychometric assessments measuring trait and state (Arterberry et al., 2014; Medvedev et al., 2017a; Paterson et al., 2017; Medvedev et al., 2018). First, a study was conducted by Arterberry et al., which utilised G-theory to examine the reliability of the Big Five Personality Inventory score (BFI; John, Donahue, & Kentle, 1991). This study included 264 participants drawn from a large public university in the American Midwest, who completed the BFI at three points in time – the baseline, after one month, and after 6 months. The results indicated that BFI score reliability was generally acceptable (G-coefficients >0.70). In other words, the

finding of this study not only indicated that the BFI was a trait measurement of personality, but also demonstrated the advantages of G-analysis to evaluate scores (Arterberry et al., 2014). Another study was carried out by Medvedev et al. (2017a), which used G-theory method to differentiate state from trait aspects in a state mindfulness measure (i.e., TMS). In their study with a sample of 55 participants, the authors examined whether the TMS measured state or trait mindfulness. The results showed that the variance attributed to a state component of mindfulness was significantly larger than that attributed to trait mindfulness, in this way providing evidence confirming that the TMS measures state mindfulness.

Moreover, Medvedev, Theadom, Barker-Collo, and Feigin (2018a) also applied G-theory to differentiate between enduring (trait) and dynamic (state) concussion symptoms in the Rivermead Post Concussion Symptoms Questionnaire (RPQ). Participants included 145 patients with a traumatic brain injury, who were assessed by the RPQ at three different times (1, 6, and 12 months after the injury). The research found that the RPQ demonstrated strong reliability in assessing enduring post-concussion symptoms. In other words, the RPQ is a means to trait measurement in assessing post-concussion symptoms. Paterson et al. (2017) also applied G-theory to distinguish temporary (state) and enduring (trait) aspects of depression in the 10-item Children's Depression Inventory (CDI-10; Kovacs, 1985), which is a widely used depression screening measure. Their sample was made up of 668 children, who were New Zealand Pacific Islanders aged 9, 11 and 14 years, and were assessed by the CDI-10. The authors found that the CDI-10 yielded a reliable measure of both state and trait aspects of depression in children because it obtained acceptable generalisability across occasions ($G = 0.79$). About two thirds of variance in the total scores were attributed to more enduring aspects of depression and one third to temporary aspects.

Nonetheless, literature reviews showed that there were no acceptable benchmarks that could be used as criteria for determining the relative proportions of state and trait variance components in a measure of either state or trait. Accordingly, in the G-study by Medvedev et al. (2017a), the authors developed the formulae to compute the state component index (SCI) and trait component index (TCI), which highlight the proportion of

variance that can be attributed to a temporal (state) and an enduring (trait) component in a measure, as follows:

$$SCI = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_t^2} \quad (18)$$

$$TCI = \frac{\sigma_t^2}{\sigma_s^2 + \sigma_t^2} \quad (19)$$

In this equation, the state variance component ($\sigma_s^2 = \sigma_{po}^2$) is known as the noise or error variance of the person x occasion interaction that affects the stability of measures (i.e., trait scores). The ratio equation (18) indicates the proportion of state to trait in both the state variance component and the trait (persons) variance component ($\sigma_t^2 = \sigma_p^2$). Essentially, the trait (persons) is the fundamental component of state variance. For the accurate measurement of the SCI, the calculation used an absolute variance value resulting from the person x occasion interaction derived from G-analysis. The SCI was developed based on the logical nature of G-theory. It was interpreted as SCI equals 1.00, meaning either that there is no trait component in the measure or that such a measure has only an individual state. However, this circumstance is less likely to occur because a trait serves as a basic indicator of a state (Buss 1989; Epstein 1984). An SCI of 0.50 can be interpreted to mean that the state component is equal to the trait component and a scale cannot be designated as measuring either a state or a trait. However, if an SCI is greater than 0.60, the measure can be regarded as having the characteristic of a state. In this case, higher scores correspond to the higher ability of a measurement to capture changes of state. Similarly, the use of the same metric for TCI can validate a measure of trait (Equation 19). A TCI close to 1.00 indicates that a measure can be considered indicative of a trait or that the scores of an instrument are consistent over time, while a TCI below 0.50 indicates that a measure is sensitive to change across occasions.

Although previous studies have demonstrated that the benefits of G-theory method in evaluating psychometric instruments outweigh its disadvantages, it is clear that only a few G-studies have implemented this novel method to distinguish between temporal and enduring aspects in a measure. Particularly, the TMS was the only mindfulness measure scrutinised using G-theory to test its ability to differentiate between trait and state

mindfulness. Further investigation with this robust psychometric method is necessary to distinguish between state and trait variance components (or vice versa) in other mindfulness scales, especially the most commonly used multidimensional assessment tool, the FFMQ.

Summary and Aim of the Present Work

The clear differentiation of trait and state in a measure may help to monitor the changes in both these aspects, and is necessary to develop and implement the most appropriate intervention. To date, the commonly used method to evaluate the temporal reliability of a measure is CTT, which is based only on test-retest correlational reliability coefficients. Importantly, the traditional CTT approach to the state-trait distinction also uses these reliability values, which tend to be low for a measure of state and high for a measure of trait. CTT method considers all potential sources of error as a single factor and does not account for variability at the level of individual items or interactions between person, item, and occasion. Thus, CTT methodology may not provide the precision necessary for the state-trait distinction, due to its limited scope. In contrast, G-theory method considers all possible sources of variance which contribute to errors in measurement. Such sources include person and other factors, including methodology (e.g., scale items) and situation (e.g., time of day, occasion), and each of them may contribute to measurement error either independently or via interactions. Therefore, G-theory is the most suitable statistical approach to distinguish state from trait aspects of a person and to rigorously examine the reliability of any psychometric instrument. The present work applies G-theory to evaluate the FFMQ's psychometric properties in order to distinguish between state and trait components of mindfulness and to establish the temporal reliability of the instrument.

Chapter 4 Generalisability Study Method and Results

The Purpose of the Study

The current study aims to utilise G-theory to investigate the reliability of the FFMQ and its short 18-item version (FFMQ-18) over time, differentiate state and trait components of mindfulness items and subscales, and identify sources of error that may impinge on measurement. The research employs a repeated-measures longitudinal design with participants assessed on three occasions, separated by equal 2-week intervals. The application of G-theory involved two parts, G-study and D-study. G-study investigated the overall generalisability of the FFMQ and its subscale scores and also evaluated sources of error variance in the measurement. D-study was subsequently carried out to assess the psychometric properties of individual items and to manipulate the measurement design to optimize the measurement (Shavelson et al., 1989; Cardinet et al., 2010).

This study was approved by the Psychology Research and Ethics Committee, University of Waikato, ethics approval application number 19:22 (see Appendix A2).

Participants

The sample included 83 university students who participated in the study on a voluntary basis and did not receive any payment or academic credit for their involvement. The sample size satisfied the requirements for reliability studies of this type of research (Shoukri, Asyali, & Donner, 2004) and was also sufficient for generalisability analysis, due to the similarity of G-coefficients and ICC reliability coefficients (Bloch & Norman, 2012). All 83 participants were New Zealand university students, including 22 men (26.5%) and 61 women (73.5%). From the total sample, ten participants (12%) engaged in regular meditation practice. Participants' ages ranged from 18 to 47 years, with a mean of 21.34 (standard deviation =5.83). Ethnic groups included 57% Caucasian, 11% Māori, 10% Pasifika, 6% Asian and 17% other.

Procedure

Participants completed the FFMQ items in class before the lecture or during a break. They were instructed to return the completed forms to the researcher, place them in a locked box at their faculty office, or use a self-addressed pre-paid envelope to mail their completed forms to the researcher's university address. Each participant was asked to fill out the same questionnaire on three occasions at 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 filled out by the same participant on three occasions. This research was not expected to involve any risk, discomfort or harm, and participants were informed about the nature of the study.

Measures

The FFMQ (Baer et al. 2006) is made up of 39 items that assess aspects of mindfulness grouped into five subscales: "Act with Awareness, Describe, Nonjudge, Nonreact, and Observe". Each individual item employs a 5-point Likert scale with options ranging from 1 = "Never or very rarely true" to 5 = "Very often or always true". There were 19 items that required reverse coding before data analysis could be conducted. After reverse coding, the total score and individual subscale scores were computed by adding together responses to the relevant items.

The FFMQ-18 has 18 items and is grouped according to the same format of five facets as the total FFMQ. This short version was proposed by Medvedev et al. (2018) after applying the Rasch method to evaluate the psychometric properties of various short versions of the FFMQ. In the current study, the FFMQ-18 scores were extracted from the total FFMQ for each participant.

Data analysis

IBM SPSS Statistics 25 software was used to compute means, Cronbach's alpha, test-retest coefficients, standard deviation (SD), and ICC for the FFMQ, FFMQ-18, and

individual subscales of the FFMQ. Missing data comprised merely 0.04%, a negligible amount, replaced using mean imputation (Huisman, 2000).

EduG 6.1-e software (Swiss Society for Research in Education Working Group, 2006) was used to conduct generalisability analyses (see Appendix C1-51). Both the D-study and G-study used two-facet design incorporating a random effects model: person (P) by item (I) by occasion (O), typically formulated as “P x I x O” (Medvedev et al., 2017a). Whereas the P and O facets are infinite, the I facet is fixed because the same set of items was used across all assessments using the FFMQ. All error variances are counted as 100% in a G-study, after controlling for person variance (P), which reflects true differences between persons. Person was the object of measurement (differentiation facet) and was not a source of error, while I and O were instrumentation facets (Vispoel et al., 2018; Cardinet et al., 2010). All estimations for the G-study were calculated following the equations in Chapter 3, including grand mean, facet means effects, variance components for each facet and their interactions, relative G-coefficient (G_r) and absolute G-coefficient (G_a).

This study computed SCI and TCI, which indicate the degree of variance attributed to a state and a trait component in a score (i.e., total FFMQ score, FFMQ-18 score, individual subscale score, individual item score), by using the formulae which were developed by Medvedev et al. (2017a) and described in the previous chapter. In D-study, variance components were obtained for each individual item and SCI values were calculated. In addition, a series of G-analyses was conducted with a combination of the most dynamic items, attempting to explore the possibility of producing a valid state scale.

Results

Descriptive statistics for the 39-item FFMQ, its subscales, and the FFMQ-18 on three occasions are presented in Table 2. As measured by Cronbach's alpha, the internal consistency of the total FFMQ over three occasions was good and 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 an ICC of 0.83, which is a more robust measure of temporal reliability. Overall, these reliability values were higher than those of the FFMQ-18 and the individual subscales of the FFMQ (Table 2).

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 to other subscales. Overall, all assessed FFMQ scales and subscales showed acceptable internal consistency and the temporal reliability expected for a trait measure. An exception was the “nonreact” subscale, which displayed the lowest Cronbach’s alpha value of 0.69 on Occasion 1 and the lowest test-retest value on Occasion 3 (0.64).

Table 2. 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 x 3 occasions).

Scale/Assessment	Occasion 1	Occasion 2	Occasion 3	ICC (95% CI)
<i>Observe</i>				
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 (r^a)	--	0.74	0.74	0.77(0.69-0.84)
<i>Describe</i>				
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 (r^a)	--	0.90	0.83	0.86(0.81-0.90)
<i>Act with Awareness</i>				
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 (r^a)	--	0.87	0.77	0.81(0.74-0.87)
<i>Nonjudge</i>				
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 (r^a)	--	0.81	0.87	0.85(0.79-0.90)
<i>Nonreact</i>				
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 (r^a)	--	0.75	0.64	0.71(0.61-0.79)
<i>FFMQ-18</i>				
Mean (SD)	57.68 (8.46)	57.33 (9.51)	57.71 (9.90)	
Cronbach's alpha	0.79	0.84	0.86	
Test-retest (r^a)	--	0.87	0.80	0.82(0.74-0.87)
<i>FFMQ Total</i>				
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 (r^a)	--	0.92	0.83	0.86(0.80-0.91)

Note. Mean differences are not significant compared to Occasion 1 (Bonferroni corrected); ^a Test-retest bivariate correlations between Occasion 1 and subsequent Occasions 2 and 3; CI=Confidence interval.

G-Study

Table 3 presents the variance components that can be attributed to person (P), item (I), and occasion (O), and their interactions, typically expressed as “P x I, P x O, I x O, P x I x O” (Medvedev et al., 2017a), together with generalisability coefficients and state and trait component indices for the FFMQ, its five subscales, and the FFMQ-18. All related EduG analyses outputs, including observation and estimation designs, ANOVA and G-study tables are included in Appendices C1-7. The best generalisability and reliability scores across occasions and persons were found for the total FFMQ, with both relative and absolute G coefficients (G_r and G_a) of 0.89 and the main source of error variance, due to P x O interaction, that accounted for 98.2% of total error. Slightly lower but still acceptable G_r and G_a values of 0.76 and 0.75, respectively, were observed for the FFMQ-18, with measurement error mainly explained by P x O and P x O x I interactions, which together explain 80% of the error variance. 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 no subscales met expectations for a reliable trait measure (Shavelson et al. 1989). The SCI, reflecting the ability of a measure to reliably assess state changes, was below expectations for a valid state measure for all individual FFMQ subscales (all SCI < 0.40). Even though TCI values for all five FFMQ subscales were high, ranging from 0.64 (“nonreact”) to 0.89 (“observe”), all subscales were affected by measurement error, the result of interaction between person, item, and occasion. This resulted in the low reliability of all subscales (all G_r < 0.50), meaning that the individual FFMQ subscales cannot be considered as reliable measures of trait mindfulness.

Table 3. *G*-study estimates for the FFMQ and FFMQ-18 and five subscales of the FFMQ standard errors (SE), Coefficient *G* relative (Gr), Coefficient *G* absolute (Ga), Trait Component Index (TCI), State Component Index (SCI), grand mean (GM), and variance components (σ^2) in %, for the Person (P) \times Occasion (O) \times Item (I) design, including interactions ($n = 83$)

Facets	FFMQ Total		FFMQ-18		Observe		Describe		Act ^a		Nonjudge		Nonreact	
	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^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	0.003	3.8
PxI	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
PxO	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
IxO	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
PxIxO	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
SE	0.080		0.137		0.224		0.238		0.235		0.221		0.279	
Gr	0.89		0.76		0.33		0.40		0.44		0.42		0.29	
Ga	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	

Note. Grand mean=3.19. Numbers in bold signify acceptable reliability/generalisability coefficients.

^aAct with awareness.

D-Study

Individual item analysis was carried out to determine variance components for individual items by excluding all others. All related EduG analyses outputs, including observation and estimation designs, and *G*-study tables, are included in Appendices C8-46. The estimates for variance of person, occasion, and person-occasion interaction, together with computed SCI, are included in Table 4. As can be seen, 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) which presented a high SCI (≥ 0.60), reflecting high sensitivity for state changes over time. On the other hand, the remaining 14 items had an SCI between the benchmarks ($0.30 < \text{SCI} < 0.60$) and cannot be clearly classified as reflecting either trait or state because they measure both

aspects. This means that there are no items with low SCI (≤ 0.30) that are least sensitive to state changes and reflect predominantly trait mindfulness.

Interestingly, all individual subscales, with the exception of the “nonjudge” subscale, consist of items which reflect their readiness to change across occasions and, in addition, include those items which are unspecified as a state or trait item. The “nonjudge” subscale is the only facet in the instrument which contains all state items. Three subscales, including “observe”, “act with awareness”, and “nonreact”, have more state items and accordingly fewer items unspecified between state and trait, whereas the “describe” subscale consists of more unspecified items than items which are identified as reflecting a state.

Furthermore, a series of generalisability analyses was conducted by combining the most dynamic items with the highest SCI, because we expected that this would result in a reliable state measure. EduG outputs for these analyses, including observation and estimation designs and G-study tables, are included in Appendices C47-51. Table 5 shows D-study results, including reliability estimates and variance components that can be 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 5, a). In the b analyses (Table 5), 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 those remaining (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 level of acceptable generalisability 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., the SCI should be above 0.60 to be considered a state measure).

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

Table 4. Variance components of Person (P), Occasion (O) and P x O interaction, along with the state component index (SCI) for each item in the FFMQ (n = 83x3)

Items / Factors	P	O	PxO	SCI
<i>Observe</i>				
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
<i>Act with Awareness</i>				
38 <i>doing things without paying attention</i> ^R	0.01	0.00	0.30	0.95
13 I am easily distracted ^R	0.31	0.01	0.25	0.45
5 my mind wanders off and I'm easily distracted ^R	0.12	0.03	0.38	0.76
8 I don't pay attention to what I'm doing ^R	0.12	0.08	0.32	0.72
34 I do jobs or tasks automatically ^R	0.23	0.07	0.21	0.48
18 I find it difficult to stay focused ^R	0.06	0.02	0.38	0.86
28 I rush through activities without being attentive ^R	0.06	0.00	0.26	0.83
23 I am "running on automatic" ^R	0.16	0.05	0.26	0.62
<i>Nonjudge</i>				
25 I shouldn't be thinking the way I'm thinking ^R	0.09	0.00	0.28	0.77
35 I judge myself as good or bad ^R	0.16	0.01	0.33	0.67
17 I make judgments about my thoughts ^R	0.16	0.00	0.34	0.67
30 I think my emotions are bad or inappropriate ^R	0.01	0.00	0.42	0.98
14 I believe my thoughts are abnormal or bad ^R	0.12	0.02	0.29	0.70
10 I shouldn't be feeling the way I'm feeling ^R	0.12	0.03	0.25	0.68
39 I disapprove of myself ^R	0.13	0.00	0.29	0.69
3 I criticize myself for inappropriate emotions ^R	0.13	0.03	0.25	0.66
<i>Describe</i>				
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 ^R	0.04	0.02	0.32	0.89
16 I have trouble thinking of the right words ^R	0.23	0.03	0.21	0.48
7 I can easily put my thoughts into words ^R	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 can't find the right words to describe sensation ^R	0.24	0.01	0.34	0.58
<i>Nonreact</i>				
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

Table 5. *D-study reliability estimates and variance components for the Person (P) × Occasion (O) × Item (I) design, including interactions, for FFMQ items combined with the highest State Component Index (SCI). One item selected per subscale (a); and items selected from the total scale (b)*

Facets	5 items (a)		5 items (b)		6 items (b)		7 items (b)		8 items (b)	
	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^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
SE	0.10		0.09		0.09		3.20		0.08	
Gr	0.61		0.59		0.37		0.08		0.47	
GM	3.25		4.14		3.20		0.48		3.13	
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	

Note: GM= Grand mean; SE=Standard error of the grand mean

Chapter 5 Discussion

Using G-theory, the current study aimed to evaluate the reliability and generalisability over time of the most widely used multidimensional assessment tool to assess mindfulness as a trait – the FFMQ and its advanced short version, the FFMQ-18. This study also aimed to make a clear distinction between state and trait aspects in individual items and subscales of the FFMQ, as well as to determine sources of error that may skew measurement. The results show that the FFMQ and FFMQ-18 are reliable in measuring trait mindfulness, due to the fact that their G-coefficients are 0.89 and 0.75, respectively, and that they exhibit a high index of trait (both TCIs=0.90) and a low index of state (SCIs=0.10). Normally, a G-coefficient of 0.70 or higher is an acceptable benchmark indicating a reliable trait measurement (Arterberry et al., 2014), and a TCI of 0.60 or higher (SCI<0.40) is considered a measure of a trait (Medvedev et al., 2017a). Together, these results support temporal stability and generalisability of the FFMQ and FFMQ-18 scores across persons and occasions. Moreover, the application of G-theory in this study demonstrates its usefulness in evaluating reliability and distinguishing more dynamic and more enduring aspects of mindfulness. In the following sections, such aspects will be discussed in greater detail.

Main Findings of G-Study

The current study shows that the proposed TCI and SCI were particularly useful for exploring the degree of 'stateness' and 'traitness' of a measure. In this design, the person facet is used as an object of measurement and thus the interaction between person and occasion would reflect a state aspect, which is often treated as a measurement error, particularly if trait measures are examined (Arterberry et al., 2014). A state measure is considered reliable if its G-coefficient is less than 0.70 and simultaneously its SCI is more than 0.60 (Medvedev et al., 2017a). Conversely, a trait measure is indicated once its G-coefficient is 0.70 or higher and its TCI equals 0.60 or higher (SCI<0.40). As stated above, the scores of both the FFMQ and its FFMQ-18 short version indicate that both measure traits because their estimates of G-coefficients, TCIs, and SCIs satisfy these requirements. Without the SCI and TCI estimates in this study, the use of all five individual subscales of the FFMQ based merely on their G-coefficients would be inaccurate because researchers or

clinicians might use each individual subscale as a state mindfulness assessment tool (G-coefficients <0.45).

Although the scores of both the FFMQ and its FFMQ-18 short version are generalisable over time and across persons, all five individual subscales of the FFMQ were also found to measure trait mindfulness, with a TCI above 0.60 (SCI below 0.40), but they appear less reliable (G-coefficients below 0.45) in comparison to the FFMQ and FFMQ-18 total. The results also show that individual subscale scores were affected by measurement error due to interactions between person, item, and occasion, presenting the highest percentage of error variance, ranging from 43% to 64% across subscales. Individual subscales were also affected by interaction error between person and item and this was particularly 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 constituted 98% of the total error variance, but its influence on overall measurement reliability was very weak, with $G \geq 0.80$ (Shavelson et al., 1989). These values support the construct validity of the FFMQ and FFMQ-18 as instruments which assess trait mindfulness, because person variance reflects a trait and explains more than 80% of differences in scores.

Furthermore, the comparison between the current research and the study by Medvedev et al. (2017a) study shows that G-theory analysis illustrates a useful method for exploring state and trait variance components in a measurement. In the work by Medvedev et al., the authors used a G-design similar to that in the current study but their research focused on state measure and different (state) parameters were expected. Their study examined the state mindfulness measure, the TMS, which has two subscales (“curiosity” and “decentering”) (Lau et al., 2006). G-coefficients for both subscales of the TMS were below the acceptable benchmark for a trait measure (i.e., G-coefficients <0.70) while SCIs of the “curiosity” and “decentering” subscales were 0.70 and 0.75, respectively (Medvedev et al.). These results led to the conclusion that TMS subscales measured state. However, although G-coefficients of all individual subscales of the FFMQ in the current study were below 0.45, the same conclusion could not be drawn because the TMS subscales obtained low G-coefficients at the same time as satisfactory SCI values (SCIs>0.60). All these

estimates from both G-studies demonstrate the advance of G-theory in evaluating a measure to assess a state or a trait compared to the CTT method, whose conditions are based merely on the test-retest coefficient.

Contrast of G-Study and CTT Findings

Technically, G-coefficients cannot be directly compared to CTT reliability estimates. Consequently, the current study aims to indicate the differences between G-theory and CTT methods in evaluating the reliability of the FFMQ. Prior to G-analysis, the traditional CTT method was employed to estimate the reliability of the instrument. Although this aspect was not the main purpose, the CTT results of this study were consistent with Baer et al. (2006) (Table 2). This means that the CTT results indicate that the FFMQ is a trait measure, because its test-retest scores remained stable over three occasions (r ranged from 0.83 to 0.92). In addition, Cronbach's alphas obtained through CTT analysis for the five subscales of the FFMQ in this study were consistent with a systematic review conducted by Park et al. (2013). In this study, Cronbach's alphas ranged from 0.64 and 0.90, while in the review by Park et al., the alphas were between 0.67 and 0.93. In addition, this study examined Cronbach's alphas over time. These were not evaluated in the earlier study and the results showed that the scale maintained internal consistency over time (four out of five subscales and the total FFMQ/FFMQ-18).

Even though CTT is currently the dominant statistical method for distinguishing between state and trait components of an assessment, the application of G-theory in this study demonstrates its superiority to CTT in assessing the reliability of individual subscales of the FFMQ. There was a noticeable difference between the temporal reliability estimates obtained through CTT and G-theory methods for all five individual subscales of the FFMQ. Test-retest reliability coefficients computed by the CTT method for all facets, except the "nonreact" subscale (test-retest coefficients = 0.64 at Occasion 3), achieved acceptable benchmarks (>0.70), which indicated that all individual subscales measuring trait mindfulness were stable over time. These CTT coefficients indicate that most FFMQ individual subscales (except "nonreact") are trait measures. Findings based on CTT were not consistent with what was found using the G-theory method. G-theory results in this study indicate that all FFMQ subscales appeared less reliable in measuring trait

mindfulness. The difference between the two statistical methods is due to the fact that CTT coefficients in this study reveal that there is no investigation at all of different error sources, such as item error, occasion error and interaction errors. CTT examines different aspects of reliability independently (using Cronbach's alpha or test-retest) but it is unable to evaluate such error sources together in a single analysis. By contrast, the G-theory method in the current study estimates precisely all possible influences on reliability (item error, occasion error, and error in interactions between item, occasion and/or person) simultaneously, providing a more rigorous evaluation of the overall reliability that permits the generalisability of the FFMQ scores across sample occasions and populations.

D-Study Findings

A D-study was conducted to examine the degree of "stateness" of each FFMQ item and to investigate the possibilities for optimising the assessment of trait mindfulness and determine the possibility of assessing state mindfulness. The D-study results indicated that even though the FFMQ demonstrated sound reliability in the assessment of trait mindfulness, it contained no items which could be considered to be predominantly trait items, due to their high TCI (Table 4). However, there were many FFMQ items with high SCIs that reflected their sensitivity to changes over time ($SCI \geq 0.60$) and the remaining items could not differentiate clearly between state and trait mindfulness. These results indicate that most FFMQ items are likely to change over time and these changes are unlikely to take place simultaneously. This means that when an item changes in one way, other items tend to change differently. Each item, however, even if it is sensitive to a state, reflects a trait component to some degree and while state-related variances possibly compensate for each other, the remaining trait components contribute to the reliable assessment of a trait in the FFMQ and its short version.

Furthermore, after evaluating all individual FFMQ items according to their ability to capture a trait or a state, a number of generalisability tests were carried out 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. However, no combination of state-sensitive items resulted in a sensitive state measure, as reflected by low SCI. These findings also support

the earlier explanation that dynamic changes in particular aspects of mindfulness do not occur at the same time but cancel each other out if diverse 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 an SCI of 0.95 (TCI=0.05) and 0.98 (TCI=0.02), respectively, indicating that to a large extent they measure the state aspect of mindfulness. However, combining these items may counterbalance the state variability of each aspect over time because they are unlikely to occur simultaneously. This idea is further supported by the results in Table 5, where an attempt was made to combine state items and derive a state measure resulting in lower SCI. These findings and their explanation agree with psychometric studies demonstrating a diminution in measurement error resulting from individual items by merging them into parcels or super-items (Taylor et al., 2017; Medvedev et al., 2018).

Although further analyses conducted in the D-study did not result in any particular scales or subscales that were able to evaluate state mindfulness, the D-study provided a useful measure of the appropriateness of the G-study assessment design. For instance, removing items which were less sensitive to change across occasions did not improve the likelihood of obtaining a state scale or subscale from the FFMQ. This means that all FFMQ items play an important role in representing the overarching construct of dispositional mindfulness and are interrelated. Eventually, removing items in individual subscales and combining most- sensitive-to-change items did not reveal any psychometric benefits for the FFMQ, suggesting that the FFMQ is the most adequate measure of dispositional or trait mindfulness in the current measurement design, and the FFMQ-18 is the second best.

Five Facets of Mindfulness

It should be noted that each of the FFMQ subscales, except for “nonjudge”, included state items and items measuring both state and trait. All “nonjudge” subscale items were sensitive to change over time, however, because their SCIs ranged from 0.66 to 0.98, but the overall subscale sensitivity to change was very low (SCI=0.19), meaning that overall, this subscale did not reflect state changes. Consistent with other findings from this study, different aspects of a non-judgemental attitude captured by individual items (e.g.,

self, emotions, thoughts) do not appear to occur together at one point in time. Similarly, combining “nonjudge” items together reduces the overall subscale sensitivity to change because state-related variances may cancel each other out (Taylor et al. 2017; Medvedev et al. 2018). However, these findings indicate that various aspects of a non-judgemental attitude are very dynamic and should be the primary focus of any MBIs because they are more amendable, while a non-judgemental attitude was consistently found to be a strong predictor of psychological symptoms (Baer et al. 2008; Medvedev et al.).

Of the five state items on the “observe” subscale, three (“I pay attention to sensations”, “I notice the sensations of my body moving,” and “I notice how emotions affect thoughts and behaviour”) clearly assessed state, due to their high SCI (0.89, 0.88, and 0.75, respectively). The estimates on this subscale were consistent with other findings of the current research because the scale’s overall SCI was low, meaning that state variance is likely to be cancelled as noise, while trait components present in each item together contribute to the overall “traitness” of this subscale. If the aim is to develop mindful observing, then focusing in the first place on emotions, sensations, and thoughts 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” are difficult to qualify as either stable trait-like or dynamic state-like aspects of mindfulness, as they reflect both components to a comparable degree. Although this subscale is of low reliability, however, the evaluation of individual items on the “observe” subscale is useful for enhancing MBIs.

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”) on the “describe” subscale clearly displayed high sensitivity to change (state), with an SCI of 0.81 and 0.89, respectively. The remaining items in this facet are unqualified items, reflecting both temporary and enduring patterns. Despite the fact that dominant items measure both state and trait mindfulness and that the overall TCI of this scale is high (TCI= 0.82) — meaning that the trait variance of this subscale likely contributes to its overall “traitness” (SCI=0.12) — it has low reliability as a trait measure according to the results of its coefficients (i.e., $G_r=0.40$). This may be

accounted for by the fact that individual items measuring the ability to describe mindfulness relate to unobservable behaviours, such as feelings, sensations, and thoughts which change over time, reflected in high measurement error due to interactions between person, item, and occasion.

In comparison with the “describe” subscale, the “act with awareness” subscale has items of varying content, with more items measuring state. There were only two “act with awareness” items — “I am easily distracted” (SCI=0.45) and “I do jobs or tasks automatically” (SCI=0.48) — that reflect both state and trait for assessing over occasions. The remaining six out of eight items on this subscale reflected state aspects of mindfulness, with three items showing high SCIs, ranging from 0.83 to 0.95. Consistent with the findings for the total FFMQ, it is obvious that this scale should measure state mindfulness because there are many state items, but in fact it measures trait.

The overall SCI of this scale is low (SCI=0.23), meaning that state variance is likely to be cancelled out as noise while trait components present in each item contribute together to the overall “traitness” of this subscale. These results reflect the true nature of the “act with awareness” subscale, as it assesses the ability to be fully aware of whatever activities a person is engaged in, because an individual tends to perform a given activity differently at different times. For example, people may engage in eating their favourite meal everyday but they may eat it at different times, and it may taste different each day, so their ability to be fully aware of this activity will vary. However, even though there are significantly more “state-like” items than items measuring both state and trait in this facet compared to other FFMQ subscales, combining these items did not result in a sensitive state measure.

On the “nonreact” subscale, there were four items with an SCI higher than the benchmark that indicated high sensitivity to change (SCI>0.60). The item most sensitive to change across occasions in this facet is “I perceive my emotions without reacting to them” (SCI=0.80). The remaining three items on this subscale cannot be psychometrically quantified because they measure both a person’s trait and state. Although the “nonreact” subscale included more items sensitive to change over time than items measuring both state and trait aspects, the overall SCI was low (0.36), meaning that this subscale did not reflect

dynamic aspects of mindfulness reliably when these items were combined. As with the other findings of the current study, “nonreact” was influenced by measurement error due to interactions between person, item, and occasion. This indicates that people may respond to the same item differently on different occasions because individual thoughts and feelings vary over time.

Implications of G-study Findings

By applying G-theory method and providing evidence that scores can be generalised across sample populations and occasions, this study demonstrates the validity and reliability of the FFMQ and its short advanced FFMQ-18 version as trait measures. As a result, researchers and those applying MBIs clinically can reliably use the FFMQ or the FFMQ-18 in assessing individual trait mindfulness. Using the total scores of the FFMQ or FFMQ-18 can assist both researchers and clinical professionals to monitor accurately the effectiveness of MBIs and determine specifically if an MBI has a long-lasting effect. If a significant trait change has taken place, this will be reflected by a significant change in the FFMQ/FFMQ-18 scores. Since the FFMQ and FFMQ-18 measure traits, a change occurring as a result of an MBI is expected to be long lasting. However, researchers should also be aware that individual FFMQ subscales are less reliable for measuring trait mindfulness because they are affected by measurement error to a higher degree, compared with the FFMQ total score. Therefore, individual FFMQ facets should be used with caution, and preference should be given to the FFMQ total score to evaluate MBIs accurately.

In addition, this study contributes to the growing body of research using G-theory to establish the difference between state and trait aspects of an assessment and their true reliability. Although G-theory was strongly recommended for various applications in psychometric work (Brennan, 2001; Bloch & Norman, 2012), however, only a few studies have applied this theory to test the reliability of state and trait measures over time (Arterberry et al. 2014; Ulvenes, Berggraf, Wampold, Hoffart & McCullough, 2014).

There are two possible reasons to explain why G-theory method has not been widely applied in psychometrics. First, because of the complexity of the calculations used in G-theory, little user-friendly software is available. Secondly, data collection is laborious

because data tends to include at least three or more points in time and hence causes high attrition rates. EduG (Cardinet et al., 2009) is the most user-friendly software for conducting G-analysis. It allows both G- and D-studies to be easily implemented but still involves a degree of complexity because data must be prepared in univariate format, which precisely matches the G-study design defined by a researcher (Cardinet et al., 2009). However, despite some weaknesses due to data preparation for G-analysis, the advantages outweigh the inconveniences, mainly because distinguishing between trait and state aspects of measurements can be carried out accurately on the basis of variance components, rather than depending on a single test-retest correlation. Such work provides more evidence for the use of the G-theory method as standard for differentiating between state and trait aspects in psychometric instruments.

Even though SCI and TCI have been recently proposed and established with their preliminary cut-off points and benchmarks (Medvedev et al., 2017a), they play a vital role in differentiating the two aspects of a measure as well as in evaluating a measurement design. SCI and TCI estimates allow us to determine to what degree a scale or an item score assesses a trait or a state. This information can be useful to modify existing assessment methods or develop new instruments. Therefore, with SCI and TCI estimates, G-theory once again shows itself to be a superior alternative for validating state and trait measures, and is seen to be especially powerful in its ability to determine the “stateness” or “traitness” of each item on a scale.

Implications of D-study findings

A decision study (D-study) involves testing various assessment designs to optimise the FFMQ scale. D-study in this research indicates that the FFMQ and the FFMQ-18 achieve the most reliable and generalisable scores across occasions and sample populations in the current measurement design. D-study results also show that individual FFMQ subscales are less reliable in measuring traits and cannot be improved. Furthermore, with obvious aspects of each individual FFMQ item revealed through D-study, further investigation was conducted by combining all state items. However, all attempts to combine such items failed to result in a reliable measure of state, which is not surprising because the FFMQ was originally conceived as a trait measure and there are no state

mindfulness components, such as “curiosity” and “decentering” in the TMS, the state measure of mindfulness (Lau et al., 2006). Therefore, it is not necessary for a state mindfulness assessment method to be developed from the FFMQ because the reliability and validity of the TMS has already been validated using G-theory (Medvedev et al., 2017a).

Even though D-study has not resulted in any psychometric benefits for the FFMQ and its individual subscales, D-study is one of the most significant components of G-analysis. If no D-study had been conducted in this research, there would always be concerns about the assessment properties of individual FFMQ items. Therefore, while it seems that D-study has not reduced the measurement error of the design and has been unable to improve an instrument by removing items inconsistent with the aim of the measure, D-study has confirmed that no better measure can be derived from the FFMQ. Through D-study, it is clear that even with items highly indicative of state, the FFMQ/ FFMQ-18 will measure a trait because state variances cancel each other out. Unlike the TMS, for example, combining state items resulted in a reliable state measure because items were conceptualised and developed for measuring state mindfulness (Medvedev et al., 2017a). The extent to which state variances cancel each other out to produce a trait measure if items are combined, can only be examined using G-theory and, in particular, D-study.

Moreover, by virtue of D-study with SCIs obtained for each FFMQ item, it is clear that there is an imbalance in number between state items and unclassified items. As can be seen from the results, more FFMQ items were sensitive to change over time ($SCI \geq 0.60$), compared to items which could not be clearly classified as reflecting either state or trait because they were measuring both aspects, 25 and 14 items, respectively. However, the overall FFMQ is still a reliable trait assessment tool. This finding is important and has implications for developing the most effective mindfulness-based interventions, because dynamic aspects of mindfulness are the most amendable. Thus, the focus on such dynamic FFMQ items should be considered because they may reveal the effectiveness of an MBI.

Limitations

Some limitations should be acknowledged. The current study was conducted with participants who were all university students, showing a degree of homogeneity and

including a large population of women. Accordingly, these results should be tested with a more diverse sample. Specifically, the gender imbalance may influence the results and it would be helpful for later studies to replicate this analysis with a more balanced sample and analyse the genders separately. However, dynamic (state) and enduring (trait) patterns appear to be universal across the human population because they are related to physiological features of human functioning, such as attention, cognition and emotion (Cannon 1926; LeDoux 2000). This means that the findings of this study may well be generalisable outside the sample population.

The FFMQ-18 was analysed using data from the full FFMQ scale, potentially a limitation because responding to items presented in a different order may influence the results. Moreover, although the FFMQ contains 19 reverse-scored items designed to reduce response bias, they may potentially affect the reliability of the scale, i.e., the obtained G-coefficients could be higher if there were 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 a high SCI (≥ 0.60), reflecting high sensitivity for state changes over time. On the other hand, the remaining 14 items had an SCI between the benchmarks ($0.30 < \text{SCI} < 0.60$) and cannot be clearly classified as reflecting either state or trait because they measure both aspects. This means that there are no items with a low SCI (≤ 0.30) that are least sensitive to state changes and predominantly reflect trait mindfulness. These findings should be verified in future research using different samples to confirm the replicability of this result.

Since only a few studies have used the criteria of SCI and TCI estimates, more research is needed with different psychometric instruments and samples to support their benchmarks and cut-off points. Moreover, some concerns related to response bias are worthy of mention. For instance, it is possible that participants' attitudes (e.g., "it is a waste of their time/ is boring" or "not appealing to their field of interest") or feelings (e.g., "how they felt when they had to repeat the same thing three times") may have affected the results. However, despite its limitations, G-theory is a robust method for establishing reliability and the sample size was large enough for analyses of this type to be generalisable for the sample population and occasions.

Directions for Further Research

The literature shows that available mindfulness scales considered to be either state or trait measures were rarely confirmed by test-retest scores, and were not robust enough to differentiate sharply between trait and state mindfulness. The weaknesses of employing test-retest scores to differentiate between trait and state were discussed earlier. Only a few studies have applied G-theory to investigate state versus trait aspects of mindfulness instruments (e.g., Medvedev et al, 2017a). Consequently, the application of the G-theory method is useful and can be recommended for examining the ability of other psychometric measures of mindfulness (e.g., CHIME, FMI) and other psychological constructs for assessing state or trait.

Conclusion

The practice of mindfulness leads to changes of both state and trait and most reports of contemporary mindfulness research employ self-report mindfulness instruments to monitor such changes. Distinguishing accurately between stable and dynamic aspects of a measure is essential for valid and reliable assessments while applying MBIs. This study has used G-theory, the most suitable psychometric method available, to distinguish between temporary and enduring components in the most widely used multidimensional mindfulness instrument, the FFMQ. The findings indicate that reliable assessment of trait mindfulness can be achieved by using the full FFMQ scale or its short FFMQ-18 version, with scores generalisable across sample populations and occasions. However, individual facet subscales of the FFMQ appear to be less reliable in measuring trait mindfulness.

Moreover, despite the fact that all FFMQ items measure state or both state and trait, the overall reliability of the FFMQ over time is strong in measuring trait mindfulness linked to a trait component inherently present in every FFMQ item to a varying degree. This also means that in each aspect over time, all FFMQ items may counterbalance state changes that are less likely to occur simultaneously. This phenomenon might also be seen with further analysis combining the most dynamic FFMQ items, which did not result in any reliable state measure. This illustrates the point that regardless of state sensitivity, all

FFMQ items play an important role in representing the overarching construct of dispositional mindfulness.

Finally, this study implies that of current assessment methods, the FFMQ is the most adequate measure of dispositional or trait mindfulness and the FFMQ-18 is the second best. Thus, clinicians and researchers would benefit from using the total scores of the FFMQ and FFMQ-18 to reliably assess trait mindfulness.

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 generalisability theory to the big five inventory. *Personality and Individual Differences, 69*, 98-103.
- Bach, P., & Hayes, S. C. (2002). The use of Acceptance and Commitment Therapy to prevent the rehospitalisation of psychotic patients: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 70*, 1129-1139.
- Baer, R. (2003). Mindfulness training as a clinical intervention: A conceptual and empirical review. *Clinical Psychology: Science and Practice, 10*(2), 125-142.
- Baer, R. A., Carmody, J., & Hunsinger, M. (2012). Weekly change in mindfulness and perceived stress in a mindfulness-based stress reduction programme. *Journal of Clinical Psychology, 68*(7), 755-765. doi:10.1002/jclp.21865
- 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.
- Baer, R., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment, 11*(3), 191-206.
- Baer, R., Smith, G., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment, 13*(1), 27-45.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist 54*, 462–479.
- Barker, H. R., Wadsworth, A. P., & Wilson, W. (1976). Factor structure of the State-Trait Anxiety Inventory in a nonstressful situation. *Journal of Clinical Psychology, 32*, 595-598.
- Barnes, V. A., Davis, H. C., Murzynowski, J. B., & Treiber, F. A. (2004). Impact of meditation on resting and ambulatory blood pressure and heart rate in youth. *Psychosomatic Medicine, 66*(6), 909-914. doi:10.1097/01.psy.0000145902.91749.35

- Barnhofer, T., Chittka, T., Nightingale, H., Visser, C., & Crane, C. (2010). State effects of two forms of meditation on prefrontal EEG asymmetry in previously depressed individuals. *Mindfulness, 1*, 21-27.
- Barry, R. J., Clark, A. R., Johnstone, S. J., Magee, C. A., & Rushby, J. A. (2007). EEG differences between eyes-closed and eyes-open conditions. *Clinical Neurophysiology, 118*, 2765-2773.
- Baum, C., Kuyken, W., Bohus, M., Heidenreich, T., Michalak, J., & Steil, R. (2010). The psychometric properties of the Kentucky Inventory of the Mindfulness Skills in clinical populations. *Assessment, 17*(2), 220-229.
- Baumeister, J., Barthel, T., Geiss, K., & Weiss, M. (2008). Influence of phosphatidylserine on cognitive performance and cortical activity after induced stress. *Nutritional Neuroscience, 11*(3), 103-110. doi:10.1179/147683008x301478
- Beach, M. C., Roter, D., Korthuis, P. T., Epstein, R. M., Sharp, V., Ratanawongsa, N., et al. (2013). A multicentre study of physician mindfulness and health care quality. *Annals of Family Medicine, 11*, 421–428. doi:10.1370/afm.1507
- Beck, A. T., & Alford, B. A. (2009). *Depression: Causes and treatment*. Philadelphia, PA: University of Pennsylvania Press.
- Becker, D. E., & Shapiro, D. H. (1981). Physiological responses to clicks during Zen, yoga, and TM meditation. *Psychophysiology, 18*, 694-699.
- Beitel, M., Ferrer, E., & Cecero, J. J. (2005). Psychological mindedness and awareness of self and others. *Journal of Clinical Psychology, 61*, 739–750. doi:10.1002/jclp.20095
- Belzer, F., Schmidt, S., Lucius-Hoene, G., Schneider, J. F., Orellana-Rios, C. L., & Sauer, S. (2013). Challenging the construct validity of mindfulness assessment—A cognitive interview study of the Freiburg mindfulness inventory. *Mindfulness, 4*(1), 33-44.
- Bergomi, C., Tschacher, W., & Kupper, Z. (2013). The assessment of mindfulness with self-report measures: Existing scales and open issues. *Mindfulness, 4*(3), 191-202.
- Bergomi, C., Tschacher, W., & Kupper, Z. (2014). Konstruktion und erste Validierung eines Fragebogens zur umfassenden Erfassung von Achtsamkeit [Construction and

- initial validation of a questionnaire for the comprehensive investigation of mindfulness]. *Diagnostica*, 60, 111–125.
- Berkovich-Ohana, A., Glicksohn, J., & Goldstein, A. (2012). Mindfulness-induced changes in gamma band activity—implications for the default mode network, self-reference and attention. *Clinical neurophysiology*, 123(4), 700-710.
- Bernier, M., Thienot, E., Codron, R., & Fournier, J. F. (2009). Mindfulness and acceptance approaches in sport performance. *Journal of Clinical Sport Psychology*, 4, 320-333.
- Bhikkhu, T. (1997). The path of concentration and mindfulness. *Insight Journal*, 29–38.
Retrieved from <https://www.buddhisticinquiry.org/article/the-path-of-concentration-and-mindfulness/>
- Bichi, A. (2016). Classical test theory: An introduction to linear modeling approach to test and item analysis. *International Journal for Social Studies*, 2, 27-33.
- Birrer, D., Röthlin, P., & Morgan, G. (2012). Mindfulness to enhance athletic performance: Theoretical considerations and possible impact mechanisms. *Mindfulness*, 3(3), 235-246. doi: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.
- Black, D. A. (2014). *Mindfulness research publications by year, 1980–2013*. Retrieved from the American Mindfulness Research Association website: <https://goamra.org/>
- Black, D. S. (2011). A brief definition of mindfulness. *Mindfulness research guide*. Accessed from <http://www.mindfulexperience.org>
- Bloch, R., & Norman, G. (2012). Generalisability theory for the perplexed: A practical introduction and guide: AMEE guide no. 68. *Medical Teacher*, 34, 960–992.
- 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.
doi:10.1177/1073191111408231
- Bond, F. W., & Bunce, D. (2000). Mediators of change in emotion-focused and problem-focused worksite stress management interventions. *Journal of Occupational Health Psychology*, 5, 156-163.

- Branstetter, A. D., Wilson, K. G., Hildebrandt, M., & Mutch, D. (2004). *Improving psychological adjustment among cancer patients: ACT and CBT*. Paper presented at a meeting of the Association for Advancement of Behavior Therapy, New Orleans.
- Bränström, R., Kvillemo, P., Brandberg, Y., & Moskowitz, J. T. (2010). Self-report mindfulness as a mediator of psychological well-being in a stress reduction intervention for cancer patients—A randomized study. *Annals of Behavioral Medicine, 39*(2), 151–161. doi: 10.1007/s12160-010-9168-6
- Brennan, R. L. (1992). *Elements of generalisability theory* (Rev. ed.). Iowa City, IA: ACT, Inc.
- Brennan, R. L. (2001). *Generalisability theory*. New York: Springer-Verlag.
- Broderick, P. C., & Metz, S. (2009). Learning to BREATHE: A pilot trial of a mindfulness curriculum for adolescents. *Advances in School Mental Health Promotion, 2*, 35–46.
- 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. doi:10.1007/s12671-014-0349-4
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822-884.
- Brown, K.W., Ryan, R.M., Creswell, J. D. (2007). Mindfulness: Theoretical foundations and evidence for its salutary effects. *Psychological Inquiry, 18* (4): 211–37. doi:10.1080/10478400701598298
- Buchheld, N., Grossman, P., & Walach, H. (2001). Measuring mindfulness in insight meditation (vipassana) and meditation-based psychotherapy: The development of the Freiburg Mindfulness Inventory (FMI). *Journal for Meditation and Meditation Research, 1*, 11-34.
- Bush, M. (2011). Mindfulness in higher education. *Contemporary Buddhism, 12*(1), 183-197. doi: 10.1080/010183-197
- Buss, A. H. (1989). Personality as traits. *American Psychologist, 44*, 1378-1388.
- Byrne, J., Hauck, Y., Fisher, C., Bayes, S., & Schutze, R. (2013). Effectiveness of a mindfulness-based childbirth education pilot study on maternal self-efficacy and

- fear of childbirth. *Journal of Midwifery & Women's Health*, 59(2), 192-197.
doi:10.1111/jmwh.12075
- Byron, G., Ziedonis, D. M., McGrath, C., Frazier, J. A., & Fulwiler, C. (2015).
Implementation of mindfulness training for mental health staff: Organizational
context and stakeholder perspectives. *Mindfulness*, 6, 861–872.
- Cahn, B. R., & Polich, J. (2006). Meditation states and traits: EEG, ERP, and neuroimaging
studies. *Psychological Bulletin*, 132, 180-211.
- Cameron, C. D., & Fredrickson, B. L. (2015). Mindfulness facets predict helping behavior
and distinct helping-related emotions. *Mindfulness*, 6, 1211–1218.
doi:10.1007/s12671-014-0383-2
- Cannon, W. B. (1926). Physiological regulation of normal states: Some tentative postulates
concerning biological homeostatics. Paris: Editions Medicales.
- Cardaciotto, L., Herbert, J. D., Forman, E. M., Moitra, E., & Farrow, V. (2008). The
assessment of present-moment awareness and acceptance: The Philadelphia
Mindfulness Scale. *Assessment*, 15(2), 204-223.
- Cardinet, J., Pini, G., & Johnson, S. (2010). *Applying generalisability theory using EduG*.
London: Routledge Academic.
- Carlson, L. E., & Brown, K. W. (2005). Validation of the Mindful Attention Awareness
Scale in a cancer population. *Journal of Psychosomatic Research*, 58(1), 29-33.
- Carlson, L. E., Speca, M., Farris, P., & Patel, K. D. (2007). One year pre–post intervention
follow-up of psychological, immune, endocrine and blood pressure outcomes of
mindfulness-based stress reduction (MBSR) in breast and prostate cancer
outpatients. *Brain, Behavior, and Immunity*, 21(8), 1038-1049.
doi:10.1016/j.bbi.2007.04.002
- Carmody, J., & Baer, R. A. (2008). Relationships between mindfulness practice and levels
of mindfulness, medical and psychological symptoms and well-being in a
mindfulness-based stress reduction programme. *Journal of Behavioral Medicine*, 31,
23–33. doi:10.1007/s10865-007-9130-7
- Cash, M., & Whittingham, K. (2010). What facets of mindfulness contribute to
psychological well-being and depressive, anxious, and stress-related
symptomatology? *Mindfulness*, 1(3), 177–182.

- Cayoun, B. (2011). *Mindfulness-integrated CBT: Principles and practice*. Oxford: John Wiley & Sons, Ltd.
- Chadwick, P., Hember, M., Mead, S., Lilley, B., & Dagnan, D. (2005). Responding mindfully to unpleasant thoughts and images: Reliability and validity of the Mindfulness Questionnaire. *Unpublished manuscript*.
- Chadwick, P., Hember, M., Symes, J., Peters, E., Kuipers, E., & Dagnan, D. (2008). Responding mindfully to unpleasant thoughts and images: Reliability and validity of the Southampton Mindfulness questionnaire (SMQ). *The British Journal of Clinical Psychology*, *47*(4), 451-455.
- Chang, V. Y., Palesh, O., Caldwell, R., Glasgow, N., Abramson, M., Luskin, F., . . . , & Koopman, C. (2004). The effects of a mindfulness-based stress reduction programme on stress, mindfulness self-efficacy, and positive states of mind. *Stress & Health*, *20*(3), 141-147. doi: 10.1002/smi.1011
- Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of state and trait: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology*, *54*(4), 541-557.
- Chapman, M., Hare, D., Caton, S., Donalds, D., McInnis, E., & Mitchell, D. (2013). The use of mindfulness with people with intellectual disabilities: A systematic review and narrative analysis. *Mindfulness*, *4*(2), 179–189.
- Chiesa, A., & Serretti, A. (2009). Mindfulness-based stress reduction for stress management in healthy people: A review and meta-analysis. *The Journal of Alternative and Complementary Medicine*, *15*(5), 593-600.
doi:10.1089/acm.2008.0495
- Chiesa, A., & Serretti, A. (2010). A systematic review of neurobiological and clinical features of mindfulness meditations. *Psychological Medicine*, *40*, 1239-1252.
- Christopher, M. S., & Gilbert, B. D. (2010). Incremental validity of components of mindfulness in the prediction of satisfaction with life and depression. *Current Psychology*, *29*(1), 10-23.
- Christopher, M. S., Charoensuk, S., Gilbert, B. D., Neary, T. J., & Pearce, K. L. (2009). Mindfulness in Thailand and the United States: A case of apples versus oranges? *Journal of Clinical Psychology*, *65*(6), 590-612.

- Christopher, M. S., Neuser, N. J., Michael, P. G., & Baitmangalkar, A. (2012). Exploring the psychometric properties of the Five Facet Mindfulness Questionnaire. *Mindfulness, 3*(2), 124–131.
- Coelho, H. F., Canter, P. H., & Ernst, E. (2007). Mindfulness-based cognitive therapy: Evaluating current evidence and informing future research. *Journal of Consulting and Clinical Psychology, 75*(6), 1000-1005.
- 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. doi: 10.1007/s12671-010-0033-2
- Cohen, J. (2017). Synesthetic perception as continuous with ordinary perception, or: We're all synesthetes now. *Sensory Blendings: New Essays on Synaesthesia*, 1-28.
- Cohen, R. J., & Swerdlik, M. E. (2010). *Psychological testing and assessment: An introduction to tests and measurement*. New-York: McGraw-Hill.
- Cohen-Katz, J., Wiley, S. D., Capuano, T., Schaufeli, D. M., & Shapiro, S. (2005). The effects of mindfulness-based stress reduction on nurse stress and burnout, part II: A quantitative and qualitative study. *Holistic Nursing Practice, 19*, 26–35.
- Cordon, S. L., & Finney, S. J. (2008). Measurement invariance of the Mindful Attention Awareness Scale across adult attachment style. *Measurement & Evaluation in Counseling & Development, 40*(4), 228-245.
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalisability: A liberalization of reliability theory. *British Journal of Statistical Psychology, 16*, 137-163.
- Dahl, J., Wilson, K. G., & Nilsson, A. (2004). Acceptance and commitment therapy and the treatment of persons at risk for long-term disability resulting from stress and pain symptoms: A preliminary randomized trial. *Behavior Therapy, 35*, 785-802.
- Dakwar, E., Mariani, J. P., & Levin, F. R. (2011). Mindfulness impairments in individuals seeking treatment for substance use disorders. *The American Journal of Drug and Alcohol Abuse, 37*(3), 165–169. doi: 10.3109/00952990.2011.553978
- Dalai Lama, Baron, R., & Gaffiney, P. (2004). *Dzogchen: Heart Essence of the great Perfection*. (2nd ed.). New York, NY: Snow Lion Publications.

- Davidson, R. J., Kabat-Zinn, J., Schumacher, J., Rosenkranz, M., Muller, D., & Santorelli, S. F. (2003). Alterations in brain and immune function produced by mindfulness meditation. *Psychosomatic Medicine*, *65*(564-570).
- 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), 185–197.
- Dekeyser, M., Raes, F., Leijssen, M., Leysen, S., & Dewulf, D. (2008). Mindfulness skills and interpersonal behaviour. *Personality and Individual Differences*, *44*, 1235–1245. doi:10.1016/j.paid.2007.11.018
- Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmalings, K. B., Gallop, R., McGlinchey, J. B., ... Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, *74*(4), 658-670.
- Dimidjian, S., & Linehan, M. M. (2003). Defining an agenda for future research on the clinical application of mindfulness practice. *Clinical Psychology: Science and Practice*, *10*, 166-171.
- Donnelly, K., James, K., & Walz, L. (2015). Group-based mindfulness practice training for individuals with a learning difficulty: A pilot study. *Clinical Psychology Forum*, *266*, 7–16.
- Dunn, B. R., Hartigan, J. A., & Mikulas, W. L. (1999). Concentration and mindfulness meditations: Unique forms of consciousness? *Applied Psychophysiology and Biofeedback*, *24*(3), 147-165.
- Epstein, S. (1984). Trait theory as personality theory: Can a part be as great as the whole? *Psychological Inquiry*, *5*, 120-122.
- Erwin, R. J., Mawhinney-Hee, M., Gur, R. C., & Gur, R. E. (1989). Effects of task and gender on EEG indices of hemispheric activation: similarities to previous rCBF findings. *Neuropsychiatry, Neuropsychology, & Behavioural Neurology*, *2*(4), 248-260.
- Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*, 903–911. doi:10.1016/j.neubiorev.2010.10.009

- Felder, J. N., Dimidjian, S., & Segal, Z. (2012). Collaboration in mindfulness-based cognitive therapy. *Journal of Clinical Psychology*, *68*(2), 179-186.
doi:10.1002/jclp.21832
- Feldman, G., Hayes, A., Kumar, S., Greeson, J., & Laurenceau, J. (2007). Mindfulness and emotion regulation: The development and initial validation of the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R). *Journal of Psychopathology & Behavioral Assessment*, *29*(3), 177-190.
- Fisak, B., & von Lehe, A. C. (2012). The relation between the five facets of mindfulness and worry in a non-clinical sample. *Mindfulness*, *3*(1), 15–21.
- Fisher, B., Allen, R., & Kose, G. (1996). The relationship between anxiety and problem solving skills in children with and without learning disabilities. *Journal of Learning Disabilities*, *29*(4), 439–446.
- Fisher, R. A. (1925). *Statistical methods for research workers*. London: Oliver & Bond.
- Fitzgibbon, S., Pope, K., Mackenzie, L., Clark, C., & Willoughby, J. (2004). Cognitive tasks augment gamma EEG power. *Clinical Neurophysiology*, *115*(8), 1802-1809.
doi:10.1016/j.clinph.2004.03.009
- Flook, L., Smalley, S. L., Kitil, M. J., Galla, B. M., Kaiser-Greenland, S., Locke, J., . . . Kasari, C. (2010). Effects of mindful awareness practices on executive functions in elementary school children. *Journal of Applied School Psychology*, *26*(1), 70-95.
doi:10.1080/15377900903379125
- Fortney, L., Luchterhand, C., Zakletskaia, L., Zgierska, A., & Rakel, D. (2013). Abbreviated mindfulness intervention for job satisfaction, quality of life, and compassion in primary care clinicians: A pilot study. *Annals of Family Medicine*, *11*, 412–420. doi:10.1370/afm.1511
- Foureur, M., Besley, K., Burton, G., Yu, N., & Crisp, J. (2013). Enhancing the resilience of nurses and midwives: Pilot of a mindfulness based programme for increased health, sense of coherence and decreased depression, anxiety and stress. *Contemporary nurse*, *45*(1), 114-125.
- 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. doi:10.1007/s12671-012-0161-y

- Frewen, P., Evans, E., Maraj, N., Dozois, D., & Partridge, K. (2008). Letting go: Mindfulness and negative automatic thinking. *Cognitive Therapy & Research, 32*(6), 758-774.
- Gardner, F. L., & Moore, Z. E. (2004). A Mindfulness-Acceptance Commitment-Based approach to athletic performance enhancement: Theoretical considerations. *Behavior Therapy, 35*, 707-723.
- Gardner, F. L., & Moore, Z. E. (2007). *The psychology of enhancing human performance. The mindfulness–acceptance–commitment approach (MAC)*. New York, NY: Springer.
- Gaudiano, B. A., & Herbert, J. D. (2006). Acute treatment of inpatients with psychotic symptoms using Acceptance and Commitment Therapy: Pilot results. *Behaviour Research and Therapy, 44*(3), 415–437. doi: 10.1016/j.brat.2005.02.007
- Gazzaniga, M. S., Ivry, R. V., Mangun, G. R., & Steven, M. S. (2009). *Cognitive neuroscience: The biology of the mind* (3rd ed.). New York: W. W. Norton & Company, Inc.
- Geiser, C., Litson, K., Bishop, J., Keller, B. T., Burns, G. L., Servera, M., & Shiffman, S. (2015). Analyzing person, situation and person \times situation interaction effects: Latent state-trait models for the combination of random and fixed situations. *Psychological Methods, 20*(2), 165–192. doi: 10.1037/met0000026
- Gelles, D. (2012). *Mindful work: How meditation is changing business from the inside out*. New York, NY: Houghton Mifflin Harcourt.
- Germer, C. K., Siegel, R. D., & Fulton, P. R. (Eds.). (2005). *Mindfulness and psychotherapy*. New York, NY: The Guilford Press.
- Glomb, T. M., Duffy, M. K., Bono, J. E., & Yang, T. (2012). Mindfulness at work. *Research in Personnel and Human Resource Management, 30*, 115–157.
- Goldin, P. R., & Gross, J. J. (2010). Effects of Mindfulness-Based Stress Reduction (MBSR) on emotion regulation in social anxiety disorder. *Emotion, 10*(1), 89-91.
- Goodman, M. J., & Schorling, J. B. (2012). A mindfulness course decreases burnout and improves well-being among healthcare providers. *International Journal of Psychiatry in Medicine, 43*(2), 119–128.

- Govern, J. M., & Marsch, L. A. (2001). Development and validation of the situational self-awareness scale. *Consciousness and Cognition, 10*, 366–378.
- Greason, P. B., & Cashwell, C. S. (2009). Mindfulness and counseling self-efficacy: The mediating role of attention and empathy. *Counselor Education and Supervision, 49*, 2–19. doi:10.1002/j.1556-6978.2009.tb00083.x
- Greco, L. A., Baer, R. A., & Smith, G. T. (2011). Assessing mindfulness in children and adolescents: Development and validation of the Child and Adolescent Mindfulness Measure (CAMM). *Psychological Assessment, 23*(3), 606–614. doi: 10.1037/a0022819
- 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. doi: 10.1037/pas0000263
- Gulliksen, H. (1950). *Theory of mental tests*. New York, NY: Wiley.
- Gunaratana, B. (2002). *Mindfulness in plain English*. Somerville, MA: Wisdom Publications.
- Haigh, E. A., Moore, M. T., Kashdan, T. B., & Fresco, D. M. (2011). Examination of the factor structure and concurrent validity of the Langer Mindfulness/Mindlessness Scale. *Assessment, 18*(1), 11-26.
- Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. (2007). The integrated trait-state model. *Journal of Research in Personality, 41*, 295–315.
- Hambleton, R. K., & Jones, R. W. (2005). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38-47.
- Hayes, S. C., Follette, V. M., & Linehan, M. M. (Eds.). (2004). *Mindfulness and acceptance: Expanding cognitive-behavioral tradition*. New York, NY: The Guilford Press.
- Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes, and outcomes. *Behaviour Research and Therapy, 44*, 1-25.

- Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York, NY: Guilford Press.
- Hewett, Z. L., Ransdell, L. B., Gao, Y., Petlichkoff, L. M., & Lucas, S. (2011). An examination of the effectiveness of an 8-week Bikram yoga programme on mindfulness, perceived stress, and physical fitness. *Journal of Exercise Science & Fitness, 9*(2), 87-92
- Hofmann, S. G., Sawyer, A. T., Witt, A., & Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*(2), 169-183.
- Hölzel, B. K., Carmody, J., Vangela, M., Congleton, C., Yerramsetia, S. M., Gard, T., & Lazar, S. W. (2011). Mindfulness practice leads to increases in regional brain gray matter density. *Psychiatry Research: Neuroimaging, 191*, 36-43.
- Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality and Quantity, 34*(4), 331-351.
- Hwang, Y., & Kearney, P. (2013). A systematic review of mindfulness intervention for individuals with developmental disabilities: Long-term practice and long-lasting effects. *Research in Developmental Disabilities, 34*(1), 314–326.
- 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*. doi: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. doi: 10.1017/iop.2015.41
- Irving, J. A., Dobkin, P. L., & Park, J. (2009). Cultivating mindfulness in health care professionals: A review of empirical studies of mindfulness-based stress reduction (MBSR). *Complementary Therapies in Clinical Practice, 15*(2), 61–66. doi:10.1016/j.ctcp.2009.01.002
- Ivanovski, B., & Malhi, G. S. (2007). The psychological and neurophysiological concomitants of mindfulness forms of meditation. *Acta Neuropsychiatrica, 19*, 76-91.

- Ives-Deliperi, V. L., Solms, M., & Meintjes, E. M. (2010). The neural substrates of mindfulness: An fMRI investigation. *Social Neuroscience*, 6(3), 231-242.
- Jacobs, R. H., Guo, S., Kaundinya, P., Lakind, D., Klein, J., Rusch, D., ... Atkins, M. (2017). A pilot study of mindfulness skills to reduce stress among a diverse paraprofessional workforce. *Journal of Child and Family Studies*, 26(9), 2579–2588. doi: 10.1007/s10826-017-0771-z
- Jacobson, N. S., Martell, C. R., & Dimidjian, S. (2001). Behavioral activation treatment for depression: Returning to contextual roots. *Clinical Psychology: Science and Practice*, 8(3), 255-270.
- Janssen, M., Heerkens, Y., Kuijer, W., Heijden, B. V. D., & Engels, J. (2018). Effects of Mindfulness-Based Stress Reduction on employees' mental health: A systematic review. *Plos One*, 13(1). doi: 10.1371/journal.pone.0191332
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big-Five Inventory-Version 4a and 54*. Berkeley, CA: Berkeley Institute of Personality and Social Research, University of California.
- Kabat-Zinn, J. (1982). An outpatient programme in behavioural medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry*, 4, 33-47.
- Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain and illness*. New York, NY: Delacorte.
- Kabat-Zinn, J. (1994). *Wherever you go, there you are: Mindfulness meditation in everyday life*. New York: Hyperion.
- Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. *Clinical Psychology: Science and Practice*, 10, 144-156.
- Kabat-Zinn, J. (Ed.). (2000). *Indra's net at work: The mindstreaming of Dharma practice in society*.
- Kabat-Zinn, J., Beall, B., & Rippe, J. (1985). A systematic mental training programme based on mindfulness meditation to optimize performance in collegiate and Olympic rowers. Poster presented at the World Congress in Sport Psychology, Copenhagen, Denmark.

- 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. doi:10.1097/00006842-199809000-00020
- Kaplan, D. (2008). *Structural equation modeling: Foundations and extensions* (2nd ed.). Sage Publications.
- Karyadi, K. A., Vanderveen, J. D., & Cyders, M. A. (2014). A meta-analysis of the relationship between trait mindfulness and substance use behaviors. *Drug and Alcohol Dependence, 143*, 1–10. doi: 10.1016/j.drugalcdep.2014.07.014
- Kasamatsu, A., & Hirai, T. (1966). An electroencephalographic study on the zen meditation (Zazen). *Folia Psychiatrica et Neurologica Japonica, 20*, 315-336.
- Kaufman, K. A., Glass, C. R., & Arnkoff, D. B. (2009). An evaluation of Mindful Sport Performance Enhancement (MSPE): A new mental training approach to promote flow in athletes. *Journal of Clinical Sport Psychology, 4*, 334–356.
- Kelly, C. (2012, April 28). O.K., Google, take a deep breath. *New York Times*. Retrieved from http://www.nytimes.com/2012/04/29/technology/google-course-asks-employees-to-take-a-deep-breath.html?pagewanted=all&_r=0
- Keng, S. L., Smoski, M. J., & Robin, C. J. (2011). Effects of mindfulness on psychological health: A review of empirical studies. *Clinical Psychology Review, 31*(6), 1041-1056.
- Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multivariate data. *Journal of Counselling and Clinical Psychology, 63*(1), 52-59.
- Keune, P. M., Bostanov, V., Hautzinger, M., & Kotchubey, B. (2011). Mindfulness-based cognitive therapy (MBCT), cognitive style, and the temporal dynamics of frontal EEG alpha asymmetry in recurrently depressed patients. *Biological Psychology, 88*(2-3), 243-252.
- Kline, P. (2013). *Handbook of Psychological Testing*. Florence: Taylor and Francis. doi: 10.4324/9781315812274
- Kline, R. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford.

- Krätgeloh, 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.
- Krasner, M. S., Epstein, R. M., Beckman, H., Suchman, A. L., Chapman, B., Mooney, C. J., & Quill, T. E. (2009). Association of an educational programme in mindful communication with burnout, empathy, and attitudes among primary care physicians. *JAMA: The Journal of the American Medical Association*, *302*, 1284–1293. doi:10.1001/jama.2009.1384
- Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In *Handbook of Survey Research, Second Edition* (pp. 263–313). Emerald Group Publishing Limited.
- 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.
- Leary, M. R. (2004). *The curse of the self: Self-awareness, egotism, and the quality of human life*. New York: Oxford University Press.
- Leary, M. R., Adams, C. E., & Tate, E. B. (2006). Hypo-egoic selfregulation: Exercising self-control by diminishing the influence of the self. *Journal of Personality*, *74*, 1803–1831.
- Ledesma, D., & Kumano, H. (2009). Mindfulness-based stress reduction and cancer: A meta-analysis. *Psycho-Oncology*, *18*(6), 571-579. doi:10.1002/pon.1400
- LeDoux, J. E. (2000) Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*,155-184.
- 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*, Article 1202. doi:10.3389/fpsyg.2014.01202
- Levy, D. M., Wobbrock, J. O., Kaszniak, A. W., & Ostergren, M. (2012, May). *The effects of mindfulness meditation training on multitasking in a high-stress information environment*. Paper presented at Graphics Interface Conference, Toronto, Canada.

- Libet, B. (2004). *Mind time: The temporal factor in consciousness*. Cambridge, MA: Harvard University Press.
- Linehan, M. M. (1993a). *Cognitive-behavioural treatment of borderline personality disorder*. New York: Guilford Press.
- Linehan, M. M. (1993b). *Skills training manual for treating borderline personality disorder*. New York, NY: Guilford Press.
- Lord, F. M., & Novick, M. R. (1968) *Statistical theory of mental test scores*. Reading, MA: Addison-Wesley.
- Lush, E., Salmon, P., Floyd, A., Studts, J. L., Weissbecker, I., & Sephton, S. E. (2009). Mindfulness meditation for symptom reduction in fibromyalgia: Psychophysiological correlates. *Journal of Clinical Psychology in Medical Settings*, 16(2), 200–207. doi:10.1007/s10880-009-9153-z
- 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. doi:10.1037/0022-006X.72.1.31
- MacDonald, H. Z., & Baxter, E. E. (in press). Mediators of the relationship between dispositional mindfulness and psychological well-being in female college students. *Mindfulness*. doi:10.1007/s12671-016-0611-z
- Mace, C. (2008). *Mindfulness and mental health: Therapy, theory and science*. New York, NY: Routledge.
- MacKillop, J., & Anderson, E. J. (2007). Further psychometric validation of the Mindful Attention Awareness Scale (MAAS). *Journal of Psychopathology & Behavioral Assessment*, 29(4), 289-293.
- Marlatt, G. A., & Kristeller, J. L. (Eds.). (1999). *Mindfulness and meditation*. Washington DC: American Psychological Association.
- 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. doi:10.1016/0306-9877(95)90299-6
- Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of empathy for ‘social pain’ and subsequent prosocial behavior. *NeuroImage*, 55, 381–388. doi:10.1016/j.neuroimage.2010.11.060

- Matvienko-Sikar, K., Lee, L., Murphy, G., & Murphy, L. (2016). The effects of mindfulness interventions on prenatal well-being: A systematic review. *Psychology & Health, 31*(12), 1415-1434. doi:10.1080/08870446.2016.1220557
- McCracken, L. M., & Yang, S.-Y. (2008). A contextual cognitive-behavioral analysis of rehabilitation workers' health and well-being: Influences of acceptance, mindfulness, and values-based action. *Rehabilitation Psychology, 53*, 479–485. doi:10.1037/a0012854
- McManus, F., Surawy, C., Muse, K., Vazquez-Montes, M., & Williams, J. M. G. (2012). A randomized clinical trial of Mindfulness-Based Cognitive Therapy versus unrestricted services for health anxiety (hypochondriasis). *Journal of Consulting and Clinical Psychology, 80*, 817–828. doi:10.1037/a0028782
- Medvedev, O. N., Krägeloh, C. U., Narayanan, A., & Siegert, R. J. (2017a). Measuring mindfulness: Applying generalisability theory to distinguish between state and trait. *Mindfulness, 8*(4), 1036-1046. doi: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. doi:10.1007/s12671-018-0917-0
- Medvedev, O. N., Siegert, R. J., Kersten, P., & Krägeloh, C. U. (2017). Improving the precision of the Five Facet Mindfulness Questionnaire using a Rasch approach. *Mindfulness, 8*(4), 995-1008. doi:10.1007/s12671-016-0676-8
- Medvedev, O. N., Theadom, A., Barker-Collo, S., & Feigin, V. (2018a). Distinguishing between enduring and dynamic concussion symptoms: Applying Generalisability Theory to the Rivermead Post Concussion Symptoms Questionnaire (RPQ). *PeerJ, 6*. doi: 10.7717/peerj.5676
- MiCBT Institute. (2019). *Mindfulness-integrated Cognitive Behaviour Therapy (MiCBT)*. Retrieved 7.01, 2019, from <http://www.mindfulness.net.au/conceptual-outline/>
- Murata, T., Koshino, Y., & Omori, M. (1994). Quantitative EEG study on Zen meditation (zaZen). *Japanese Journal of Psychiatry and Neurology, 48*, 881-890.
- Nickerson, R. S. (Ed.). (1978). *Attention and Performance VIII*. New Jersey: Lawrence Erlbaum Associates, Inc.

- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. 3rd edition. New York, NY: McGraw-Hill.
- Nyanaponika, T. (1973). *The heart of Buddhist meditation*. New York, NY: Weiser Books.
- Nyklíček, I., & Kuijpers, K. F. (2008). Effects of Mindfulness-Based Stress Reduction intervention on psychological well-being and quality of life: Is increased mindfulness indeed the mechanism? *Annals of Behavioral Medicine*, *35*, 331–340. doi:10.1007/s12160-008-9030-2
- Olendzki, A. (Ed.). (2005). *The roots of mindfulness*. New York, NY: Guilford.
- Pagnoni, G., & Cekic, M. (2007). Age effects on gray matter volume and attentional performance in Zen meditation. *Neurobiology of Aging*, *28*, 1623–1627.
- Pagnoni, G., Cekic, M., & Guo, Y. (2008). ‘Thinking about notthinking’: Neural correlates of conceptual processing during Zen meditation. *PLOS One*, *3*(9), e3083.
- Park, T., Reilly-Spong, M., & Gross, C. R. (2013). Mindfulness: A systematic review of instruments to measure an emergent patient-reported outcome (PRO). *Quality of Life Research*, *22*, 2639-2659.
- Paterson, J., Medvedev, O. N., Sumich, A., Tautolo, E., Krägeloh, C. U., Sisk, R., . . . Siegert, R. J. (2017). Distinguishing transient versus stable aspects of depression in New Zealand Pacific Island children using Generalisability Theory. *Journal of Affective Disorders*, *227*, 698-704. doi:10.1016/j.jad.2017.11.075
- Praissman, S. (2008). Mindfulness-based stress reduction: A literature review and clinician’s guide. *Journal of the American Academy of Nurse Practitioners*, *20*(4), 212–216. doi:10.1111/j.1745-7599.2008.00306.x
- Ramanaiah, N. V., Franzen, M., & Schill, T. (1983). A psychometric study of the State-Trait Anxiety inventory. *Personality Assessment*, *47*, 531–535.
- Schonert-Reichl, K. A., & Lawlor, M. S. (2010). The effects of a mindfulness-based education programme on pre- and early adolescents’ well-being and social and emotional competence. *Mindfulness*, *1*(3), 137-151. doi:10.1007/s12671-010-0011-8
- Schonert-Reichl, K. A., Oberle, E., & Lawlor, M. S. (2015). Enhancing cognitive and social–emotional development through a simple-to-administer mindfulness-based

- school programme for elementary school children: A randomized controlled trial. *Developmental Psychology*, 51, 52–66.
- Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2002). *Mindfulness based cognitive therapy for depression*. New York, NY: Guilford Press.
- Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2013). *Mindfulness-Based Cognitive Therapy for depression* (2nd ed.). New York, NY: Guilford Press.
- Shapiro, S. L., & Carlson, L. E. (2009). *The art and science of mindfulness: Integrating mindfulness into psychology and the helping professions*. Washington, DC: American Psychological Association. doi:org/10.1037/11885-000
- 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. doi:10.1037/1072-5245.12.2.164
- Shapiro, S. L., Brown, K. W., & Biegel, G. M. (2007). Teaching selfcare to caregivers: Effects of mindfulness-based stress reduction on the mental health of therapists in training. *Training and Education in Professional Psychology*, 1, 105.
- Shapiro, S. L., Schwartz, G. E., & Bonner, G. (1998). Effects of mindfulness-based stress reduction on medical and premedical students. *Journal of Behavioral Medicine*, 21, 581–599.
- Shavelson, R.J., Webb, N.M., Rowley, G.L. (1989). Generalisability theory. *American Psychologist*, 44, 599–612.
- Siebelink, N. M., Asherson, P., Antonova, E., Bögels, S. M., Speckens, A. E., Buitelaar, J. K., & Greven, C. U. (2019). Genetic and environmental aetiologies of associations between dispositional mindfulness and ADHD traits: A population-based twin study. *European Child & Adolescent Psychiatry*, 1-11. doi: 10.1007/s00787-019-01279-8
- Siegel, R. D., Germer, C. K., & Olendzki, A. (2009). Mindfulness: What is it? Where did it come from? *Clinical Handbook of Mindfulness*, 17-35. doi:10.1007/978-0-387-09593-6_2
- Simons, D. J., & Chabris, C. F. (1999) Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28, 1059-1074.

- Singh, N., Wahler, R., Adkins, A., & Myers, R. (2003). Soles of the feet: A mindfulness-based self-control intervention for aggression by an individual with mild mental retardation and mental illness. *Research in Developmental Disabilities, 24*(3), 158–169.
- Somer, E., Somer, L., & Jopp, D. S. (2016). Childhood antecedents and maintaining factors in maladaptive daydreaming. *The Journal of Nervous and Mental Disease, 204*(6), 471–478. doi: 10.1097/nmd.0000000000000507
- Spearman, C. E. (1904). 'General intelligence' objectively determined and measured. *American Journal of Psychology, 5*, 201-293.
- 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, CA: Consulting Psychologists Press.
- Stelmack, R. (2004). On personality and arousal: A historical perspective on Eysenck and Zuckerman. *On the Psychobiology of Personality, 17–28*. doi: 10.1016/b978-008044209-9/50003-8
- Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment, 8*(2), 78-98.
- Strowger, M., Kiken, L. G., & Ramcharran, K. (2018). Mindfulness meditation and physical activity: Evidence from 2012 National Health Interview Survey. *Health Psychology, 37*(10), 924-928.
- 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. doi:10.1016/0895-4356(96)00016-9
- Suen, H. K., and Lei, P. W. (2007). Classical versus generalisability theory of measurement. *Educational Measurement, 4*, 1-13.
- Takahashia, T., Murataa, T., Hamadab, T., Masao Omoria, Kosakaa, H., Kikuchic, M., ... Wadaa, Y. (2004). Changes in EEG and autonomic nervous activity during meditation and their association with personality traits. *International Journal of Psychophysiology, 55*, 199-207.

- Tanay, G., & Bernstein, A. (2013). State Mindfulness Scale (SMS): Development and initial validation. *Psychological Assessment, 25*(4), 1286-1299.
- Tang, Y., Holzel, B., & Posner, M. (2015). The neuroscience of mindfulness meditation. *Nature Reviews Neuroscience, 16*, 213–225.
- Teasdale, J. D., Segal, Z. V., Williams, J. M. G., Ridgeway, V. A., Soulsby, J. M., & Lau, M. A. (2000). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *Journal of Consulting and Clinical Psychology, 68*, 615–623.
- Tellegen, A. (1982). *Content categories: Absorption items* (Revised). Unpublished manuscript, University of Minnesota, Minneapolis.
- Thornton, V., Williamson, R., & Cooke, B. (2017). A mindfulness-based group for young people with learning disabilities: A pilot study. *British Journal of Learning Disabilities, 45*(4), 259-265. doi:10.1111/bld.12203
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2012). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
- Trapnell, P. D., & Campbell, J. D. (1999). Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology, 76*, 284–304.
- Twohig, M. P., Hayes, S. C., & Masuda, A. (2006). Increasing willingness to experience obsessions: Acceptance and commitment therapy as a treatment for obsessive compulsive disorder. *Behavior Therapy, 37*(1), 3-13.
- Ulvenes, P. G., Berggraf, L., Wampold, B. E., Hoffart, A., Stiles, T., & McCullough, L. (2014). Orienting patient to affect, sense of self, and the activation of affect over the course of psychotherapy with cluster C patients. *Journal of Counseling Psychology, 61*(3), 315.
- Vieten, C., & Astin, J. (2008). Effects of a mindfulness-based intervention during pregnancy on prenatal stress and mood: Results of a pilot study. *Archives of Women's Mental Health, 11*(1), 67-74. doi:10.1007/s00737-008-0214-3
- Vispoel, W. P., Morris, C. A., & Kilinc, M. (2018). Applications of generalisability theory and their relations to classical test theory and structural equation modeling. *Psychological Methods, 23*(1), 1-26. doi:10.1037/met0000107

- Vøllestad, J., Sivertsen, B., & Nielsen, G. F. (2011). Mindfulness-Based Stress Reduction for patients with anxiety disorders: Evaluation in a randomized controlled trial. *Behaviour Research and Therapy*, *49*, 281– 288. doi:10.1016/j.brat.2011.01.007
- Walach, H., Buchheld, N., Buttenmuller, V., Kleinknecht, N., & Schmidt, S. (2006). Measuring mindfulness-the Freiburg Mindfulness Inventory (FMI). *Personality and Individual Differences*, *40*(8), 1543-1555.
- Wallace, B. A. (1999). The Buddhist tradition of Samatha: Methods for refining and examining consciousness. *Journal of Consciousness Studies*, *6*, 175-187.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measure of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070.
- Whimbey, A., Vaughan, G. M., & Tatsuoka, M. M. (1967). Fixed effects vs Random effects: Estimating variance components from mean squares. *Perceptual and Motor Skills*, *25*(2), 668–668. doi: 10.2466/pms.1967.25.2.668
- Whittle, P. (2013). Contrast brightness and ordinary seeing. In Gilchrist, A. L. (Ed.), *Lightness, brightness and transparency* (pp. 123-170). Psychology Press.
- Will, A., Rancea, M., Monsef, I., Wöckel, A., Engert, A., & Skoetz, N. (2015). Mindfulness-based stress reduction for women diagnosed with breast cancer. *Cochrane Database of Systematic Reviews*. doi:10.1002/14651858.cd011518
- Williams, M., Leumann, E., & Cappeller, C. (2004). *A Sanskrit-English dictionary, etymologically and philologically arranged with special reference to cognate Indo-European languages*. New Delhi: Bharatiya Granth Niketan.
- Wilson, J. D., Vo, H., Matson, P., Adger, H., Barnett, G., & Fishman, M. (2017). Trait mindfulness and progression to injection use in youth with opioid addiction. *Substance Use & Misuse*, *52*(11), 1486-1493. doi:10.1080/10826084.2017.1289225
- Wolever, R. Q., Bobinet, K. J., McCabe, K., Mackenzie, E. R., Fekete, E., Kusnick, C. A., & Baime, M. (2012). Effective and viable mind-body stress reduction in the workplace: A randomized controlled trial. *Journal of Occupational Health Psychology*, *17*, 246– 258.

- Wray, I (2004). The development of a state measure of mindfulness. Unpublished doctoral thesis, University of Sheffield, England. Retrieved from <http://etheses.whiterose.ac.uk/3072/1/411985.pdf>
- Yildiran, H., & Holt, R. R. (2015). Thematic analysis of the effectiveness of an inpatient mindfulness group for adults with intellectual disabilities. *British Journal of Learning Disabilities*, 43(1), 49–54.
- Zoogman, S., Goldberg, S. B., Hoyt, W. T., & Miller, L. (2015). Mindfulness interventions with youth: A meta-analysis. *Mindfulness*, 6(2), 290-302.

Appendix A1

Mindfulness
<https://doi.org/10.1007/s12671-020-01324-7>

ORIGINAL PAPER



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

Quoc C. Truong¹ · Christian U. Krägeloh² · Richard J. Siegert² · Jason Landon² · Oleg N. Medvedev¹

© 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

¹ School of Psychology, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand

² Department of Psychology, Auckland University of Technology, Auckland, New Zealand

Published online: 03 February 2020

Springer

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 μ 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:

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

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

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

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

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

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

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 ($\sigma_\delta^2 = \frac{\sigma_p^2}{n_i} + \frac{\sigma_{po}^2}{n_o} + \frac{\sigma_{pio}^2}{n_i n_o}$; 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 (Φ) coefficient, which is obtained after applying Whimby'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):

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

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., ≥ 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_r* and *G_a*) 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_r* and *G_a* 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)
Observe				
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>) ^a	–	0.74	0.74	0.77(0.69–0.84)
Describe				
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>) ^a	–	0.90	0.83	0.86(0.81–0.90)
Act with awareness				
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>) ^a	–	0.87	0.77	0.81(0.74–0.87)
Nonjudge				
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>) ^a	–	0.81	0.87	0.85(0.79–0.90)
Nonreact				
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>) ^a	–	0.75	0.64	0.71(0.61–0.79)
FFMQ-18				
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>) ^a	–	0.87	0.80	0.82(0.74–0.87)
FFMQ total				
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>) ^a	–	0.92	0.83	0.83(0.80–0.91)

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

^a 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	
	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^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	0.99		0.99		0.79		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 (≥ 0.80) reflecting high sensitivity for state changes over time. On the other end, there are nine items with low SCI (≤ 0.50) that are least sensitive to state changes and reflecting predominantly trait mindfulness. All other items had SCI between these benchmarks ($0.50 < \text{SCI} < 0.80$) 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
Observe				
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
Act with awareness				
38 <i>doing things without paying attention</i> ^R	0.01	0.00	0.30	0.95
13 I am easily distracted ^R	0.31	0.01	0.25	0.45
5 my mind wanders off and I'm easily distracted ^R	0.12	0.03	0.38	0.76
8 I do not pay attention to what I'm doing ^R	0.12	0.08	0.32	0.72
34 I do jobs or tasks automatically ^R	0.23	0.07	0.21	0.48
18 I find it difficult to stay focused ^R	0.06	0.02	0.38	0.86
28 I rush through activities without being attentive ^R	0.06	0.00	0.26	0.83
23 I am "running on automatic" ^R	0.16	0.05	0.26	0.62
Nonjudge				
25 I should not be thinking the way I'm thinking ^R	0.09	0.00	0.28	0.77
35 I judge myself as good or bad ^R	0.16	0.01	0.33	0.67
17 I make judgments about my thoughts ^R	0.16	0.00	0.34	0.67
30 I think my emotions are bad or inappropriate ^R	0.01	0.00	0.42	0.98
14 I believe my thoughts are abnormal or bad ^R	0.12	0.02	0.29	0.70
10 I should not be feeling the way I'm feeling ^R	0.12	0.03	0.25	0.68
39 I disapprove of myself ^R	0.13	0.00	0.29	0.69
3 I criticize myself for inappropriate emotions ^R	0.13	0.03	0.25	0.66
Describe				
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 ^R	0.04	0.02	0.32	0.89
16 I have trouble thinking of the right words ^R	0.23	0.03	0.21	0.48
7 I can easily put my thoughts into words ^R	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 ^R	0.24	0.01	0.34	0.58
Nonreact				
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	
	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^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 vary 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 (≥ 0.60) reflecting high sensitivity for state changes over time. On the other hand, the remaining fourteen items had SCI between the benchmarks ($0.30 < \text{SCI} < 0.60$) 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 (≤ 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 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., & Caims, 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.

Appendix A2

School of Psychology
The University of Waikato
Private Bag 3105
Hamilton, New Zealand

Phone DDI +64-7-856 2889
Facsimile 64-7-858 5132



24 June 2019

Quoc Cuong Truong,
C/- School of Psychology
University of Waikato
Private Bag 3105
Hamilton 3240

Dear James,

Ethics Approval Application – # 19:22

Title: Applying Generalisability Theory (G-Theory) to differentiate between trait and state aspects of mindfulness operationalised by the Five Facet Mindfulness Questionnaire (FFMQ)

Thank you for your ethics application submitted for approval which has been fully considered and approved by the Psychology Research and Ethics Committee on the condition that the amendment is made to the application as stated by Reviewer 1 – ‘Please remove Mohi Rua as a cultural advisor or a person providing cultural support – that was not /is not his role, he simply provided feedback about the application’.

Please note that approval is for three years.

If any modifications are required to your application, e.g., nature, content, location, procedures or personnel these will need to be submitted to the Convenor of the Committee.

I wish you success with your research.

Yours sincerely



Professor Nicola Starkey
Convenor
Psychology Research and Ethics Committee
School of Psychology
University of Waikato

Appendix B

Five Facet Mindfulness Questionnaire

Description:

This instrument is based on a factor analytic study of five independently developed mindfulness questionnaires. The analysis yielded five factors that appear to represent elements of mindfulness as it is currently conceptualized. The five facets are observing, describing, acting with awareness, non-judging of inner experience, and non-reactivity to inner experience. More information is available in:

Please rate each of the following statements using the scale provided. Write the number in the blank that best describes your own opinion of what is generally true for you.

1	2	3	4	5
never or very rarely true	rarely true	sometimes true	often true	very often or always true

- _____ 1. When I'm walking, I deliberately notice the sensations of my body moving.
- _____ 2. I'm good at finding words to describe my feelings.
- _____ 3. I criticize myself for having irrational or inappropriate emotions.
- _____ 4. I perceive my feelings and emotions without having to react to them.
- _____ 5. When I do things, my mind wanders off and I'm easily distracted.
- _____ 6. When I take a shower or bath, I stay alert to the sensations of water on my body.
- _____ 7. I can easily put my beliefs, opinions, and expectations into words.
- _____ 8. I don't pay attention to what I'm doing because I'm daydreaming, worrying, or otherwise distracted.
- _____ 9. I watch my feelings without getting lost in them.
- _____ 10. I tell myself I shouldn't be feeling the way I'm feeling.

- _____ 11. I notice how foods and drinks affect my thoughts, bodily sensations, and emotions.
- _____ 12. It's hard for me to find the words to describe what I'm thinking.
- _____ 13. I am easily distracted.
- _____ 14. I believe some of my thoughts are abnormal or bad and I shouldn't think that way.
- _____ 15. I pay attention to sensations, such as the wind in my hair or sun on my face.
- _____ 16. I have trouble thinking of the right words to express how I feel about things
- _____ 17. I make judgments about whether my thoughts are good or bad.
- _____ 18. I find it difficult to stay focused on what's happening in the present.
- _____ 19. When I have distressing thoughts or images, I "step back" and am aware of the thought or image without getting taken over by it.
- _____ 20. I pay attention to sounds, such as clocks ticking, birds chirping, or cars passing.
- _____ 21. In difficult situations, I can pause without immediately reacting.
- _____ 22. When I have a sensation in my body, it's difficult for me to describe it because I can't find the right words.
- _____ 23. It seems I am "running on automatic" without much awareness of what I'm doing.
- _____ 24. When I have distressing thoughts or images, I feel calm soon after.
- _____ 25. I tell myself that I shouldn't be thinking the way I'm thinking.
- _____ 26. I notice the smells and aromas of things.
- _____ 27. Even when I'm feeling terribly upset, I can find a way to put it into words.
- _____ 28. I rush through activities without being really attentive to them.
- _____ 29. When I have distressing thoughts or images I am able just to notice them without reacting.
- _____ 30. I think some of my emotions are bad or inappropriate and I shouldn't feel them.
- _____ 31. I notice visual elements in art or nature, such as colors, shapes, textures, or patterns of light and shadow.

- _____ 32. My natural tendency is to put my experiences into words.
- _____ 33. When I have distressing thoughts or images, I just notice them and let them go.
- _____ 34. I do jobs or tasks automatically without being aware of what I'm doing.
- _____ 35. When I have distressing thoughts or images, I judge myself as good or bad, depending what the thought/image is about.
- _____ 36. I pay attention to how my emotions affect my thoughts and behavior.
- _____ 37. I can usually describe how I feel at the moment in considerable detail.
- _____ 38. I find myself doing things without paying attention.
- _____ 39. I disapprove of myself when I have irrational ideas.

Scoring Information:

Observe items:

1, 6, 11, 15, 20, 26, 31, 36

Describe items:

2, 7, 12R, 16R, 22R, 27, 32, 37

Act with Awareness items:

5R, 8R, 13R, 18R, 23R, 28R, 34R, 38R

Nonjudge items:

3R, 10R, 14R, 17R, 25R, 30R, 35R, 39R

Nonreact items:

4, 9, 19, 21, 24, 29, 33

Reference:

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

Appendix C1

EduG analyses output for the total FFMQ, including observation and estimation designs, ANOVA and G-study tables.

Observation and Estimation Designs								
Facet	Label	Levels	Univ.	Reduction (levels to exclude)				
Person	P	83	INF					
Item	I	39	39					
Occasion	O	3	INF					

Analysis of variance								
Source	SS	df	MS	Components				
				Random	Mixed	Corrected	%	SE
P	563.199	82	6.868	0.049	0.052	0.052	4.7	0.009
I	97.985	38	2.579	-0.017	-0.017	-0.017	0.0	0.005
O	3.726	2	1.863	-0.001	0.000	0.000	0.0	0.001
PI	3814.938	3116	1.224	0.120	0.120	0.120	10.6	0.012
PO	119.248	164	0.727	-0.004	0.019	0.019	1.7	0.002
IO	492.724	76	6.483	0.068	0.068	0.068	6.0	0.013
PIO	5392.302	6232	0.865	0.865	0.865	0.865	77.0	0.015
Total	10484.122	9710					100	
							%	

G Study Table (Measurement design P/IO)						
Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.052		
	I		(0.000)	0.0
	O		0.000	1.8
	PI	(0.000)	0.0	(0.000)	0.0
	PO	0.006	100.0	0.006	98.2
	IO		(0.000)	0.0
	PIO	(0.000)	0.0	(0.000)	0.0
Sum of variances	0.052		0.006	100%	0.006	100%
Standard deviation	0.229		Relative SE: 0.079		Absolute SE: 0.080	
Coef_G relative	0.89					
Coef_G absolute	0.89					

Grand mean for levels used: 3.189

Variance error of the mean for levels used: 0.001

Standard error of the grand mean: 0.029

Appendix C2

EduG analyses output for the FFMQ-18, including observation and estimation designs, ANOVA and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 3 4 5 6 8 11 13 14 16 20 21 25 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

Analysis of variance

Source	SS	df	MS	Components				SE
				Random	Mixed	Corrected	%	
P	332.685	82	4.057	0.054	0.057	0.057	5.0	0.012
I	73.608	17	4.330	-0.003	-0.003	-0.003	0.0	0.007
O	4.782	2	2.391	-0.002	0.000	0.000	0.0	0.001
PI	1725.115	1394	1.238	0.112	0.112	0.112	9.9	0.018
PO	128.996	164	0.787	-0.006	0.017	0.017	1.5	0.005
IO	165.298	34	4.862	0.048	0.048	0.048	4.2	0.014
PIO	2511.591	2788	0.901	0.901	0.901	0.901	79.4	0.024
Total	4942.073	4481					100 %	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.057		
	I		(0.000)	0.0
	O		(0.000)	0.0
	PI	0.003	18.9	0.003	18.4
	PO	0.006	30.6	0.006	29.8
	IO		0.000	2.6
	PIO	0.009	50.5	0.009	49.2
Sum of variances	0.057		0.018	100%	0.019	100%
Standard deviation	0.239		Relative SE: 0.135		Absolute SE: 0.137	
Coef_G relative	0.76					
Coef_G absolute	0.75					

Grand mean for levels used: 3.182

Variance error of the mean for levels used: 0.001

Standard error of the grand mean: 0.037

Appendix C3

EduG analyses output for “observe” subscale, including observation and estimation designs, ANOVA and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	2 3 4 5 7 8 9 10 12 13 14 16 17 18 19 21 22 23 24 25 27 28 29 30 32 33 34 35 37 38 39
Occasion	O	3	INF	

Analysis of variance

Source	SS	df	MS	Components				SE
				Random	Mixed	Corrected	%	
P	138.530	82	1.689	0.020	0.024	0.024	2.1	0.012
I	14.386	7	2.055	-0.023	-0.023	-0.023	0.0	0.011
O	4.314	2	2.157	-0.008	-0.006	-0.006	0.0	0.005
PI	756.281	574	1.318	0.154	0.154	0.154	13.8	0.028
PO	123.102	164	0.751	-0.013	0.009	0.009	0.8	0.011
IO	102.858	14	7.347	0.078	0.078	0.078	7.0	0.031
PIO	981.725	1148	0.855	0.855	0.855	0.855	76.3	0.036
Total	2121.197	1991					100 %	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.024		
	I		(0.000)	0.0
	O		(0.000)	0.0
	PI	0.016	32.9	0.016	31.2
	PO	0.003	6.2	0.003	5.9
	IO		0.003	5.3
	PIO	0.029	60.9	0.029	57.7
Sum of variances	0.024		0.048	100%	0.050	100%
Standard deviation	0.154		Relative SE: 0.218		Absolute SE: 0.224	
Coef_G relative	0.33					
Coef_G absolute	0.32					

Grand mean for levels used: 3.147

Variance error of the mean for levels used: 0.004

Standard error of the grand mean: 0.059

Appendix C4

EduG analyses output for “describe” subscale, including observation and estimation designs, ANOVA and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 3 4 5 6 8 9 10 11 13 14 15 17 18 19 20 21 23 24 25 26 28 29 30 31 33 34 35 36 38 39
Occasion	O	3	INF	

Analysis of variance

Source	SS	df	MS	Components				SE
				Random	Mixed	Corrected	%	
P	177.526	82	2.165	0.032	0.037	0.037	3.3	0.015
I	27.261	7	3.894	-0.012	-0.012	-0.011	0.0	0.011
O	1.088	2	0.544	-0.009	-0.007	-0.007	0.0	0.003
PI	784.281	574	1.366	0.189	0.189	0.189	17.0	0.029
PO	135.995	164	0.829	0.004	0.024	0.024	2.2	0.012
IO	86.807	14	6.201	0.065	0.065	0.065	5.8	0.026
PIO	916.776	1148	0.799	0.799	0.799	0.799	71.7	0.033
Total	2129.734	1991					100 %	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.037		
	I		(0.000)	0.0
	O		(0.000)	0.0
	PI	0.019	35.4	0.019	34.0
	PO	0.008	14.9	0.008	14.3
	IO		0.002	3.9
	PIO	0.027	49.8	0.027	47.8
Sum of variances	0.037		0.055	100%	0.057	100%
Standard deviation	0.192		Relative SE: 0.234		Absolute SE: 0.238	
Coef_G relative	0.40					
Coef_G absolute	0.39					

Grand mean for levels used: 3.178

Variance error of the mean for levels used: 0.003

Standard error of the grand mean: 0.058

Appendix C5

EduG analyses output for “act with awareness” subscale, including observation and estimation designs, ANOVA and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 2 3 4 6 7 9 10 11 12 14 15 16 17 19 20 21 22 24 25 26 27 29 30 31 32 33 35 36 37 39
Occasion	O	3	INF	

Analysis of variance

Source	SS	df	MS	Components				SE
				Random	Mixed	Corrected	%	
P	180.262	82	2.198	0.039	0.041	0.041	3.6	0.015
I	20.497	7	2.928	-0.028	-0.028	-0.027	0.0	0.015
O	7.519	2	3.760	-0.009	-0.006	-0.006	0.0	0.006
PI	667.128	574	1.162	0.096	0.096	0.096	8.4	0.026
PO	161.148	164	0.983	0.014	0.036	0.036	3.1	0.014
IO	133.790	14	9.556	0.105	0.105	0.105	9.1	0.041
PIO	1002.210	1148	0.873	0.873	0.873	0.873	75.8	0.036
Total	2172.554	1991					100 %	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.041		
	I		(0.000)	0.0
	O		(0.000)	0.0
	PI	0.010	19.1	0.010	17.8
	PO	0.012	23.3	0.012	21.8
	IO		0.004	6.5
	PIO	0.030	57.6	0.030	53.9
Sum of variances	0.041		0.052	100%	0.055	100%
Standard deviation	0.203		Relative SE: 0.227		Absolute SE: 0.235	
Coef_G relative	0.44					
Coef_G absolute	0.43					

Grand mean for levels used: 3.174

Variance error of the mean for levels used: 0.005

Standard error of the grand mean: 0.068

Appendix C6

EduG analyses output for “nonjudge” subscale, including observation and estimation designs, ANOVA and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 2 4 5 6 7 8 9 11 12 13 15 16 18 19 20 21 22 23 24 26 27 28 29 31 32 33 34 36 37 38
Occasion	O	3	INF	

Analysis of variance

Source	SS	df	MS	Components				SE
				Random	Mixed	Corrected	%	
P	160.716	82	1.960	0.033	0.035	0.035	3.2	0.014
I	11.436	7	1.634	-0.011	-0.011	-0.011	0.0	0.007
O	2.694	2	1.347	-0.004	-0.003	-0.003	0.0	0.003
PI	667.148	574	1.162	0.082	0.082	0.082	7.5	0.026
PO	151.140	164	0.922	0.001	0.024	0.024	2.2	0.014
IO	58.543	14	4.182	0.039	0.039	0.039	3.6	0.018
PIO	1051.623	1148	0.916	0.916	0.916	0.916	83.5	0.038
Total	2103.299	1991					100 %	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.035		
	I		(0.000)	0.0
	O		(0.000)	0.0
	PI	0.008	17.6	0.008	17.1
	PO	0.008	16.9	0.008	16.5
	IO		0.001	2.7
	PIO	0.031	65.5	0.031	63.7
Sum of variances	0.035		0.048	100%	0.049	100%
Standard deviation	0.187		Relative SE: 0.218		Absolute SE: 0.221	
Coef_G relative	0.42					
Coef_G absolute	0.42					

Grand mean for levels used: 3.260

Variance error of the mean for levels used: 0.002

Standard error of the grand mean: 0.048

Appendix C7

EduG analyses output for “nonreact” subscale, including observation and estimation designs, ANOVA and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 2 3 5 6 7 8 10 11 12 13 14 15 16 17 18 20 22 23 25 26 27 28 30 31 32 34 35 36 37 38 39
Occasion	O	3	INF	

Analysis of variance

Source	SS	df	MS	Components				SE
				Random	Mixed	Corrected	%	
P	172.836	82	2.108	0.025	0.030	0.030	2.6	0.017
I	10.095	6	1.683	-0.021	-0.021	-0.021	0.0	0.010
O	21.674	2	10.837	0.007	0.009	0.009	0.8	0.014
PI	673.429	492	1.369	0.179	0.179	0.179	15.4	0.032
PO	170.516	164	1.040	0.030	0.051	0.051	4.4	0.017
IO	77.161	12	6.430	0.067	0.067	0.067	5.8	0.029
PIO	817.315	984	0.831	0.831	0.831	0.831	71.1	0.037
Total	1943.027	1742					100 %	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.030		
	I		(0.000)	0.0
	O		0.003	3.8
	PI	0.022	30.0	0.022	27.8
	PO	0.017	23.7	0.017	22.0
	IO		0.003	3.5
	PIO	0.033	46.3	0.033	42.9
Sum of variances	0.030		0.072	100%	0.078	100%
Standard deviation	0.173		Relative SE: 0.268		Absolute SE: 0.279	
Coef_G relative	0.29					
Coef_G absolute	0.28					

Grand mean for levels used: 3.187

Variance error of the mean for levels used: 0.007

Standard error of the grand mean: 0.083

Appendix C8

EduG analyses output for Item 1 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.04398		
	I	
	O		0.03389	9.0
	PI	
	PO	0.34094	100.0	0.34094	91.0
	IO	
	PIO	
Sum of variances	0.04398		0.34094	100%	0.37483	100%
Standard deviation	0.20972		Relative SE: 0.58390		Absolute SE: 0.61224	
Coef_G relative	0.11					
Coef_G absolute	0.11					

Grand mean for levels used: 3.13655

Variance error of the mean for levels used: 0.03853

Standard error of the grand mean: 0.19629

Appendix C9

EduG analyses output for Item 2 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.07057		
	I	
	O		0.00047	0.2
	PI	
	PO	0.30742	100.0	0.30742	99.8
	IO	
	PIO	
Sum of variances	0.07057		0.30742	100%	0.30790	100%
Standard deviation	0.26566		Relative SE: 0.55446		Absolute SE: 0.55489	
Coef_G relative	0.19					
Coef_G absolute	0.19					

Grand mean for levels used: 2.94779

Variance error of the mean for levels used: 0.00503

Standard error of the grand mean: 0.07091

Appendix C10

EduG analyses output for Item 3 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 10 11 12 13 14 15 16 17 18 20 22 23 25 26 27 28 30 31 32 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute	
P	0.03568			
	I		(0.00000)	0.0	
	O		0.00553	6.7	
	PI		0.02742	37.3	0.02742	33.3
	PO		0.00835	11.4	0.00835	10.1
	IO			0.00337	4.1
	PIO		0.03768	51.3	0.03768	45.8
Sum of variances	0.03568		0.07345	100%	0.08235	100%	
Standard deviation	0.18888		Relative SE: 0.27103		Absolute SE: 0.28697		
Coef_G relative	0.33						
Coef_G absolute	0.30						

Grand mean for levels used: 3.18273

Variance error of the mean for levels used: 0.01021

Standard error of the grand mean: 0.10107

Appendix C11

EduG analyses output for Item 4 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.11074		
	I	
	O		0.00488	1.1
	PI	
	PO	0.43555	100.0	0.43555	98.9
	IO	
	PIO	
Sum of variances	0.11074		0.43555	100%	0.44043	100%
Standard deviation	0.33277		Relative SE: 0.65996		Absolute SE: 0.66365	
Coef_G relative	0.20					
Coef_G absolute	0.20					

Grand mean for levels used: 3.21285

Variance error of the mean for levels used: 0.01146

Standard error of the grand mean: 0.10707

Appendix C12

EduG analyses output for Item 5 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.12024		
	I	
	O		0.03339	8.1
	PI	
	PO	0.37893	100.0	0.37893	91.9
	IO	
	PIO	
Sum of variances	0.12024		0.37893	100%	0.41232	100%
Standard deviation	0.34675		Relative SE: 0.61557		Absolute SE: 0.64212	
Coef_G relative	0.24					
Coef_G absolute	0.23					

Grand mean for levels used: 3.18876
 Variance error of the mean for levels used: 0.03940
 Standard error of the grand mean: 0.19849

Appendix C13

EduG analyses output for Item 6 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute	
P	0.22994			
	I		
	O		0.00962	4.1	
	PI		
	PO		0.22332	100.0	0.22332	95.9
	IO		
	PIO			
Sum of variances	0.22994		0.22332	100%	0.23293	100%	
Standard deviation	0.47952		Relative SE: 0.47256		Absolute SE: 0.48263		
Coef_G relative	0.51						
Coef_G absolute	0.50						

Grand mean for levels used: 3.18474

Variance error of the mean for levels used: 0.01508; Standard error of the grand mean: 0.12279

Appendix C14

EduG analyses output for Item 7 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.35508		
	I	
	O		0.00359	1.9
	PI	
	PO	0.18918	100.0	0.18918	98.1
	IO	
	PIO	
Sum of variances	0.35508		0.18918	100%	0.19277	100%
Standard deviation	0.59588		Relative SE: 0.43495		Absolute SE: 0.43906	
Coef_G relative	0.65					
Coef_G absolute	0.65					

Grand mean for levels used: 3.12851

Variance error of the mean for levels used: 0.01015

Standard error of the grand mean: 0.10074

Appendix C15

EduG analyses output for Item 8 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.12322		
	I	
	O		0.07624	19.0
	PI	
	PO	0.32403	100.0	0.32403	81.0
	IO	
	PIO	
Sum of variances	0.12322		0.32403	100%	0.40027	100%
Standard deviation	0.35103		Relative SE: 0.56923		Absolute SE: 0.63267	
Coef_G relative	0.28					
Coef_G absolute	0.24					

Grand mean for levels used: 3.19679

Variance error of the mean for levels used: 0.08163

Standard error of the grand mean: 0.28571

Appendix C16

EduG analyses output for Item 9 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.22093		
	I	
	O		0.02264	6.0
	PI	
	PO	0.35353	100.0	0.35353	94.0
	IO	
	PIO	
Sum of variances	0.22093		0.35353	100%	0.37617	100%
Standard deviation	0.47003		Relative SE: 0.59458		Absolute SE: 0.61333	
Coef_G relative	0.38					
Coef_G absolute	0.37					

Grand mean for levels used: 3.32530

Variance error of the mean for levels used: 0.02956

Standard error of the grand mean: 0.17194

Appendix C17

EduG analyses output for Item 10 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 2 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.11549		
	I	
	O		0.02723	9.8
	PI	
	PO	0.24988	100.0	0.24988	90.2
	IO	
	PIO	
Sum of variances	0.11549		0.24988	100%	0.27711	100%
Standard deviation	0.33983		Relative SE: 0.49988		Absolute SE: 0.52641	
Coef_G relative	0.32					
Coef_G absolute	0.29					

Grand mean for levels used: 3.27711

Variance error of the mean for levels used: 0.03163

Standard error of the grand mean: 0.17786

Appendix C18

EduG analyses output for Item 11 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.17788		
	I	
	O		0.03304	10.9
	PI	
	PO	0.27084	100.0	0.27084	89.1
	IO	
	PIO	
Sum of variances	0.17788		0.27084	100%	0.30388	100%
Standard deviation	0.42176		Relative SE: 0.52042		Absolute SE: 0.55126	
Coef_G relative	0.40					
Coef_G absolute	0.37					

Grand mean for levels used: 3.21285

Variance error of the mean for levels used: 0.03845

Standard error of the grand mean: 0.19608

Appendix C19

EduG analyses output for Item 12 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.03835		
	I	
	O		0.02359	6.9
	PI	
	PO	0.31911	100.0	0.31911	93.1
	IO	
	PIO	
Sum of variances	0.03835		0.31911	100%	0.34270	100%
Standard deviation	0.19583		Relative SE: 0.56490		Absolute SE: 0.58541	
Coef_G relative	0.11					
Coef_G absolute	0.10					

Grand mean for levels used: 3.34940

Variance error of the mean for levels used: 0.02790

Standard error of the grand mean: 0.16702

Appendix C20

EduG analyses output for Item 13 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.31164		
	I	
	O		0.00968	3.7
	PI	
	PO	0.25002	100.0	0.25002	96.3
	IO	
	PIO	
Sum of variances	0.31164		0.25002	100%	0.25971	100%
Standard deviation	0.55824		Relative SE: 0.50002		Absolute SE: 0.50961	
Coef_G relative	0.55					
Coef_G absolute	0.55					

Grand mean for levels used: 3.18474
 Variance error of the mean for levels used: 0.01645
 Standard error of the grand mean: 0.12825

Appendix C21

EduG analyses output for Item 14 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.12425		
	I	
	O		0.01866	6.1
	PI	
	PO	0.28790	100.0	0.28790	93.9
	IO	
.....	PIO		
Sum of variances	0.12425		0.28790	100%	0.30656	100%
Standard deviation	0.35250		Relative SE: 0.53656		Absolute SE: 0.55368	
Coef_G relative	0.30					
Coef_G absolute	0.29					

Grand mean for levels used: 3.08835

Variance error of the mean for levels used: 0.02363

Standard error of the grand mean: 0.15371

Appendix C22

EduG analyses output for Item 15 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 2 3 4 5 6 7 8 9 10 11 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.03835		
	I	
	O		0.02359	6.9
	PI	
	PO	0.31911	100.0	0.31911	93.1
	IO	
	PIO	
Sum of variances	0.03835		0.31911	100%	0.34270	100%
Standard deviation	0.19583		Relative SE: 0.56490		Absolute SE: 0.58541	
Coef_G relative	0.11					
Coef_G absolute	0.10					

Grand mean for levels used: 3.34940
 Variance error of the mean for levels used: 0.02790
 Standard error of the grand mean: 0.16702

Appendix C23

EduG analyses output for Item 16 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute	
P	0.22661			
	I		
	O		0.03430	14.0	
	PI		
	PO		0.21068	100.0	0.21068	86.0
	IO	
.....	PIO		
Sum of variances	0.22661		0.21068	100%	0.24498	100%	
Standard deviation	0.47604		Relative SE: 0.45900		Absolute SE: 0.49495		
Coef_G relative	0.52						
Coef_G absolute	0.48						

Grand mean for levels used: 3.17269

Variance error of the mean for levels used: 0.03957

Standard error of the grand mean: 0.19892

Appendix C24

EduG analyses output for Item 17 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.16196		
	I	
	O		0.00498	1.5
	PI	
	PO	0.33505	100.0	0.33505	98.5
	IO	
	PIO	
Sum of variances	0.16196		0.33505	100%	0.34003	100%
Standard deviation	0.40245		Relative SE: 0.57883		Absolute SE: 0.58312	
Coef_G relative	0.33					
Coef_G absolute	0.32					

Grand mean for levels used: 3.29317

Variance error of the mean for levels used: 0.01097

Standard error of the grand mean: 0.10473

Appendix C25

EduG analyses output for Item 18 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.06323		
	I	
	O		0.01794	4.6
	PI	
	PO	0.37563	100.0	0.37563	95.4
	IO	
	PIO	
Sum of variances	0.06323		0.37563	100%	0.39357	100%
Standard deviation	0.25145		Relative SE: 0.61289		Absolute SE: 0.62736	
Coef_G relative	0.14					
Coef_G absolute	0.14					

Grand mean for levels used: 3.20080

Variance error of the mean for levels used: 0.02323

Standard error of the grand mean: 0.15241

Appendix C26

EduG analyses output for Item 19 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute	
P	0.35674			
	I		
	O		0.01270	6.2	
	PI		
	PO		0.19212	100.0	0.19212	93.8
	IO		
.....	PIO			
Sum of variances	0.35674		0.19212	100%	0.20482	100%	
Standard deviation	0.59728		Relative SE: 0.43831		Absolute SE: 0.45257		
Coef_G relative	0.65						
Coef_G absolute	0.64						

Grand mean for levels used: 3.23695

Variance error of the mean for levels used: 0.01931

Standard error of the grand mean: 0.13897

Appendix C27

EduG analyses output for Item 20 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.34499		
	I	
	O		0.01146	5.2
	PI	
	PO	0.20942	100.0	0.20942	94.8
	IO	
	PIO	
Sum of variances	0.34499		0.20942	100%	0.22088	100%
Standard deviation	0.58736		Relative SE: 0.45763		Absolute SE: 0.46998	
Coef_G relative	0.62					
Coef_G absolute	0.61					

Grand mean for levels used: 3.12048

Variance error of the mean for levels used: 0.01814

Standard error of the grand mean: 0.13469

Appendix C28

EduG analyses output for Item 21 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolut e
P	0.17127		
	I	
	O		0.09745	28.7
	PI	
	PO	0.24258	100.0	0.24258	71.3
	IO	
	PIO	
Sum of variances	0.17127		0.24258	100%	0.34003	100%
Standard deviation	0.41385		Relative SE: 0.49252		Absolute SE: 0.58312	
Coef_G relative	0.41					
Coef_G absolute	0.33					

Grand mean for levels used: 3.12450

Variance error of the mean for levels used: 0.10243

Standard error of the grand mean: 0.32005

Appendix C29

EduG analyses output for Item 22 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.24140		
	I	
	O		0.01282	3.7
	PI	
	PO	0.33792	100.0	0.33792	96.3
	IO	
	PIO	
Sum of variances	0.24140		0.33792	100%	0.35074	100%
Standard deviation	0.49133		Relative SE: 0.58131		Absolute SE: 0.59223	
Coef_G relative	0.42					
Coef_G absolute	0.41					

Grand mean for levels used: 3.28916
 Variance error of the mean for levels used: 0.01980
 Standard error of the grand mean: 0.14070

Appendix C30

EduG analyses output for Item 23 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute	
P	0.16035			
	I		
	O		0.05296	17.1	
	PI		
	PO		0.25628	100.0	0.25628	82.9
	IO	
.....	PIO		
Sum of variances	0.16035		0.25628	100%	0.30924	100%	
Standard deviation	0.40044		Relative SE: 0.50624		Absolute SE: 0.55609		
Coef_G relative	0.38						
Coef_G absolute	0.34						

Grand mean for levels used: 3.23293

Variance error of the mean for levels used: 0.05798

Standard error of the grand mean: 0.24079

Appendix C31

EduG analyses output for Item 24 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute	
P	0.27686			
	I		
	O		0.01553	5.9	
	PI		
	PO		0.24686	100.0	0.24686	94.1
	IO		
.....	PIO			
Sum of variances	0.27686		0.24686	100%	0.26238	100%	
Standard deviation	0.52618		Relative SE: 0.49685		Absolute SE: 0.51223		
Coef_G relative	0.53						
Coef_G absolute	0.51						

Grand mean for levels used: 3.18474
 Variance error of the mean for levels used: 0.02184
 Standard error of the grand mean: 0.14777

Appendix C32

EduG analyses output for Item 25 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 27 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolut e
P	0.08669		
	I	
	O		0.00330	1.1
	PI	
	PO	0.28452	100.0	0.28452	98.9
	IO	
.....	PIO		
Sum of variances	0.08669		0.28452	100%	0.28782	100%
Standard deviation	0.29443		Relative SE: 0.53340		Absolute SE: 0.53649	
Coef_G relative	0.23					
Coef_G absolute	0.23					

Grand mean for levels used: 3.32530

Variance error of the mean for levels used: 0.00777

Standard error of the grand mean: 0.08815

Appendix C33

EduG analyses output for Item 26 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 27 28 29 30 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.25928		
	I	
	O		0.01992	6.0
	PI	
	PO	0.31074	100.0	0.31074	94.0
	IO	
	PIO	
Sum of variances	0.25928		0.31074	100%	0.33066	100%
Standard deviation	0.50920		Relative SE: 0.55744		Absolute SE: 0.57503	
Coef_G relative	0.45					
Coef_G absolute	0.44					

Grand mean for levels used: 3.15663

Variance error of the mean for levels used: 0.02678

Standard error of the grand mean: 0.16366

Appendix C34

EduG analyses output for Item 27 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 28 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.23685		
	I	
	O		0.02762	8.5
	PI	
	PO	0.29902	100.0	0.29902	91.5
	IO	
.....	PIO		
Sum of variances	0.23685		0.29902	100%	0.32664	100%
Standard deviation	0.48667		Relative SE: 0.54682		Absolute SE: 0.57152	
Coef_G relative	0.44					
Coef_G absolute	0.42					

Grand mean for levels used: 3.10040
 Variance error of the mean for levels used: 0.03408
 Standard error of the grand mean: 0.18460

Appendix C35

EduG analyses output for Item 28 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 29 30 31 32 33 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolut e
P	0.05559		
	I	
	O		(0.00000)	0.0
	PI	
	PO	0.26377	100.0	0.26377	100.0
	IO	
	PIO	
Sum of variances	0.05559		0.26377	100%	0.26377	100%
Standard deviation	0.23577		Relative SE: 0.51359		Absolute SE: 0.51359	
Coef_G relative	0.17					
Coef_G absolute	0.17					

Grand mean for levels used: 2.91566
 Variance error of the mean for levels used: 0.00385
 Standard error of the grand mean: 0.06203

Appendix C36

EduG analyses output for Item 29 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 30 31 32 33 34 35 36 37 38 39
	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute	
P	0.12254			
	I		
	O		0.00855	3.1	
	PI		
	PO		0.27123	100.0	0.27123	96.9
	IO	
.....	PIO		
Sum of variances	0.12254		0.27123	100%	0.27979	100%	
Standard deviation	0.35006		Relative SE: 0.52080		Absolute SE: 0.52895		
Coef_G relative	0.31						
Coef_G absolute	0.30						

Grand mean for levels used: 3.15261

Variance error of the mean for levels used: 0.01330

Standard error of the grand mean: 0.11532

Appendix C37

EduG analyses output for Item 30 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 31 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.00970		
	I	
	O		0.00379	0.9
	PI	
	PO	0.42058	100.0	0.42058	99.1
	IO	
	PIO	
Sum of variances	0.00970		0.42058	100%	0.42436	100%
Standard deviation	0.09848		Relative SE: 0.64852		Absolute SE: 0.65143	
Coef_G relative	0.02					
Coef_G absolute	0.02					

Grand mean for levels used: 3.28916
 Variance error of the mean for levels used: 0.00897
 Standard error of the grand mean: 0.09472

Appendix C38

EduG analyses output for Item 31 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 32 33 34 35 36 37 38 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.21006		
	I	
	O		0.02274	6.7
	PI	
	PO	0.31729	100.0	0.31729	93.3
	IO	
	PIO	
Sum of variances	0.21006		0.31729	100%	0.34003	100%
Standard deviation	0.45832		Relative SE: 0.56328		Absolute SE: 0.58312	
Coef_G relative	0.40					
Coef_G absolute	0.38					

Grand mean for levels used: 3.25301
Variance error of the mean for levels used: 0.02909
Standard error of the grand mean: 0.17057

Appendix C39

EduG analyses output for Item 32 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 33 34 35 36 37 38 39
	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute	
P	0.36512			
	I		
	O		0.02287	10.7	
	PI		
	PO		0.19132	100.0	0.19132	89.3
	IO		
.....	PIO			
Sum of variances	0.36512		0.19132	100%	0.21419	100%	
Standard deviation	0.60425		Relative SE: 0.43740		Absolute SE: 0.46281		
Coef_G relative	0.66						
Coef_G absolute	0.63						

Grand mean for levels used: 3.25703

Variance error of the mean for levels used: 0.02958

Standard error of the grand mean: 0.17198

Appendix C40

EduG analyses output for Item 33 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 34 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.17323		
	I	
	O		0.01252	4.5
	PI	
	PO	0.26593	100.0	0.26593	95.5
	IO	
	PIO	
Sum of variances	0.17323		0.26593	100%	0.27845	100%
Standard deviation	0.41621		Relative SE: 0.51568		Absolute SE: 0.52768	
Coef_G relative	0.39					
Coef_G absolute	0.38					

Grand mean for levels used: 3.07229

Variance error of the mean for levels used: 0.01781

Standard error of the grand mean: 0.13346

Appendix C41

EduG analyses output for Item 34 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 35 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.23107		
	I	
	O		0.06524	23.4
	PI	
	PO	0.21321	100.0	0.21321	76.6
	IO	
	PIO	
Sum of variances	0.23107		0.21321	100%	0.27845	100%
Standard deviation	0.48070		Relative SE: 0.46175		Absolute SE: 0.52768	
Coef_G relative	0.52					
Coef_G absolute	0.45					

Grand mean for levels used: 3.20080

Variance error of the mean for levels used: 0.07059

Standard error of the grand mean: 0.26569

Appendix C42

EduG analyses output for Item 35 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 36 37 38 39
	O	3	INF	
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.16373		
	I	
	O		0.00926	2.8
	PI	
	PO	0.32542	100.0	0.32542	97.2
	IO	
	PIO	
Sum of variances	0.16373		0.32542	100%	0.33467	100%
Standard deviation	0.40463		Relative SE: 0.57045		Absolute SE: 0.57851	
Coef_G relative	0.33					
Coef_G absolute	0.33					

Grand mean for levels used: 3.32932

Variance error of the mean for levels used: 0.01515

Standard error of the grand mean: 0.12308

Appendix C43

EduG analyses output for Item 36 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 37 38 39
	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.08272		
	I	
	O		0.05631	18.3
	PI	
	PO	0.25159	100.0	0.25159	81.7
	IO	
	PIO	
Sum of variances	0.08272		0.25159	100%	0.30790	100%
Standard deviation	0.28761		Relative SE: 0.50159		Absolute SE: 0.55489	
Coef_G relative	0.25					
Coef_G absolute	0.21					

Grand mean for levels used: 3.16064
 Variance error of the mean for levels used: 0.06033
 Standard error of the grand mean: 0.24563

Appendix C44

EduG analyses output for Item 37 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 38 39
	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.23602		
	I	
	O		0.02545	8.2
	PI	
	PO	0.28512	100.0	0.28512	91.8
	IO	
.....	PIO		
Sum of variances	0.23602		0.28512	100%	0.31058	100%
Standard deviation	0.48582		Relative SE: 0.53397		Absolute SE: 0.55729	
Coef_G relative	0.45					
Coef_G absolute	0.43					

Grand mean for levels used: 3.18072

Variance error of the mean for levels used: 0.03173

Standard error of the grand mean: 0.17813

Appendix C45

EduG analyses output for Item 38 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolut e	
P	0.01479			
	I		
	O		0.00121	0.4	
	PI		
	PO		0.30267	100.0	0.30267	99.6
	IO		
	PIO		
Sum of variances	0.01479		0.30267	100%	0.30388	100%	
Standard deviation	0.12162		Relative SE: 0.55016		Absolute SE: 0.55126		
Coef_G relative	0.05						
Coef_G absolute	0.05						

Grand mean for levels used: 3.27309

Variance error of the mean for levels used: 0.00503

Standard error of the grand mean: 0.07094

Appendix C46

EduG analyses output for Item 39 of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute	
P	0.12704			
	I		
	O		(0.00000)	0.0	
	PI		
	PO		0.28795	100.0	0.28795	100.0
	IO		
.....	PIO			
Sum of variances	0.12704		0.28795	100%	0.28795	100%	
Standard deviation	0.35643		Relative SE: 0.53661		Absolute SE: 0.53661		
Coef_G relative	0.31						
Coef_G absolute	0.31						

Grand mean for levels used: 3.28514
 Variance error of the mean for levels used: 0.00500
 Standard error of the grand mean: 0.07071

Appendix C47

EduG analyses output for the combination of the highest State Component Index items from each subscale of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	2 3 5 6 7 8 9 10 11 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 31 32 33 34 35 36 37 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.10958		
	I		(0.00000)	0.0
	O		0.00783	10.1
	PI	(0.00000)	0.0	(0.00000)	0.0
	PO	(0.00000)	0.0	(0.00000)	0.0
	IO		0.00104	1.3
	PIO	0.06900	100.0	0.06900	88.6
Sum of variances	0.10958		0.06900	100%	0.07787	100%
Standard deviation	0.33103		Relative SE: 0.26267		Absolute SE: 0.27904	
Coef_G relative	0.61					
Coef_G absolute	0.58					

Grand mean for levels used: 3.25221

Variance error of the mean for levels used: 0.01102

Standard error of the grand mean: 0.10497

Appendix C48

EduG analyses output for the combination of the five highest State Component Index items of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	2 3 4 5 6 7 9 10 11 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 31 32 33 34 35 36 37 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.09142		
	I		(0.00000)	0.0
	O		(0.00000)	0.0
	PI	(0.00000)	0.0	(0.00000)	0.0
	PO	(0.00000)	0.0	(0.00000)	0.0
	IO		0.00606	8.6
	PIO	0.06409	100.0	0.06409	91.4
Sum of variances	0.09142		0.06409	100%	0.07015	100%
Standard deviation	0.30236		Relative SE: 0.25317		Absolute SE: 0.26487	
Coef_G relative	0.59					
Coef_G absolute	0.57					

Grand mean for levels used: 3.24900

Variance error of the mean for levels used: 0.00793

Standard error of the grand mean: 0.08907

Appendix C49

EduG analyses output for the combination of the six highest State Component Index items of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	2 3 4 5 6 7 8 9 10 11 13 14 16 17 19 20 21 22 23 24 25 26 27 28 29 31 32 33 34 35 36 37 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.03479		
	I		0.00107	1.6
	O		0.00494	7.5
	PI	0.00011	0.2	0.00011	0.2
	PO	0.00774	13.3	0.00774	11.8
	IO		0.00129	2.0
	PIO	0.05036	86.5	0.05036	76.9
Sum of variances	0.03479		0.05820	100%	0.06549	100%
Standard deviation	0.18651		Relative SE: 0.24124		Absolute SE: 0.25591	
Coef_G relative	0.37					
Coef_G absolute	0.35					

Grand mean for levels used: 3.19946

Variance error of the mean for levels used: 0.00841

Standard error of the grand mean: 0.09172

Appendix C50

EduG analyses output for the combination of the seven highest State Component Index items of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person Item	P	83	INF	
	I	39	39	2 3 4 5 6 7 8 9 10 11 13 14 16 17 19 20 21 22 23 24 25 26 27 29 31 32 33 34 35 36 37 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.05243		
	I		0.00184	0.8
	O		0.00277	6.4
	PI	0.00073	1.5	0.00073	0.1
	PO	0.00884	17.7	0.00884	11.2
	IO		0.00107	2.0
	PIO	0.04025	80.8	0.04025	79.5
Sum of variances	0.03243		0.04983	100%	0.05551	100%
Standard deviation	0.18007		Relative SE: 0.22323		Absolute SE: 0.23562	
Coef_G relative	0.48					
Coef_G absolute	0.45					

Grand mean for levels used: 3.20491

Variance error of the mean for levels used: 0.00668

Standard error of the grand mean: 0.08170

Appendix C51

EduG analyses output for the combination of the eight highest State Component Index items of the FFMQ, including observation and estimation designs and G-study table.

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	83	INF	
Item	I	39	39	3 4 5 6 7 8 9 10 11 13 14 16 17 19 20 21 22 23 24 25 26 27 29 31 32 33 34 35 36 37 39
Occasion	O	3	INF	

G Study Table (Measurement design P/IO)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.03469		
	I		0.00180	4.1
	O		0.00237	5.4
	PI	0.00081	2.1	0.00081	1.8
	PO	0.00415	10.6	0.00415	9.4
	IO		0.00081	1.8
	PIO	0.03425	87.4	0.03425	77.5
Sum of variances	0.03469		0.03921	100%	0.04419	100%
Standard deviation	0.18626		Relative SE: 0.19802		Absolute SE: 0.21021	
Coef_G relative	0.47					
Coef_G absolute	0.44					

Grand mean for levels used: 3.13253

Variance error of the mean for levels used: 0.00587

Standard error of the grand mean: 0.07660