

Measuring Trait Mindfulness: How to Improve the Precision of the Mindful Attention Awareness Scale Using a Rasch Model

Oleg N. Medvedev¹ · Richard J. Siegert¹ · Xuan Joanna Feng¹ · D. Rex Billington¹ · Jin Young Jang² · Christian U. Krägeloh¹

Published online: 16 October 2015
© Springer Science+Business Media New York 2015

Abstract The Mindful Attention Awareness Scale (MAAS) is the most widely used mindfulness scale to date, with validation studies indicating acceptable reliability and convergent validity. However, recent evidence indicates that the ability of the MAAS items to precisely discriminate between mindfulness levels is compromised. To improve item functioning and precision of the MAAS, responses of 250 participants to the scale were subjected to Rasch analysis. To improve disordered thresholds, items were re-scored, and each item was tested for Differential Item Functioning. Where misfit to Rasch model expectations was identified, items were removed and the effect on individual item-fit estimates was tested. Uniform rescaling of all items was the best solution to order thresholds of all items and to improve overall goodness of fit to the Rasch model. Satisfactory model fit was achieved after removing the misfitting items 6 and 15 and combining the locally dependent items 7 and 8. Functioning of MAAS items can be improved substantially by several minor modifications to scoring algorithms without the need to modify the current response format. Precision of the instrument can be improved further by using the ordinal-to-interval conversion tables presented here.

Keywords Mindfulness · Measurement · Mindful Attention and Awareness Scale · Rasch Analysis · Psychometrics

Introduction

Recent years have seen a surge of interest in ‘mindfulness’, on the one hand, as a component in clinical interventions for a wide range of psychological and health conditions (Chiesa and Serretti 2010; Goldin and Gross 2010; Zoogman et al. 2015) and, on the other, as a trait or general capacity that is linked to optimal psychological well-being (Keng et al. 2011). Mindfulness has been described as ‘paying attention in a particular way: on purpose, in the present moment, and non-judgmentally’ (Kabat-Zinn 1994, p.4). As a broader disposition, mindfulness is the ability to be aware of external and internal experiences as *phenomena* without *automatically* using existing cognitive schemas that help construct our conceptual world and its objects (Olendzki 2005). Mindfulness is our immediate, instant contact with reality that is not contaminated by habitual and conceptual cognitions and allows a greater clarity of consciousness and more objectively based responses (Brown et al. 2007). However, mindfulness definitions used in psychology may not be adequate to capture more complex traditional understanding of mindfulness (Chiesa 2013).

The clinical *practice* of mindfulness is generally associated with mindfulness-based interventions (MBIs), such as mindfulness-based stress reduction (Kabat-Zinn 1982, 1990) and mindfulness-based cognitive therapy (Segal et al. 2002). Here, participants typically complete a number of exercises, such as meditation, body scan, hatha yoga, mindful walking, and mindful eating. The rapidly growing evidence base for the therapeutic application of mindfulness techniques includes alleviating symptoms and enhancing the coping abilities of people suffering from anxiety, stress, depression, emotional

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

✉ Christian U. Krägeloh
chris.krageloh@aut.ac.nz

¹ Department of Psychology, Faculty of Health & Environmental Sciences, Auckland University of Technology, North Shore Campus, Private Bag 92006, Auckland 1142, New Zealand

² Institute of Mind Humanities, Wonkwang University, Iksan, South Korea

instability, substance abuse, post-traumatic stress disorder, borderline personality disorder, psychophysiological disorders, and suicidal/self-harm behavior (Chiesa and Serretti 2010; Hofmann et al. 2010; Ivanovski and Malhi, 2007). Also, MBIs were shown to increase psychological well-being (Josefsson et al. 2014; Bennet and Dorje 2015). Regular meditation practice and the concomitantly repeated elevations of state mindfulness have recently been shown to lead to increases in trait mindfulness (Kiken et al. 2015). Trait mindfulness is often conceptualized as a *capacity* and involves enhanced emotional regulation (Chambers et al. 2009; Lyvers et al. 2014) and meta-emotion strategies, which are also predictive of psychological well-being (Mitmansgruber et al. 2009). In addition to participants' prior experience with mindfulness practice, one's disposition to mindfulness is an important baseline variable to control in evaluations of MBI (Visted et al. 2015). Trait mindfulness is also increasingly recognized as an important predictor variable in models of psychological well-being (e.g. Brown and Kasser 2005; Pearson et al. 2015).

A number of self-report questionnaires are available to assess levels of mindfulness. Table 1 provides a list of some of the most commonly cited instruments with their demonstrated psychometric properties (Baer et al. 2004, 2006; Brown and Ryan 2003; Cardaciotto et al. 2008; Chadwick et al. 2008; Feldman et al. 2007; Fresco et al. 2007; Haigh et al. 2011; Lau et al. 2006; Walach et al. 2006). Measurement of mindfulness is a relatively new research area, with the first self-report mindfulness measure, the Freiburg Mindfulness Inventory (FMI), published in 2001 (Buchheld et al. 2001; Walach et al. 2006). Of all mindfulness instruments, the Mindful Attention Awareness Scale (MAAS) (Brown and Ryan 2003) is the most widely cited, and its psychometric properties are supported by a larger number of studies than for any other instrument (Park et al. 2013). The second most cited instrument, the Five Facets Mindfulness Questionnaire (FFMQ) (Baer et al. 2006), and the Kentucky Inventory of Mindfulness Skills (KIMS) (Baer et al. 2004) both proposed a multidimensional profile of mindfulness skills (Table 1). Temporal reliability (test-retest) has only been reported for the MAAS (Brown and Ryan 2003), the KIMS (Baer et al. 2004), and the State Mindfulness Scale (SMS) (Tanay and Bernstein 2013).

Perhaps the attractiveness of the MAAS (Brown and Ryan 2003) is related to its simple unidimensional structure and relative brevity. The MAAS is a 15-item self-report questionnaire of trait mindfulness that uses a 6-point Likert-scale response format (1='almost always' to 6='almost never'). A total score is calculated as the mean of responses to all items, with a higher score corresponding to a greater mindfulness level. Example items are as follows: 'I rush through activities without being really attentive to them' and 'I find myself preoccupied with the future or the past'. The MAAS is not fully consistent with mindfulness definitions used in

psychology as it focuses on attention/awareness to the present moment but lacks items distinctly measuring non-judgemental attitude (Kabat-Zinn 1994; Bergomi et al. 2013). However, based on the assumption that mindless states are more common (Brown and Ryan 2003), MAAS items ask individuals about lack of mindfulness, which means that the instrument may serve as an indirect assessment of self-criticism (Bergomi et al. 2013). Converging evidence supports good internal reliability and satisfactory external reliability over a 4-week interval (Table 1). Tests of convergent validity of the MAAS by comparing it with a number of other mindfulness measures showed positive correlations in the range from weak to moderate (Baer et al. 2006; Brown and Ryan 2003; Christopher and Gilbert 2010). Construct validity of the MAAS was supported by its positive correlations with measures of positive affect, well-being and openness, and negative correlations with stress, anxiety, rumination, and neuroticism (Baer et al. 2006; Brown and Ryan 2003; Carlson and Brown 2005; Christopher and Gilbert 2010; Frewen et al. 2008). According to expectations, significantly higher MAAS scores were found for experienced meditators compared to novices (Cordon and Finney 2008).

A number of studies confirmed the proposed unidimensional structure of the MAAS (Brown and Ryan 2003; Carlson and Brown 2005; Christopher et al. 2009; MacKillop and Anderson 2007). However, the factor loading of some MAAS items (items 5, 6, and 13) were occasionally reported to be below 0.30 (Table 2). Specific investigations of the performance of individual MAAS items also revealed some issues. In a study using Item Response Theory, Van Dam et al. (2010) used category response curves (CRC) to demonstrate functioning of MAAS items that included thresholds between pairs of adjacent response options for each item. CRC shows the probability of a person to select a specific response category based on estimation of their latent trait (i.e. mindfulness). The findings showed that only 6 out of 15 items (items 4, 7, 8, 9, 10, and 14) have equally distributed thresholds indicating that only these items are capable to adequately discriminate between different mindfulness levels across available response options. In addition, the relative contribution or ability of each item to assess a latent trait has been examined. Only five items (items 7, 8, 9, 10, and 14) together explain about 66 % of the information related to the latent variable mindfulness (Van Dam et al. 2010). Table 2 shows that items with lower loadings on the first principal component (e.g. item 6) also explain a relatively lower amount of information related to the latent trait. These findings suggest that further research is necessary to investigate the functioning of individual items, which can be conducted using Rasch analysis, a technique that is particularly suited for this purpose (Tennant and Conaghan 2007).

The distinct advantages of Rasch analysis over classical psychometric methods have been discussed in detail

Table 1 Properties of commonly used mindfulness scales including: number of subscales and items, internal consistency (Cronbach's alpha), test-retest reliability, correlations with other mindfulness measures, and number of citations in Google Scholar and Web of Science (28 July 2015)

Scale	Reference (original)	Subscales (items)	Cronbach's alpha	Reliability Test-retest	Validity of Construct	Google Scholar	Web of Science
MAAS	Brown and Ryan (2003)	1 (15)	$\alpha=0.78-0.92$	ICC=0.81; $p<0.001$	FMI, KIMS, MMS, CAMS-R, SMQ: $r=0.14-0.51$; $p<0.05$	3695	1333
FFMQ	Baer et al. (2006)	5 (39)	$\alpha=0.67-0.93$	Not reported	Based on FMI, KIMS, MMS, SMQ, and MAAS items.	1928	764
KIMS	Baer et al. (2004)	4 (39)	$\alpha=0.72-0.97$	$r=0.81-0.86$, but observe $r=0.65$	FMI, SMQ, SAMS-R $r=0.51-0.67$	1138	412
TMS	Lau et al. (2006)	2 (13)	$\alpha=0.85-0.91$	Not reported	KIMS, MAAS, FFMQ, CAMS-R, SMQ, FMI $r=0.10-0.74$; $p<0.05$	490	166
FMI	Buchheld et al. (2001) Walach et al. (2006)	1 (30) 1 (14)	$\alpha=0.80-0.94$ $\alpha=0.86$	Not reported Not reported	KIMS, MAAS, SMQ, and CAM-R: $r=0.31-0.60$	419	186
CAMS-R	Feldman et al. (2007)	1 (12)	$\alpha=0.61-0.81$	Not reported	KIMS, MAAS, FMI and SMQ: $r=0.51-0.67$; $p<0.05$	366	134
PHLMS	Cardaciotto et al. (2008)	2 (20)	$\alpha=0.75-0.91$	Not reported	KIMS, MAAS: $r=0.38-0.61$; $p<0.05$	287	101
SMQ	Chadwick et al. (2008)	1 (16)	$\alpha=0.82-0.89$	Not reported	KIMS, MAAS, FMI: $r=0.38-0.61$; $p<0.05$	207	58
MMS	Haigh et al. (2011)	4 (21)	$\alpha=0.45-0.86$	Not reported	Not reported	24	10
SMS	Tanay and Bernstein (2013)	2 (23)	$\alpha=0.95$	$r=0.64-0.65$	TMS, FFMQ: $r=0.31-0.47$ MAAS: $r=0.00-0.07$	13	2

The scales are listed in descending order of number of citations

ICC inter-class correlation coefficient, MAAS Mindful Attention Awareness Scale, FFMQ Five Facets Mindfulness Questionnaire, KIMS Kentucky Inventory of Mindfulness Skills, TMS the Toronto Mindfulness Scale, FMI the Freiburg Mindfulness Inventory, CAMS-R the Cognitive and Affective Mindfulness Scale-Revised, PHLMS the Philadelphia Mindfulness Scale, SMQ the Southampton Mindfulness Questionnaire, MMS the Mindfulness/Mindlessness Scale, SMS the State Mindfulness Scale

elsewhere (Rasch 1960; Wilson 2005; Wright and Stone 1979). Essentially, Rasch analysis provides a template to convert ordinal-level data to interval level, which improves precision of measurement, provided a measure is unidimensional (Rasch 1961). This process involves the testing of several attributes such as potential item bias, local independence assumptions including unidimensionality, appropriate stochastic ordering of items and response option ordering in polytomous items (Tennant and Conaghan 2007). When these prerequisites are satisfied and the data thus fit a Rasch model, participants can be ordered according to their *ability* on a scale measuring the latent construct (in this case mindfulness). Also, items are ordered by *difficulty*, which refers to the mindfulness level assessed by an item. The same log-odds interval scale is used to order both the items and the participants. This allows graphical representation of a participant-item threshold

distribution, reflecting how well the range of the sample ability fits to the items' difficulty. Applying Rasch analysis to mindfulness measures would be beneficial through the identification of item bias or Differential Item Functioning (DIF). This refers to the situation where participants with the same ability on the latent construct, but from different groups (e.g. males and females), respond differently to an item.

Taken together, MAAS is the most commonly used mindfulness scale, perhaps to a large extent because it is a brief, well-validated instrument with good psychometric properties that can be applied to a wide range of clinical and non-clinical populations. However, recent evidence suggests that only a small sub-set of MAAS items adequately discriminate between mindfulness levels (Van Dam et al. 2010). Rasch analysis is a suitable method to investigate the performance of individual items to discriminate on their overarching construct, but to date,

Table 2 Range of the reported loadings on the first principal component and the total amount of information explained by each MAAS item on the latent variable mindfulness

Item	Range of factor loading ^a	Total information ^b (%)
1 I could be experiencing some emotion and not be conscious of it until some time later	0.43–0.50	2.20
2 I break or spill things because of carelessness, not paying attention, or thinking of something else	0.36–0.57	2.41
3 I find it difficult to stay focused on what's happening in the present	0.51–0.67	5.88
4 I tend to walk quickly to get where I'm going without paying attention to what I experience along the way	0.41–0.62	3.84
5 I tend not to notice feelings of physical tension or discomfort until they really grab my attention	0.27–0.51	4.63
6 I forget a person's name almost as soon as I've been told it for the first time	0.26–0.49	1.37
7 It seems I am "running on automatic" without much awareness of what I'm doing	0.59–0.80	13.18
8 I rush through activities without being really attentive to them	0.68–0.76	17.60
9 I get so focused on the goal I want to achieve that I lose touch with what I am doing right now to get there	0.38–0.72	9.85
10 I do jobs or tasks automatically, without being aware of what I'm doing	0.69–0.74	13.92
11 I find myself listening to someone with one ear, doing something else at the same time	0.45–0.56	3.01
12 I drive places on "automatic pilot" and then wonder why I went there	0.46–0.62	4.63
13 I find myself preoccupied with the future or the past	0.28–0.54	2.42
14 I find myself doing things without paying attention	0.71–0.78	11.77
15 I snack without being aware that I'm eating.	0.36–0.62	3.29

^a The range of factor loading presented here is based on a systematic review by Park et al. (2013)

^b Total information in percent on the latent trait (mindfulness) measured by each MAAS item is based on the Item Response Theory analysis conducted by Van Dam et al. (2010)

Rasch analysis has only been applied to the Spanish version (Inchausti et al. 2014) and not to an English-language version or a non-clinical sample. The aim of the present study is to apply Rasch analysis to explore strategies to improve the psychometric properties of the MAAS.

Method

Participants

The present study analyzed data from 250 participants, based on the recommended optimal sample size estimates for the most purposes of Rasch analysis (Linacre 1994). For the present Rasch analysis, the sample included a randomly selected sub-set of 125 from a sample of New Zealand university students and a randomly selected sub-set of 125 participants from a New Zealand-wide postal survey to control for possible DIF effects. Also, we aimed to make the results applicable to both students and general adult populations in line with the original validation study (Brown and Ryan 2003). The total sample size of the university student sample was 253 (79.1 % females, 19.0 % males, 2.0 % missing gender identification). Ages ranged from 18 to 59 with a mean age of 23.33 (SD=7.73). Ethnic groups include 51.8 % Caucasians, 5.5 % Māori, 7.1 % Pasifika, 16.5 % Asian, and 17.9 % of other unspecified ethnicities. The total sample size of the national general population survey was 436, of whom 155 (35.6 %) indicated that

they were male and 280 (64.2 %) that they were female. Ages ranged from 18 to 91, with a mean of 52.87 and a standard deviation of 17.05. The majority self-classified as Caucasian (81.9 %), 8.5 % as Māori, 2.8 % as Pasifika, 2.8 % as Asian, and the remainder as other ethnicities. After merging the two sub-sets of 125 respondents, the mean age was 38.20 years (SD=20.01). To investigate DIF, three age categories were created: 18–21 ($n=86$), 22–50 ($n=83$), and 51–88 ($n=76$). The number of males was 63 and number of females 185 (missing gender data=5), and those regularly engaging in mindfulness practice was 84, as opposed to 154 not engaging in regular practice.

Procedure

The present study collected responses from university students as well as from the general population. For the student sample, potential participants were approached in class with permission of a paper coordinator and invited to complete the survey and to hand the survey back to the researchers or submit it to a locked collection box at their respective faculty. Students completed the questionnaire in class before the lecture or during the break. To obtain a sample from the general population, a questionnaire was posted to a sample of 4000 individuals randomly selected from the national electoral roll. Respondents returned completed questionnaires using an enclosed self-addressed pre-paid return envelope. The

response rate was 11 %. The authors' institutional ethics committee approved this study.

Measures

The Mindfulness Attention Awareness Scale (MAAS) (Brown and Ryan 2003) is a 15-item self-report questionnaire that has been described as measuring trait mindfulness (Siegling and Petrides 2014). Sample questions include 'I find it difficult to stay focused on what's happening in the present' and 'I do jobs or tasks automatically, without being aware of what I'm doing'. All items use a 6-point Likert-scale response format (1= 'almost always' to 6= 'almost never'). Internal consistency (Cronbach's alpha) in the current data set was 0.87.

Data Analysis

Descriptive statistics and reliability analyses of the MAAS were conducted using IBM SPSS v.22. Data were then formatted and saved as an ASCII file to be imported into the software RUMM2030 for Rasch analysis (Andrich et al. 2009). The likelihood-ratio test was conducted on the initial output analysis and should indicate appropriateness of the unrestricted (Partial-Credit) version of the model. Rasch analysis includes the following sequential steps (Siegert et al. 2010):

1. A test for overall data fit to the Rasch model.
2. Identifying MAAS items with disordered thresholds and rescoreing them.
3. Deletion of items with poorest fit to the Rasch model.
4. Re-testing individual item fits and overall fit to the Rasch model.
5. Analysis of DIF for gender, age, sample population, and practice experience.
6. Unidimensionality test.
7. Examination of local dependency based on the residual correlation matrix.
8. Distribution analysis of the participant-item thresholds.
9. An associating analysis between the best-fitting MAAS items (1–8 steps) and all 15 MAAS items.
10. Comparison of Rasch results with traditional psychometric tests (e.g. item-to-total correlations and factor analysis).

Rasch analysis is an iterative procedure that starts with testing the overall fit to the model using the *threshold map* of the software output to identify any items showing disordered thresholds. A threshold is disordered when participants' higher capacity on a construct (e.g. mindfulness) is not consistently reflected in progressively higher scores on the ordinal-scale response options for that specific item. To correct disordered thresholds, response options may be collapsed.

Usually, one or two items with disordered thresholds are rescored at a time, and, at each step, goodness of fit to the model is re-tested. In the same way, after thresholds are satisfactorily adjusted, poorly fitting items are removed one at a time and the overall fit is re-calculated. The iterative procedure is completed upon meeting the following criteria: Overall model and individual item fit are both satisfactory and unidimensionality of the scale is clearly evident. The Rasch model requires the chi-square statistics for overall model fit, reflecting the item-trait interaction, to be non-significant ($p > 0.05$, Bonferroni adjusted), and overall item and person fit residuals should have a standard deviation close to 1.00 and a mean close to 0.00 in case of an excellent fit. At the individual item level, fit residuals for individual items should be in the range between -2.50 and $+2.50$. Also, no significant differences (Bonferroni adjusted) should be evident between gender, age, and experience groups if DIF is tested.

Dimensionality of MAAS was tested by examining the first principal component of the residuals (Smith 2002). Then, person-locations for two groups of items with the highest positive and the highest negative loadings on the first principal component after removing the Rasch (e.g. mindfulness) factor were compared by an independent-samples *t* test. A scale is considered unidimensional if the percentage of significant *t* test comparisons outside -1.96 to 1.96 confidence intervals is below 5 %. More accurate estimation is conducted using a binominal test of proportions, which assesses the exact amount of acceptable deviations based on sample size. Unidimensionality is confirmed if the lower bound of a binominal confidence interval computed for the number of observed significant *t* tests overlaps the 5 % cutoff point. To test if there is any significant difference to participant estimates for the shorter version, a paired-samples *t* test was conducted to compare participant estimates for the full scale, and the short version after metric compatibility of both scales was established by calibration and rescoreing of response options.

Results

Preliminary Test of the Overall fit to the Rasch Model

The person separation index (PSI) of 0.88 indicated good reliability. However, unsatisfactory overall fit to the model was evident ($\chi^2(45) = 146.71$, $p < 0.001$), and items 2, 5, 12 and 15 displayed clearly disordered thresholds (Table 3, test 1). Therefore, rescoreing of the MAAS items was conducted prior to any further analyses.

Rescoreing of MAAS Items

Iterative rescoreing of the MAAS items showed that optimal ordering of thresholds and goodness of fit could be achieved

Table 3 Summary of fit statistics for the original MAAS version (1), after uniform item rescoring (2), after removing items 6 and 15 (3), and after combining items 7 and 8 into subtest (4)

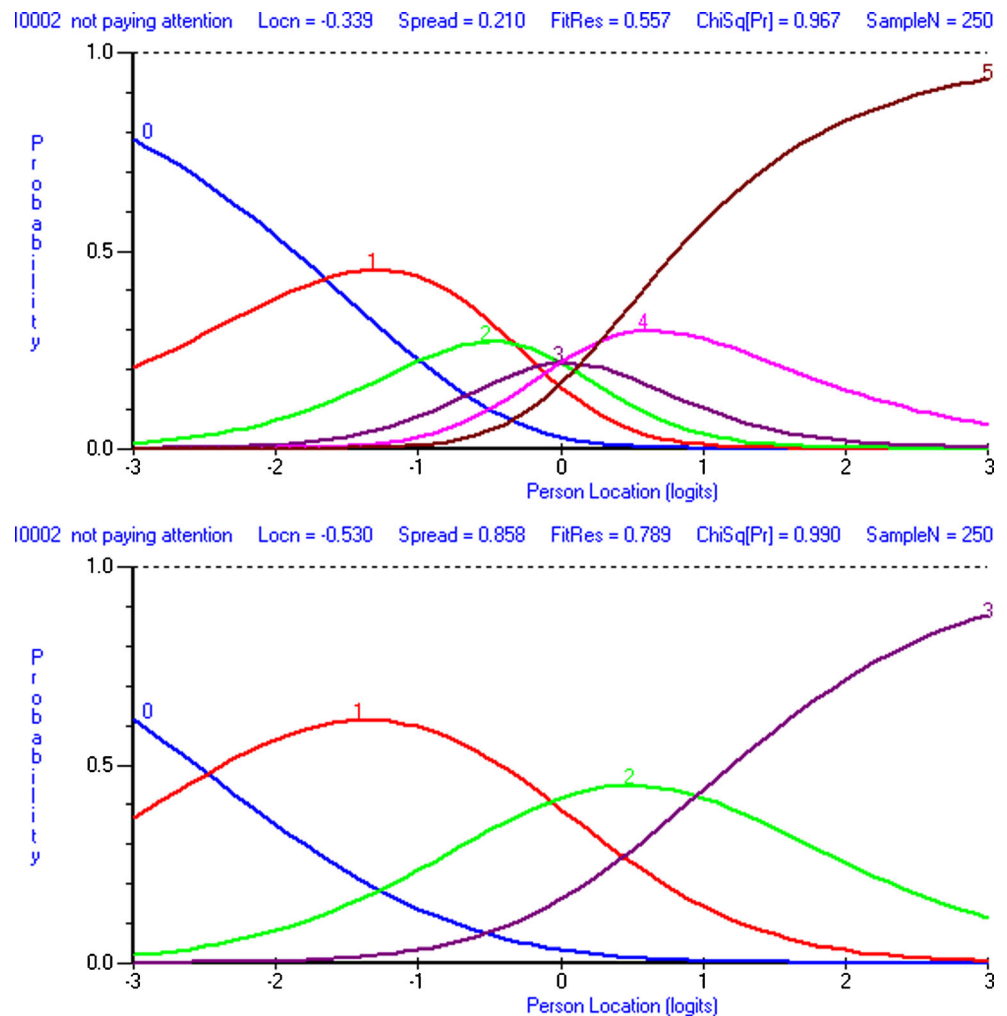
Tests	Item residual		Person residual		Goodness of fit		Person separation	Independent <i>t</i> -test	
	Value	SD	Value	SD	χ^2 (df)	<i>p</i>	Value	%	95 % CI
1	0.44	2.26	-0.21	1.50	147 (45)	<0.001	0.88	6.36	0.03–0.09
2	0.16	1.90	-0.29	1.42	118 (45)	<0.001	0.86	7.20	0.04–0.10
3	0.11	1.38	-0.36	1.42	61 (39)	<0.01	0.87	7.20	0.04–0.10
4	0.11	1.30	-0.36	1.38	47 (36)	0.11	0.87	7.20	0.04–0.10

using uniform rescoring of all the 6-point Likert scale items by collapsing response category 2 (*very frequently*) with 3 (*somewhat frequently*), and category 4 (*somewhat infrequently*) with 5 (*very infrequently*). Figure 1 shows an example of the effect of rescoring item 2 on the category response probability curves. All disordered thresholds became ordered after uniform rescoring, and the overall fit to the model was also improved, although still not at acceptable levels ($\chi^2(45)=118.14$, $p<0.001$, Table 3, test 2). Therefore, it was decided to remove items with poorest fit (highest fit residuals) one at a time, with subsequent tests of fit to the model.

Removing non-fitting Items

Fit residuals for all 15 items were analysed after rescoring all items. Items 6 (*forgetting names*) and 15 (*snacking without awareness*) displayed the largest fit residuals and highest chi-square values, indicating poor fit to the Rasch model, and were removed before the analyses continued. Table 4 includes the chi-square values for all MAAS items from the initial test before rescoring (1) and after rescoring and removing non-fitting items 6 and 15 (test 3). Large chi-square values are associated with poor fit to the Rasch model. Also, Table 4

Fig 1 Item category probability curves illustrating disordered thresholds for MAAS item 2 before rescoring (*top panel*) and orderly thresholds after rescoring (*bottom panel*)



includes the location of each item in log units of probability, or *logits*, which indicates relative *difficulty* of every item on the Rasch scale. Higher positive values signify difficult items (e.g. item 13 *preoccupied with future or past*) meaning that only few individuals obtained higher scores and negative values correspond to less difficult items with more people having higher scores (e.g. Item 2 *not paying attention*). Substantial improvement of fit was noted after removing items 6 and 15, which both had the lowest loading on the first principal component and item-to-total correlations (Table 4). However, chi-square for overall person-item interaction was still significant ($\chi^2(39)=61.01, p<0.01$, Table 3, test 3). At this stage, all the remaining 13 items had satisfactory model fit, with fit residuals below 2.50. Therefore, local dependency between items was investigated because it affects estimations of both discrimination parameters and test information.

Local Dependency

The residual correlations between items were analyzed, and the highest correlation was found between items 7 and 8 (0.28). The residual correlations above 0.20 are generally considered as indicative for local dependency between items. To confirm this observation, the correlation matrix between all items was also examined and showed the highest correlation between items 7 and 8 (0.66). Together, these observations confirmed local dependency, and items 7 and 8 were therefore combined into a single testlet. This solution provided a desirable alternative to achieve a good fit to the Rasch model

($\chi^2(36)=46.79, p>0.05$, Table 3, test 4) without excluding further MAAS items. This solution was replicated with the full sample size ($n=689$), showing identical issues with non-fitting items and local dependency.

Test for Unidimensionality

The set of person estimates from the three items with the highest positive loadings on the first principal component were compared with the set of estimates from the three items with the highest negative loadings. Out of 250 *t* test comparisons between both sets calibrated to the same metric, 17 tests (6.8 %) were significant. A binominal test was conducted to estimate the exact amount of acceptable deviations based on sample size. The calculated value of the binominal 95 % confidence interval (CI) for the observed proportion overlapped 5 % on the lower bound and thus confirmed unidimensionality of the current solution (Table 3).

Differential Item Functioning

Differential Item Functioning (DIF) was analyzed by controlling for gender, age, sample (students versus general population), and practice factors. Significant DIF effect of the sample population was found for items 3 (*difficult to focus on present*) ($F(1249)=11.29, p<0.001$) and 5 (*not notice physical tension*) ($F(1249)=12.25, p<0.001$), Bonferroni adjusted. However, graphical examination showed that for item 3, the differences between samples

Table 4 Corrected item-to-total correlation for MAAS items, item loadings on the first principal component (PC), item difficulty (location), item-fit residuals, chi-square from the initial analysis before rescoring (1), and chi-square values after rescoring and removing Items 6 and 15 (3)

Item	Item-to-total correlation	Item loadings on the 1st PC	Item difficulty (location)	Item-fit residuals	Chi-square (1)	Chi-square (3)
1	Not conscious of emotions	0.42	0.50	-0.33	1.68	1.97
2	Not paying attention	0.47	0.55	-0.34	0.26	1.55
3	Difficult to focus on present	0.59	0.70	-0.14	-1.04	6.44
4	Walk without paying attention	0.57	0.66	0.34	0.24	3.97
5	Not notice physical tension	0.51	0.59	-0.07	0.27	5.44
6	Forgetting names	0.27	0.34	0.69	6.12	-
7	Running on automatic	0.70	0.77	0.11	-1.56	7.92
8	Rush not attentive to activities	0.72	0.80	-0.11	-2.55	6.40
9	Focused on goal achievement	0.60	0.67	0.07	-0.07	4.68
10	Do tasks automatically	0.54	0.62	-0.17	0.01	0.79
11	Listening and doing something	0.42	0.50	0.28	2.58	7.85
12	Drive on 'automatic pilot'	0.58	0.66	-0.35	-1.51	0.99
13	Preoccupied with future or past	0.52	0.59	0.38	1.24	2.76
14	Doing things without attention	0.70	0.76	-0.16	-2.30	10.25
15	Snack without awareness	0.34	0.41	-0.21	2.86	-

were not consistent across observed confidence intervals. Therefore, only item 5 was split for sample DIF resulting in the same item measuring each population independently. Also, the effect of age on DIF was significant for item 5 ($F(2249)=7.90, p<0.001$), Bonferroni adjusted. However, graphical examination revealed that the respective observed means are not systematically different across observed confidence intervals for any of the age groups. No other significant effects on DIF were observed for other personal factors.

Item-person Threshold Distribution

Figure 2 shows the person-item threshold distribution where person ability and item difficulty are plotted on the same logit scale for the final solution (Analysis 4). *Ability* refers to mindfulness that is the latent trait measured by the MAAS. Person-threshold distribution is close to normal, with evidence of a small ceiling effect indicating limited ability of the MAAS to discriminate between higher mindfulness levels. However, the item-threshold distribution satisfactorily covers most people’s abilities of both students and national sample population on the latent trait, and there was no evidence of a floor effect.

Equating Test

A paired-samples *t* test was conducted to compare the means of person estimates from the full 15-item MAAS and the 13-item version. A significant difference was found between the

person estimates of the two versions ($t(250)=1.96, p<0.01$), indicating significant change in the ability of the 13-item version to discriminate between individual mindfulness levels compared to the original 15-item version. This confirms that the implemented modifications led to an improved solution for the MAAS.

Item-to-total Correlations and Loadings on the First Principal Component

Item-to-total correlations and loadings on the first principal component for all the original MAAS items were computed in IBM SPSS for a comparison with Rasch results and are included in Table 4. It shows that the excluded items 6 and 15 have the lowest values for both parameters, confirming that these items are less consistent with the latent construct represented by the remaining items.

Ordinal-to-interval Conversion Table

Table 5 shows how raw scores can be converted from an ordinal scale to an interval scale. The raw scores shown here are after items 6 and 15 have been removed and after response categories 2 and 3 as well as 4 and 5 have been merged. Researchers who have already used the MAAS to collect data can apply the results of this study as follows: items 6 and 15 should be dropped and the response categories *almost always* should be recoded as 0, *very frequently* and *somewhat frequently* as 1, *somewhat infrequently* and *very infrequently* as 2, and *almost never* as 3. Then, the 13 item responses (range

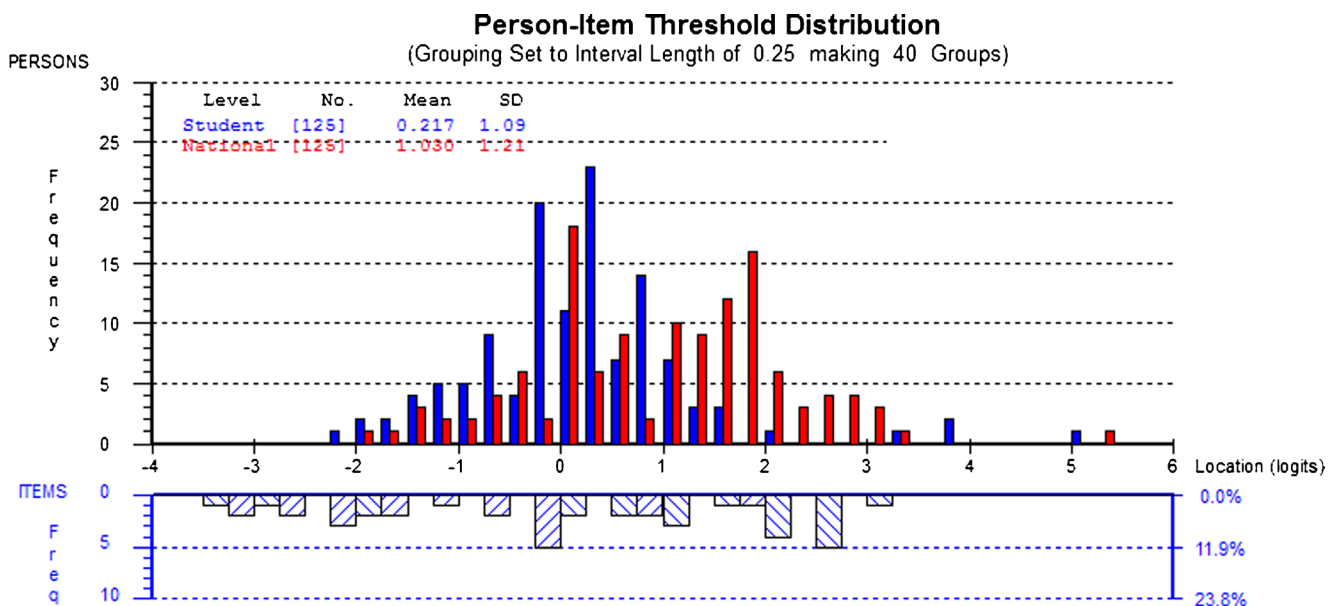


Fig 2 Person-item threshold distribution for the final solution ($N=250$), 13 MAAS items including both students (*blue*) and national sample (*red*) thresholds

Table 5 Converting from a uniformly re-scored 13-item raw score (0 to 39) to an interval scale in logit units and in mean scores ranging from 1 to 6 for the national and student populations

	Raw score ^a		Interval measure				Raw score ^a		Interval measure			
			Students sample		National sample				Students sample		National sample	
			Logit	Mean	Logit	Mean			Logit	Mean	Logit	Mean
0			-5.45	1.00	-5.42	1.00	20	0.13	3.61	0.19	3.63	
1			-4.59	1.40	-4.55	1.40	21	0.30	3.69	0.36	3.71	
2			-3.97	1.69	-3.94	1.69	22	0.47	3.77	0.54	3.79	
3			-3.53	1.90	-3.49	1.90	23	0.64	3.85	0.71	3.87	
4			-3.17	2.07	-3.13	2.07	24	0.81	3.94	0.88	3.95	
5			-2.86	2.21	-2.82	2.22	25	0.99	4.02	1.06	4.03	
6			-2.59	2.34	-2.55	2.34	26	1.16	4.10	1.23	4.12	
7			-2.34	2.46	-2.30	2.46	27	1.34	4.18	1.41	4.20	
8			-2.11	2.57	-2.07	2.57	28	1.53	4.27	1.59	4.28	
9			-1.89	2.67	-1.86	2.67	29	1.71	4.36	1.78	4.37	
10			-1.69	2.76	-1.65	2.76	30	1.91	4.45	1.97	4.46	
11			-1.49	2.85	-1.45	2.86	31	2.11	4.54	2.17	4.55	
12			-1.30	2.94	-1.26	2.95	32	2.32	4.64	2.38	4.65	
13			-1.12	3.03	-1.07	3.04	33	2.55	4.75	2.60	4.76	
14			-0.93	3.12	-0.89	3.12	34	2.80	4.87	2.85	4.87	
15			-0.75	3.20	-0.70	3.21	35	3.08	5.00	3.12	5.00	
16			-0.57	3.29	-0.52	3.29	36	3.40	5.15	3.45	5.15	
17			-0.39	3.37	-0.34	3.38	37	3.81	5.34	3.85	5.34	
18			-0.22	3.45	-0.16	3.46	38	4.39	5.61	4.43	5.61	
19			-0.05	3.53	0.01	3.54	39	5.22	6.00	5.25	6.00	

Note: ^a The following uniform rescaling of response options for all 13 items is required before converting into an interval scale: 1 to 0; 2 to 1; 3 to 1; 4 to 2; 5 to 2; 6 to 3. The 13-item raw score is calculated as the sum of re-scored values from all MAAS items except for Items 6 and 15

of scores 0 to 39) should be summed. Next, Table 5 should be used to convert these scores to means on a 1-to-6 scale similar to the original MAAS scoring system. By using the conversion table provided here, users are able to increase the precision of the MAAS. Considering the above-reported DIF by sample, separate conversion tables are presented for use with student and with general population samples. Note that the ordinal-to-interval scale conversion proposed here does not require altering the response format of the scale but only involves a different scoring algorithm. These conversion tables were replicated with the full sample size ($n=689$), showing almost identical results.

Discussion

The MAAS (Brown and Ryan 2003) is the most widely used scale to measure trait mindfulness, despite the fact that its psychometric properties have not been fully clarified. Recently, Van Dam et al. (2010) reported results that challenged the ability of the MAAS to adequately discriminate between mindfulness levels, as only 6 out of the total 15 items had ordered thresholds, and 5 of these items represented

approximately two thirds of the total information of the scale. The present Rasch analysis added to the limited number of studies that have investigated the performance of individual MAAS items in detail. While 2 items (item 6 *forgetting names* and item 15 *snacking without awareness*) had to be removed, the functioning of the remaining 13 items could be improved substantially by uniform rescaling. The psychometric properties of the MAAS following these adjustments are thus robust, and the precision of the scale can be further improved by using the ordinal-to-interval conversion algorithm in Table 5. This increased precision is not only desirable for studies that investigate the effects of clinical interventions on trait mindfulness but it also means that parametric statistics may now be used to analyse MAAS data.

Compared to Van Dam et al. (2010), who reported disordered thresholds for items 1, 2, 3, 5, 6, 11, 12, 13, and 15, we only found clearly disordered thresholds for items 2, 5, 12, and 15. However, the thresholds for items 6, 11, and 13 were only marginally acceptable. For that reason, we decided to use uniform rescaling of all items. This improved the thresholds of all items and goodness of fit to the Rasch model. This solution also ensures that our proposed rescaling algorithm is easy to apply and suitable for users unfamiliar with Rasch analysis.

The removed items (*forgetting names* and *snacking without awareness*) were found to have item-to-total correlations and factor loadings that were clearly lower than those of other items and were also found to explain only a relatively small amount of information on the latent trait (Table 3). Possible reasons are that forgetting names may be more related to assessment of verbal memory and might only indirectly refer to mindfulness. This item may also be biased as the effort to remember a difficult (e.g. foreign) name might interfere with attention available to the present moment (Nickerson 1978). The item ‘I snack without being aware that I’m eating’ may not work well because it appears relatively unlikely that people are unaware of a whole episode of snacking, even though they may be absentminded during some periods during their snacking. Also, in a multicultural sample such as the present one, snacking habits may be very varied, and some people may prefer ‘eating’ to ‘snacking’, as the latter may be associated with unhealthy food. Future research may investigate to what extent re-wording of these items may improve the psychometric properties of the 15-item MAAS. However, until this work has been conducted, we recommend using the 13-item version with our proposed scoring algorithm.

Van Dam et al. (2010) argued that the ability of the MAAS to discriminate between mindfulness levels is impaired due to application of items measuring *mindlessness* because an individual without special training is not capable of accurately registering mindless states. However, Brown and Ryan (2003) insisted that mindlessness is more common among the general population, and hence, they should have the ability to adequately report it. The results of the present study support the construct validity of the MAAS in the 13-item format and suggest that items measuring lack of mindfulness might be adequate to reliably assess the construct if uniform item rescoring is applied. Moreover, the ordinal MAAS scale can be converted to an interval scale without changing the response format (Table 5), which accounts for DIF between students and general adult populations and provides interval level scores for each sample.

The following limitations need to be acknowledged. Even though the sample reflects New Zealand’s diversity of ethnic groups, no efforts were made to purposively sample underrepresented groups. The response rate of 11 % for the national electoral roll sample was very low, which may have amplified self-selection biases. However, such response rates are not uncommon for research of this nature in New Zealand (Hill et al. 2014; Krägeloh et al. 2013), and the above DIF analyses explored in detail any effects by demographic variables. Additionally, achieving a suitable fit to the Rasch model required rescoring of all items, which makes scoring the scale somewhat more complex. Nonetheless, converging from ordinal to an interval level scale can be conducted for both students and general population in logit units and in mean values using the same metric as the original scale for easy

comparison. The benefits of this certainly outweigh this inconvenience, and the authors may be contacted for assistance with data conversion. The readers are advised to refer to the present version of the MAAS as ‘the 13-item version’ to differentiate it from the original 15-item version.

Trait mindfulness has emerged as an important contributor to health and well-being, and its accurate measurement represents an on-going challenge. The current study used Rasch analysis to address previously reported limitations of the widely used MAAS trait mindfulness instrument. We demonstrated that item functioning and precision of the MAAS can be enhanced to satisfy the expectations of a unidimensional Rasch model using uniform rescoring of item response categories. Two items significantly affected individual estimates and appeared less consistent with the latent trait. The precision of the MAAS can be optimized by discarding these two items and by using the ordinal-to-interval conversion tables published here.

Acknowledgements The authors thank Jason Landon and Penny Neave for helping with data collection. This study is a part of doctoral work of the first author funded by the Vice-Chancellor’s Scholarship of the Auckland University of Technology. This article was supported by a National Research Foundation of Korea (NRF) grant (NRF-2010-361-A00008) funded by the Korean Government (MEST).

Compliance with Ethical Standards We declare that we have no conflicts of interest in connection with this paper. The study was conducted in compliance with the guidelines of the Auckland University of Technology Ethics Committee.

References

- Andrich, D., Sheridan, B., & Luo, G. (2009). *RUMM 2030*. Perth: RUMM Laboratory.
- 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.
- 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.
- Bennett, K., & Dorjee, D. (2015). The impact of a mindfulness-based stress reduction course (MBSR) on well-being and academic attainment of sixth-form students. *Mindfulness*, doi: 10.1007/s12671-015-0430-7
- 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.
- Brown, K. W., & Kasser, T. (2005). Are psychological and ecological well-being compatible? The role of values, mindfulness, and lifestyle. *Social Indicators Research, 74*(2), 349–368.
- 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), 822–848.
- Brown, K. W., Ryan, R. M., & Creswell, J. D. (2007). Mindfulness: theoretical foundations and evidence for its salutary effects. *Psychological Inquiry, 18*(4), 211–237.
- 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.
- 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.
- 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.
- 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.
- Chambers, R., Gullone, E., & Allen, N. B. (2009). Mindfulness emotion regulation: an integrative review. *Clinical Psychology Review*, 29(6), 560–572.
- Chiesa, A. (2013). The difficulty of defining mindfulness: current thought and critical issues. *Mindfulness*, 4(3), 255–268.
- Chiesa, A., & Serretti, A. (2010). A systematic review of neurobiological and clinical features of mindfulness meditations. *Psychological Medicine*, 40(8), 1239–1252.
- 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., & 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.
- 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.
- Feldman, G., Hayes, A., Kumar, S., Greeson, J., & Laurenceau, J.-P. (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.
- Fresco, D. M., Moore, M. T., van Dulmen, M. H., Segal, Z. V., Ma, S. H., Teasdale, J. D., & Williams, J. M. (2007). Initial psychometric properties of the experiences questionnaire: validation of a self-report measure of decentering. *Behavior Therapy*, 38(3), 234–246.
- Frewen, P. A., Evans, E. M., Maraj, N., Dozois, D. J. A., & Partridge, K. (2008). Letting go: mindfulness and negative automatic thinking. *Cognitive Therapy & Research*, 32(6), 758–774.
- 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.
- 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.
- Hill, E. M., Billington, R., & Krägeloh, C. (2014). Noise and diminished health: testing moderators and meditators of the relationship. *Noise & Health*, 16(68), 47–56.
- Hofmann, S. G., Sawyer, A. T., Witt, A. 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.
- Inchausti, F., Prieto, G., & Delgado, A. R. (2014). Rasch analysis of the Spanish version of the Mindful Attention Awareness Scale (MAAS) in a clinical sample. *Revista de Psiquiatría y Salud Mental*, 7(1), 32–41.
- Ivanovski, B., & Malhi, G. S. (2007). The psychological and neurophysiological concomitants of mindfulness forms of meditation. *Acta Neuropsychiatrica*, 19(2), 76–91.
- Josefsson, T., Lindwall, M., & Broberg, A. G. (2014). The effects of a short-term mindfulness based intervention on self-reported mindfulness, decentering, executive attention, psychological health, and coping style: examining unique mindfulness effects and mediators. *Mindfulness*, 5(1), 18–35.
- Kabat-Zinn, J. (1982). An outpatient program in behavioural medicine for chronic pain patients based on the practice of mindfulness meditation: theoretical considerations and preliminary results. *General Hospital Psychiatry*, 4(1), 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, USA: Delacorte.
- Kabat-Zinn, J. (1994). *Wherever you go, there you are: mindfulness meditation in everyday life*. New York, USA: Hyperion.
- Keng, S.-L., Smoski, M. J., & Robins, C. J. (2011). Effects of mindfulness on psychological health: a review of empirical studies. *Clinical Psychology Review*, 31(6), 1041–1056.
- Kiken, L. G., Garland, E. L., Bluth, K., Palsson, O. S., & Gaylord, S. A. (2015). From a state to a trait: trajectories of state mindfulness in meditation during intervention predict changes in trait mindfulness. *Personality and Individual Differences*, 81, 41–46.
- Krägeloh, C. U., Kersten, P., Billington, D. R., Hsu, P. H.-C., Shepherd, D., Landon, J., & Feng, X. J. (2013). Validation of the WHOQOL-BREF quality of life questionnaire for general use in New Zealand: confirmatory factor analysis and Rasch analysis. *Quality of Life Research*, 22(6), 1451–1457.
- Lau, M. A., Bishop, S. R., Segal, Z. V., Buis, T., Anderson, N. D., Carlson, L. E., Shapiro, S., Carmody, J., Abbey, S., & Devins, G. (2006). The Toronto Mindfulness Scale: development and validation. *Journal of Clinical Psychology*, 62(12), 1445–1467.
- Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7, 328.
- Lyvers, M., Makin, C., Toms, E., Thorberg, F. A., & Samios, C. (2014). Trait mindfulness in relation to emotional self-regulation and executive function. *Mindfulness*, 5(6), 619–625.
- 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.
- Mitmansgruber, H., Beck, T. N., Höfer, S., & Schüßler, G. (2009). When you don't like what you feel: experiential avoidance, mindfulness and meta-emotion in emotion regulation. *Personality and Individual Differences*, 46(4), 448–453.
- Nickerson, R. S. (Ed.). (1978). *Attention and performance VIII*. New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.
- Oleznick, A. (Ed.). (2005). *The roots of mindfulness*. New York: Guilford.
- 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(10), 2639–2659.
- Pearson, M. R., Brown, D. B., Bravo, A. J., & Witkiewitz, K. (2015). Staying in the moment and finding purpose: the associations of trait mindfulness, decentering, and purpose in life with depressive symptoms, anxiety symptoms, and alcohol-related problems. *Mindfulness*, 6(3), 645–653.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test*. Copenhagen: Danish Institute for Educational Research.
- Rasch, G. (1961). *On general laws and the meaning of measurement in psychology*. Berkeley, California: University of California Press. Symposium conducted at the meeting of the Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.
- Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2002). *Mindfulness based cognitive therapy for depression*. New York, USA: Guilford Press.

- Siebert, R. J., Tennant, A., & Turner-Stokes, L. (2010). Rasch analysis of the Beck Depression Inventory-II in a neurological rehabilitation sample. *Disability and Rehabilitation*, 32(1), 8–17.
- Siegling, A. B., & Petrides, K. V. (2014). Measures of trait mindfulness: convergent validity, shared dimensionality, and linkages to the five-factor model. *Frontiers in Psychology*, 5, 1–8.
- Smith, E. V. (2002). Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 205–231.
- Tanay, G., & Bernstein, A. (2013). State Mindfulness Scale (SMS): development and initial validation. *Psychological Assessment*, 25(4), 1286–1299.
- Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism*, 57(8), 1358–1362.
- Van Dam, N. T., Earleywine, M., & Borders, A. (2010). Measuring mindfulness? An item response theory analysis of the Mindful Attention Awareness Scale. *Personality and Individual Differences*, 49(7), 805–810.
- Visted, E., Völlestad, J., Nielsen, M. B., & Nielsen, G. H. (2015). The impact of group-based mindfulness training on self-reported mindfulness: a systematic review and meta-analysis. *Mindfulness*, 6(3), 501–522.
- Walach, H., Buchheld, N., Buttenmüller, V., Kleinknecht, N., & Schmidt, S. (2006). Measuring mindfulness—the Freiburg Mindfulness Inventory (FMI). *Personality and Individual Differences*, 40(8), 1543–1555.
- Wilson, M. (2005). *Constructing measures*. Mahwah, NJ: LEA.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.
- Zoogman, S., Goldberg, S. B., Hoyt, W. T., & Miller, L. (2015). Mindfulness interventions with youth: a meta-analysis. *Mindfulness*, 6(2), 290–302.