

Article

Measuring Mindfulness: A Rasch Analysis of the Freiburg Mindfulness Inventory

Sebastian Sauer ^{1,2,*}, Harald Walach ^{3,4}, Martin Offenbächer ¹, Siobhan Lynch ⁵ and Niko Kohls ^{1,2,4}

¹ Generation Research Program, Human Science Center, Ludwig-Maximilians-Universität, Prof.-Max-Lange-Platz 11, 83646 Bad Tölz, Germany; E-Mails: offenbaecher@med.uni-muenchen.de (M.O.); kohls@grp.hwz.uni-muenchen.de (N.K.)

² Peter Schilffarth Institute, Prof.-Max-Lange-Platz 16, 83646 Bad Tölz, Germany

³ Institute of Transcultural Health Studies, European University Viadrina, Frankfurt (Oder), Postfach 1786, 15207 Frankfurt (Oder), Germany; E-Mail: walach@europa-uni.de

⁴ Brain, Mind and Healing Program, Samuelli Institute, 1737 King Street, Suite 600, Alexandria, VA 22314, USA

⁵ Institute of Psychiatry, King's College London, Stamford St, London SE1 8WA, UK; E-Mail: smlynch79@gmail.com

* Author to whom correspondence should be addressed; E-Mail: sauer@grp.hwz.uni-muenchen.de; Tel.: 00498041799290; Fax: 004980417992911.

Received: 31 October 2011 / Accepted: 1 December 2011 / Published: 8 December 2011

Abstract: The objective of the study was to assess the psychometric properties of the Freiburg Mindfulness Inventory (FMI-14) using a Rasch model approach in a cross-sectional design. The scale was administered to $N = 130$ British patients with different psychosomatic conditions. The scale failed to show clear one-factoriality and item 13 did not fit the Rasch model. A two-factorial solution without item 13, however, appeared to fit well. The scale seemed to work equally well in different subgroups such as patients with or without mindfulness practice. However, some limitations of the validity of both the one-factorial and the two-factorial version of the scale were observed. Sizeable floor and ceiling effects limit the diagnostical use of the instrument. In summary, the study demonstrates that the two-factorial version of the FMI-13 shows acceptable approximation to Rasch requirements, but is in need of further improvement. The one-factorial solution did not fit well, and cannot be recommended for further use.

Keywords: mindfulness; measurement; assessment; Rasch; item response theory

Abbreviations

CTT: classical test theory; DIF: differential item functioning; DMS: Developmental Mindfulness Survey; IRT: Item Response Theory; FMI: Freiburg Mindfulness Inventory; MAAS: Mindfulness Attention Awareness Scale; MBI: Mindfulness-based interventions. MNSQ: mean square; RSM: rating scale model.

1. Introduction/Background

Broadly conceptualized, mindfulness can be described as a non-elaborative, non-judgmental, present-centered awareness in which each thought, feeling, or sensation that arises in the attentional field is acknowledged and accepted as it is [1]. Through the pioneering work of Jon Kabat-Zinn and others, mindfulness has been introduced into secular settings, such as the modern Western medical system, and has recently become a focus of interest in the health sciences. Consequently, there has been a surge of interest in mindfulness-based interventions (MBI) from both clinical and health psychology. Several meta-analyses have documented that MBI are an efficient tool for reducing clinical symptoms and generic distress in clinical and non-clinical populations [2,3]. Additionally, recent neuroscientific advancements have made it possible to identify neural correlates of mindfulness [4-6].

MBI seem to be not only an effective intervention for clinical and non-clinical populations but also a way for scrutinizing the neuropsychological mechanisms underlying attentional and emotional processes. However, a number of crucial questions concerning the field of mindfulness research have not been adequately addressed. One of the most pressing issues is associated with the development of an internally and externally valid assessment instrument. In short, Mindfulness has predominantly been assessed by means of questionnaire instruments [7-9], although alternatives have been proposed [10]. Approximately a dozen mindfulness questionnaires have been published [11] such as the Mindfulness Attention and Awareness Scale (MAAS) [7] or the Freiburg Mindfulness Inventory (FMI) [12]. The existing questionnaires have only partly, but by no means comprehensively, been validated.

The FMI (14 item version; FMI-14) is a widely used instrument that assesses trait mindfulness that has been validated in a number of studies, but so far not been subjected to more rigorous psychometric analyses such as Item Response Theory (IRT) analysis [12-15]. Both a one-factorial and a two-factorial version have been proposed for the FMI [9], as the latent component structure of mindfulness is still a subject of ongoing debate. However, most questionnaire instruments attempting to measure mindfulness explicitly or implicitly draw upon an attention factor and an acceptance factor. Even instruments consisting of several scales such as Baer's Five Facet Mindfulness Questionnaire may be interpreted on the basis of a higher order two factorial taxonomy with different layers (attention facet: describe, act with awareness, observe; acceptance facet: nonreactivity, nonjudging) [8,16]. One exception to this latent structure model is the MAAS scale which explicitly models mindfulness as consisting solely of an attention factor [7]. The FMI-14 seems to be a promising

psychometric instrument to assess mindfulness in clinical and non-clinical populations. Recent experimental research has suggested that the effects of mindfulness may be blurred if a one-factorial interpretation of mindfulness is used [17]. As both a one and a two-factorial solution have been proposed for the FMI, this instrument seems to be an ideal candidate for scrutinizing the factorial structure of mindfulness [9]. As further empirical support for a certain factorial structure may have implications for our understanding of mindfulness, there is need of a further and more sophisticated examination of measurement instruments exploring mindfulness.

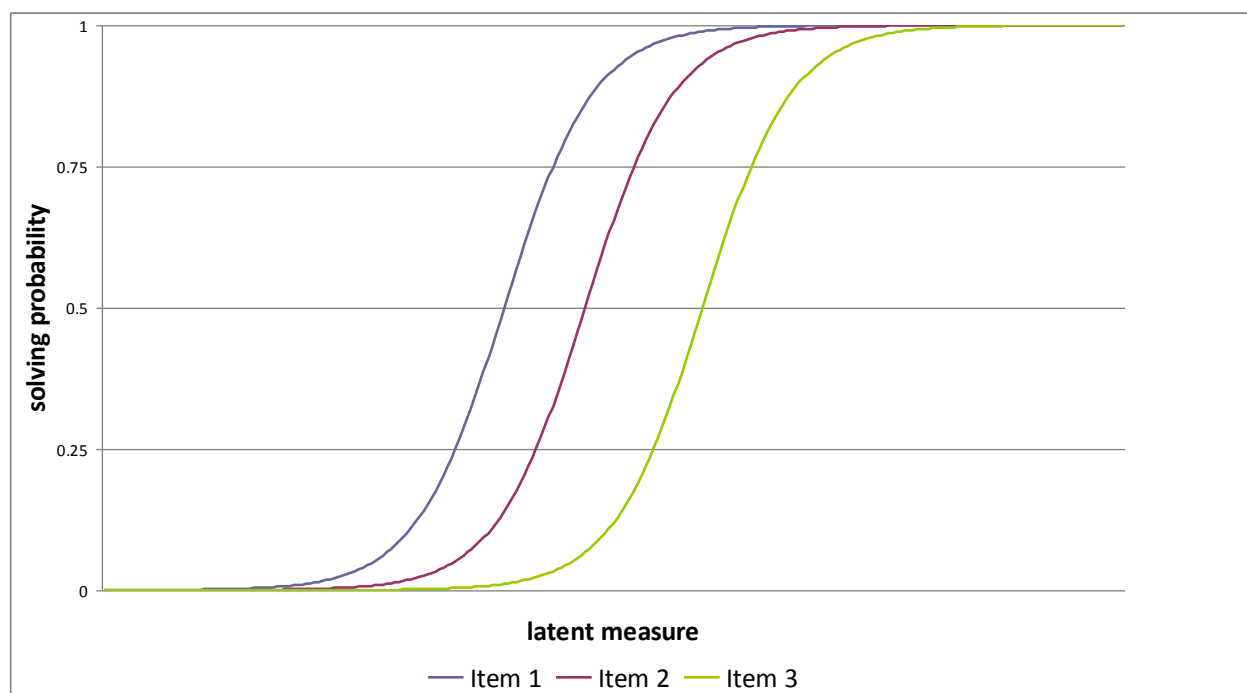
1.1. The Rasch Model

A practical alternative to Classical Test Theory (CTT) construction methods is the Rasch model that belongs to the family of IRT models. IRT models generally formulate the probability of endorsing a given item as a function of the latent trait such as mindfulness [18,19]. A number of IRT models have been proposed, such as the Rasch model or the Birnbaum model (the former being a special case of the latter) [18,19]. Different IRT models rely on different measurement assumptions and hence different models are used for different purposes. The Rasch model allows the explicit testing of basic statistical properties that cannot be investigated by means of CTT [20,21]. For example, one can explicitly test by means of Rasch analysis, if it is legitimate to collapse single item scores into a sum score. Similarly, only if Rasch-type models corroborate the claim that measurement on an interval scaled level is achieved, can additive or multiplicative computations be justified [21-23]. From a mathematical point of view, the Rasch model tests whether the relationship between item responses and a latent trait such as mindfulness adheres to an s-formed curve, or ogive—similar to a cumulative normal curve (Figure 1). CTT models implicitly assume a linear relationship, although the underlying model is only rarely tested in an explicit manner. Moreover, linear models pose logical restrictions [24], and ogive-type forms are frequently more plausible, particularly with regard to psychological variables. This is because ogive-type models assume that an increase in extreme (high or low) levels of a respective latent variable has only a small impact on the general probability associated with endorsement of the respective item, whereas for intermediate levels of the same latent trait characteristic, an increase in the trait is associated with a substantial increase in the endorsement probability. The Rasch model may therefore be seen as a useful complementary approach to standard CTT.

It should be noted that other item response theory models, such as the Birnbaum model, do not fit the measurement axioms that allow for explicitly testing interval level measurement [22,25]. However, the conjoint measurement theory, as presented by Luce and Tukey [26], and formalized by Krantz and colleagues [27], lays out the conditions for psychological variables to be measured on an interval level. It has been shown that, while the Rasch model adheres to its axioms in many circumstances, other IRT models are not in line with the conjoint measurement theory [23]. Rasch models have been employed in a substantial number of psychometric studies [28-31]. However, in research related to mindfulness measurement, most questionnaires have been developed using CTT approaches. Two exceptions where IRT models have been used are the Developmental Mindfulness Survey (DMS) [32], which was developed using a Rasch model approach, and the psychometric investigation of the MAAS which has been validated using the Graded Response Model [33]. The authors of the DMS reported excellent fit

values, although this scale has not been externally validated nor employed in other published studies. The MAAS is probably one of the most widely used instruments in the field, but IRT results suggest that the scale may measure the construct of mindlessness rather than mindfulness.

Figure 1. Rasch item function curves for three items with different difficulties.



Notes: The ordinate depicts the probability of endorsing (or solving) an item. The abscissa depicts the latent measure (such as mindfulness).

1.2. Aims of the Present Research

More psychometric research is needed to clarify contemporary research questions related both to the conceptualization and operationalization of mindfulness. Given the high interest in mindfulness research, existing instruments should be tested by means of methods complementary to CTT, such as the Rasch modeling, as this will facilitate a deeper understanding of mindfulness and its measurement. This article subjects the Freiburg Mindfulness Inventory to a Rasch analysis in order to tackle the questions related to its factorial structure from a different perspective than CTT. Due to the increasing application of MBI in clinical contexts, we will focus on the applicability of the FMI in clinical populations.

2. Methods

2.1. Participants and Procedure

The sample consisted of a clinical population of $N = 130$ British patients (female = 112, male = 17, missing = 1) who were treated in a single private medical clinic for integrative medicine (Southampton, UK). The aim of the study was to investigate the impact of spiritual and exceptional experiences on health, thereby particularly focusing on the moderating role of mindfulness [34].

Seventy-one participants were diagnosed with chronic fatigue syndrome, 29 with irritable bowel syndrome, 9 with migraine, and 29 with other symptoms (not specified). Six hundred questionnaires were disseminated by post along with an introduction letter, a consent form and a pre-paid envelope for returning the survey. In total, 130 of 600 questionnaires were returned together with a signed consent form, leading to an overall return rate of 22%. Mean age was 46.9 years (SD = 14.7). Sixty-nine of the participants were married, 35 were single, 25 divorced, and 1 person was widowed. Seventy-nine had children, and 49 did not (2 responses missing). Twenty participants had no educational degree, 5 had high school degree, 16 undergraduate degree, 69 graduate, 5 held a PhD, and 15 did not report on educational degrees. Slightly more than half of the participants (69) reported following a spiritual or meditative practice such as Yoga regularly, as opposed to 61 who did not. All participants had to give informed consent before they were allowed to participate in the study. The study was approved by the relevant internal review board of the respective university.

2.2. Measure

The items of the FMI-14 are shown in Table 1. Response options range from 1 (*almost never*) to 4 (*almost always*). The empirical alpha of the FMI-14 in this sample was $r = 0.88$. The FMI captures trait mindfulness and has been shown to have good psychometric properties including a high internal consistency (alpha of 0.86 in an initial validation study), and it has been shown to correlate positively with health indicators [13–15]. Furthermore, the scale was able to differentiate between mindfulness practitioners and non-practitioners [12]. A recent confirmatory factor analysis showed good fit indices [9]. However, the confirmatory factor analysis has also suggested that there may be ambiguity with regard to the dimensional structure of the instrument as the authors found comparable evidence for both an one-factorial and a two-factorial solution [9].

2.3. Measurement Properties of the Rasch Model

The Rasch model, part of the IRT model family, can be understood as a stochastic approximation to conjoint measurement theory that allows for explicitly testing the existence of interval metric measurement in a given data set [35]. The Rasch model assumes that the probability of a subject endorsing an item can be exclusively seen as a function of the distance between the item difficulty and the person ability on a linear scale [20]. Rasch models possess the feature of “specific objectivity” indicating that individual measures can be compared independently of the subset of relevant items (item invariance). Similarly, item difficulties are considered to be independent of the sample in which the questionnaire was administered if the Rasch model can be corroborated (person invariance).

More formally, within the dichotomous Rasch model, the probability p of person i solving item j , is in its most simple form a logistic function of the difference of ability θ (or attitude *etc.*) and item difficulty β on a linear scale. The response to the item is often denoted by X ; endorsing a dichotomous item (“yes”) can be denoted by $X = 1$, whereas non-endorsing (“no”) can be denoted by $X = 0$:

$$p(X_{ij} = 1 | \theta, \beta) = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \quad (1)$$

Hence, the probability of endorsing a given item is a function of the ability of that person and the respective difficulty of the item. The probability of endorsing the item monotonically increases with the respective ability of the person increasing. Both item as well as person parameter values in the context of Rasch analysis are measured in “logit” units. For example, an item parameter is calculated as the natural logarithm of the odds-ratio (*i.e.*, the probability of solving an item divided by the probability of not solving an item). A logit of 0 equals a probability of 50% of endorsing a given item, whereas a probability of 75% (90%) equals approx. a logit value of 1.1 (2.2). Probabilities smaller than 50% are identical with regard to the size of logit, but they are indicated by negative signs (*e.g.*, logit of 25% is approx. -1.1 logits). The curve-linear relation between probability of endorsing an item and the associated amount of the person ability can be seen in Figure 1.

Table 1. Item difficulties sorted by item hierarchy (easy and difficult items).

Nr.	Items	Item Difficulty (in logits)	Factor
“EASY” ITEMS		−0.32 *	
5	I pay attention to what’s behind my actions.	−1.00	Presence
1	I am open to the experience of the present moment.	−0.79	Presence
7	I feel connected to my experience in the here-and-now.	−0.45	Presence
2	I sense my body, whether eating, cooking, cleaning or talking.	−0.10	Presence
3	When I notice an absence of mind, I gently return to the experience of the here and now.	0.13	Presence
10	I watch my feelings without getting lost in them.	0.31	Acceptance
“DIFFICULT” ITEMS		0.24 *	
4	I am able to appreciate myself.	−0.20	Acceptance
8	I accept unpleasant experiences.	−0.15	Acceptance
14	I am able to smile when I notice how I sometimes make life difficult.	0.03	Acceptance
6	I see my mistakes and difficulties without judging them.	0.15	Acceptance
11	In difficult situations, I can pause without immediately reacting.	0.27	Acceptance
12	I experience moments of inner peace and ease, even when things get hectic and stressful.	0.27	Acceptance
9	I am friendly to myself when things go wrong.	0.30	Acceptance
13	I am impatient with myself and with others.	1.23	Acceptance

Notes: * overall difficulty for “easy” and “difficult” items, respectively. Both categories of easy and difficult items are sorted by their difficulty in ascending order. In contrast, “Nr” of the item refers to the position of the item in the questionnaire.

2.4. Analysis

From the various Rasch models available Andrich’s Rating Scale Model (RSM) was employed [36]. This model seemed ideal for our data set, as it allows a) for the analysis of a scale consisting of items with more than two answer options (*i.e.*, polytomous items), and b) with the same number of answer options across all items. Additionally, in contrast to Master’s Partial Credit Model [37], the RSM is appropriate for smaller samples sizes such as in the present study [38]. We used the software program Winsteps 3.92 for the statistical analyses [21,39]. Alpha was set at .01 for each test, if not explicitly

stated otherwise. We followed the recommendations for Rasch analysis and model test recommended by Bond and Fox [21], and Tennant and Connaghan [40].

2.5. Rasch Validity

Substantial differences between the expected and empirically observed matrix values generally indicate a poor model fit. The INFIT and OUTFIT coefficients are commonly employed statistical parameters for describing the fit / misfit between empirically observed data and a theoretical model. The OUTFIT coefficient is defined as the mean squared deviations (MNSQ) from the model's expected values that can easily be affected by outliers (hence the name "OUT"FIT). The INFIT coefficient weights the MNSQ by the variance of observed values and is more robust towards outliers as it is more strongly determined by values within the interquartile distance (hence "IN"FIT). The INFIT coefficient is generally preferred for assessing the model fit. A value of 1 indicates an ideal fit whereas values greater than 1.5 are considered as indicators of poor fit [21]. The respective threshold of 1.5 suggests that the data would exhibit more than 50% variability than actually predicted by the model.

It is reasonable to assume for polytomous item formats that, for example, to endorse the answer options *fairly often* or *almost always* a higher level of ability is needed compared to the answers options *rarely* or *occasionally*. However, this hypothesis can and should be empirically tested. Rasch analysis provides the category thresholds coefficients τ that indicate the ability level at which the probability for two adjacent categories is 50%. If the relation between categories decoding higher ability levels and their respective τ form are well-ordered (*i.e.*, transitive), then the polytomous format can be seen as established.

A common test for the presumed one-factoriality of the Rasch model is a principal component analysis based on the residuals after the extraction of the Rasch factor. If one-factoriality can be observed, then no further factors should be present. Since concise and universal criteria have yet to be established for determining whether a potential additional dimension should to be considered, results were interpreted according to the recommendations of Linacre [2010]. Therefore >60% of variance explained by the Rasch factor and <5% explained by the largest potential additional dimension was considered to indicate a good fit. Additionally, an eigenvalue <3 indicates that the potential second dimension has only marginal explanatory power.

From a psychometric point of view a good measurement instrument should be unaffected by specific sample characteristics such as age or gender. Lack of this property is called "Differential Item Functioning" (DIF). In non-technical terms, items should be equally difficult for all subgroups such as women or men, *i.e.*, no DIF should be present. We included DIF analyses for diagnosis in our analysis as it is highly desirable that mindfulness instruments can be used irrespectively of such sample characteristics. Furthermore, as mindfulness instruments are frequently administered to individuals both with and without proficiency in mindfulness and systematic meditative or spiritual practice or lack thereof, it is important that the instrument is "fair" to both groups with regard to the difficulty level. Lack of DIF between two groups is indicative of both groups using the same skill for the instrument. We tested DIF for the variables "mindfulness practice" and "spiritual practice" (both binary criterions). Finally we included central demographic variables (age and education). Due to the

gender imbalance (117 female vs. 12 male), we did not compute DIF for gender. To control for alpha error inflation, we applied Bonferroni correction (*i.e.*, adjusting the alpha in each DIF variable by the amount of tests).

The hierarchy of item difficulties was addressed as a further aspect for evaluating the validity of the scale. We hypothesized that items associated with the presence facet were easier to endorse than items related to the acceptance facet. This is based on the assertion of Kohls *et al.* [9] that the presence facet builds up earlier in the course of meditative proficiency than the acceptance facet.

2.6. Rasch Reliability

Similar to the concept of reliability in classical test theory, Rasch measurement also provides reliability estimates. However, in the context of Rasch reliability, estimates are computed separately for item and persons parameters, whereas in CTT only item reliability is considered. Rasch reliability is calculated as the ratio of explained to total variance, with the resulting value ranging from 0 to 1 and can be interpreted in a similar vein to internal consistency (Cronbach's alpha) in CTT where values of $r > 0.7$ are deemed acceptable, and values $r > 0.8$ are considered to be good [21].

3. Results

3.1. Uni-dimensional Model

The one-factorial model with 14 items of the FMI was tested as it is probably the most widely used form of the FMI [12–14]. As previous research provided evidence for both the one-factorial as well as a two-factorial solution, we initially examined the dimensionality. The Rasch factor explained 47% of variance (expected variance explained: 46%) with an eigenvalue of 12.5. The second greatest factor explained 11% of variance with an eigenvalue of 3.0. Hence, clear evidence of a second factor seems to be present. Items 1, 2, and 3 exhibited positive loadings higher than .40 on the second factor, and for another four items negative loadings higher than 0.40 (10, 4, 9, 11) on the same second factor were observed. This second factor can tentatively be labeled “present moment attention”. The Rasch factor seemed to describe a latent factor embracing an emotionally neutral, non-judgmental stance. This suggests that a one-factorial solution does not fit with the data. As all items showed good fit values except for item 13 (INFIT MNSQ = 2.23; see Table 2) we decided to exclude this item from further analyses. One obvious reason for the misfit of item 13 is that it is the only negatively coded item, which is also placed at the end of the scale.

3.2. Two-Factorial Model with 13 Items

A Rasch model for each of the two proposed subfacets of the FMI, presence (items 1,2,3,5,7,10) and acceptance (items 4,6,8,9,11,12,14) was then tested separately, as originally suggested by Kohls *et al.* [9]. For the presence facet, one-factoriality and local independence could be confirmed, as the Rasch factor accounted for 72% (expected: 72%) with an eigenvalue of 15.2. The second potential factor accounted for only 9% of the variance (eigenvalue = 1.9) and was therefore not considered to be of practical relevance. Item fit indices were good with a maximum INFIT MNSQ of 1.3, and the overall INFIT MNSQ was very high (0.99). The item answer category thresholds were found to be

ordered and evenly distributed. DIF was not present for age when we compared two groups on the basis of median split halves. DIF was also not found for diagnosis, education, mindfulness practice, and spiritual practice (such as prayer or meditation in general). The reliability was acceptable (0.78; separability: 1.89), although the scale was probably too easy for the sample (mean person parameter 0.53). Strong floor and ceiling effects were present as demonstrated by the higher dispersion of person parameters (−3.45 to 3.64) compared to item parameters (−0.75 to 0.81), hence limiting the discriminative capabilities of the instrument. However, taken together the presence facet exhibited good adherence to the Rasch model although reliability could be improved.

Table 2. Item characteristics of revised model (FMI-13) sorted by item difficulty.

Item	Score	β	SE	INFIT	OUTFIT		
				MNSQ	INFIT z	MNSQ	OUTFIT z
10	178	0.43	0.13	0.93	−0.60	0.92	−0.60
9	178	0.42	0.13	0.89	−0.90	0.87	−1.10
12	176	0.40	0.13	0.98	−0.10	0.98	−0.10
11	178	0.39	0.13	0.96	−0.30	0.94	−0.50
6	187	0.26	0.13	1.04	0.40	1.05	0.50
3	190	0.23	0.13	1.13	1.10	1.20	1.50
14	191	0.13	0.15	1.08	0.70	1.07	0.60
2	204	−0.01	0.12	1.28	2.10	1.29	1.80
8	207	−0.06	0.13	1.21	1.70	1.20	1.60
4	214	−0.11	0.12	0.73	−2.50	0.67	−2.70
7	234	−0.38	0.12	0.76	−2.10	0.72	−2.10
1	265	−0.75	0.13	0.92	−0.60	0.93	−0.40
5	258	−0.96	0.14	1.00	0.00	0.98	−0.10

Notes: Score: Score per item across all persons. β : parameter for item difficulty. INFIT MNSQ: item fit, adjusted for outlier. INFIT z: z-transformed INFIT MNSQ. OUTFIT MNSQ: Unweighted item fit. OUTFIT z: z-transformed OUTFIT MNSQ.

Similar results were obtained for the acceptance facet. One-factoriality could also be confirmed. The Rasch factor accounted for 61% of the variance (eigenvalue 10.9). The second strongest factor accounted for only 9% of the variance (eigenvalue = 1.6). The item fit indices were found to be good with a maximum INFIT mnsq of 1.2 (item 8); the mean INFIT mnsq was very high (0.99). The difficulty of the response options were ascending in the expected order (order of difficulty of response options: *rarely* – *occasionally* – *fairly often* – *almost always*). No DIF was found to be present for age (median split), diagnoses, education, mindfulness practice, or spiritual practice. Rasch reliability was good with $r = 0.81$ (separability 2.04) as was the targeting (*i.e.*, mean person parameter of −0.01 was very close to mean item difficulty, anchored at 0.00). However, strong floor and ceiling effects remained as the person parameter distribution was much more dispersed (−3.99 to 3.98) than the item parameter distribution (−0.46 to 0.30). The two factors correlated substantially (0.68; $p < 0.001$); a value similar to that found by Kohls *et al.* [9].

3.3. Hierarchy of Item Difficulties

To determine whether the hypothesized hierarchy of item difficulties could be confirmed in the data, a one-factorial Rasch model without item 13 was computed. Again, this model failed to adhere to the assumption of one-dimensionality. Despite this shortcoming, it may allow for a hierarchy of item difficulty to be derived. Ranking items by their relative difficulty may prove heuristically useful to deepening understanding of the construct of mindfulness. The mean difficulty for the presence items was -0.32 whereas the mean difficulty for the acceptance items was 0.24 (see Table 2). This substantial difference of approximately a 0.5 logits suggests that, as expected, presence items are easier to endorse than acceptance items.

4. Discussion

The purpose of this research was to scrutinize psychometric properties of the 14 item version of the FMI using a Rasch model approach. Although further work is needed to gain a better understanding as to whether mindfulness can be measured validly, our findings indicate the FMI-14 with a two-factor interpretation should be favored over the unidimensional interpretation. Dimensionality analysis revealed that the one-factorial FMI-14 model did not fit the Rasch model well as a second factor existed. However, the second factor could be identified as the presence facet of mindfulness. Model fit indicators suggested that the scale contains one particularly misfitting item (item 13) that threatens the validity of the scale, and it is suggested that this item should be deleted, reformulated and / or replaced. Item and person fit indicators speak clearly in favor of the two-factorial solution. Additionally, category order and local independence assumptions were also confirmed in the two-factorial solution.

In more lay terms, our results corroborate the two-factorial solution of the FMI brought forward by Kohls *et al.* [9] but not the one-factorial solution. As Rasch analysis is a stringent way to establish the validity of a psychometric measurement, this study provides further evidence that mindfulness can be measured validly using the two factor version of the FMI. The two-factorial FMI may thus be seen as an internally valid measure of mindfulness. This is in line with the recent research conducted by Sauer *et al.* [17] who found that the effects of mindfulness, as measured by the FMI, can be detected in implicit emotional measures when the two factorial solution of the FMI is employed. The difference in difficulty of the presence and acceptance items may support the hypothesis that, in the course of acquiring proficiency in mindfulness, the ability to be present is established before the ability to accept things is built up. This novel finding is an additional support for the two-factorial conceptualization of mindfulness that allows the development of a testable prospective model of how proficiency in mindfulness may be obtained. To our knowledge, no research has, until now, explicitly investigated differential difficulties related to potential subfacets of mindfulness, although the majority of researchers would probably agree that mindfulness embraces both attentional and emotional components [1,41]. It should be stressed, however, that the proposed hierarchy in item difficulty is only tentative and in need of additional empirical corroboration.

The fact that item 13 was the only item with poor fit warrants further investigations as it offers potential insights into the structure of mindfulness. It is well known that negatively coded items, especially if there are only a few and placed at the end of the questionnaire, may be confusing for the

respondents [42,43]. Additionally, it is also possible that the item did not confuse the respondents, but that “mindlessness” may actually not be seen as an inverse conceptualization of mindfulness, but rather a (partly) different construct in its own right. In fact, the authors of the MAAS, a scale consisting of only negatively coded items (*i.e.*, “mindlessness”), reported that they found quite different results in an initial positively coded version of their scale during the construction process [7]. The only moderate correlation between the MAAS and other scales, such as the FMI or the Kentucky Inventory of Mindfulness Skills, may be interpreted in light of this hypothesis [8]. More evidence in favor of this hypothesis has been published recently. Haigh *et al.* [44] reported, in a study related to the construction of the Langer Mindfulness/Mindlessness Scale that they were able to detect both a mindfulness and a mindlessness factor. Future research should consider that “mindfulness questionnaires” and “mindlessness questionnaires” may not assess the same latent construct. As this issue is of central importance for the field of mindfulness, further research is needed to deepen our understanding of the construct and to devise appropriate measurement instruments.

Of theoretical interest for the measurement of mindfulness is that no DIF was present for individuals with or without mindfulness training. This finding corroborates the assumption raised by the authors that the FMI-14 is semantically robust. There was no evidence that individuals with diverging sociodemographic characteristics such as age or diagnoses had a different understanding of the items. Thus, we opine that there is no need for researchers to calculate mindfulness scores separately for different subgroups (such as individuals with or without mindfulness experience).

However, a number of limitations of the present study need to be borne in mind. First and foremost, only some aspects relevant for confirming Rasch properties were scrutinized in this article. For example, further groups of persons (*e.g.*, individuals with different diagnoses or subjects with another language background) should be included in future DIF analysis. It should also be recalled that the sample size was comparatively small, although still within acceptable boundaries according to Linacre [38]. Nevertheless, replication in a larger sample would be desirable. With regard to the generalizability of the results, it must be noted that the sample consisted of older and well educated participants. Thus, different results may be observed in data of younger and / or less educated participants. Furthermore, as in many survey-based studies, low return may potentially bias the results. Finally, it needs to be emphasized that even excellent internal validity is no assurance that a given scale will also exert good external validity.

Nevertheless, reflecting the findings of this study, we would like to offer some recommendations for improving the scale. Firstly, more items should be developed with high or low difficulties so that the scale will be able to gauge individuals outside an intermediate level of mindfulness. This is important because limited differentiation capabilities may attenuate existing effects. In the present state, the FMI-14 (both the one and two-factorial solution) may not be capable of detecting strong effects potentially attributable to mindfulness based interventions. Hence, mindfulness researchers would profit from more sensitive measurement instruments capable of detecting differences between individuals' high and low in mindfulness to a larger degree. Furthermore, we suggest item 13 should be excluded from the scale.

5. Conclusions

In summary, we conclude that the two-factorial solution of the FMI-13 (not the FMI-14) shows acceptable approximation to Rasch requirements. Improvements, as outlined above, are strongly warranted and may yield a reliable and internally valid measurement device for the measurement of mindfulness.

Acknowledgments

Sebastian Sauer and Niko Kohls are supported by the Peter-Schilffarth-Institute. Harald Walach and Niko Kohls are supported by the Samuelli Institute. This work was made possible by the US Army Medical Research and Materiel Command under Award No. W81XWH-06-1-0279 to the Samuelli Institute. The views, opinions and/or findings contained in this report are those of the authors and should not be construed as an official DoD position, policy or decision unless so designated by other documentation.

References and Notes

1. Bishop, S.R.; Lau, M.; Shapiro, S.; Carlson, L.E.; Anderson, N.D.; Carmody, J.; Segal, Z.V.; Abbey, S.; Speca, M. Mindfulness: A proposed operational definition. *Clin. Psychol. Sci. Pract.* **2004**, *11*, 230-241.
2. Hofmann, S.G.; Sawyer, A.T.; Witt, A.A.; Oh, D. The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *J. Consult. Clin. Psychol.* **2010**, *78*, 169-183.
3. Chiesa, A.; Calati, R.; Serretti, A. Does mindfulness training improve cognitive abilities? A systematic review of neuropsychological findings. *Clin. Psychol. Rev.* **2011**, *31*, 449-464.
4. Cahn, B.R.; Polich, J. Meditation states and traits: Eeg, erp, and neuroimaging studies. *Psychol. Bull.* **2006**, *132*, 180-211.
5. Hölzel, B.; Carmody, J.; Vangel, M.; Congleton, C.; Yerramsetti, S.M.; Gard, T.; Lazar, S.W. Mindfulness practice leads to increases in regional brain gray matter density. *Psychiatr. Res. Neuroimaging* **2011**, *191*, 36-43.
6. Hölzel, B.; Ott, U.; Gard, T.; Hempel, H.; Weygandt, M.; Morgen, K.; Vaitl, D. Investigation of mindfulness meditation practitioners with voxel-based morphometry. *Soc. Cogn. Affect. Neurosci.* **2008**, *3*, 55-61.
7. Brown, K.W.; Ryan, R.M. The benefits of being present: Mindfulness and its role in psychological well-being. *J. Pers. Soc. Psychol.* **2003**, *84*, 822-848.
8. Baer, R.A.; Smith, G.T.; Hopkins, J.; Krietemeyer, J.; Toney, L. Using self-report assessment methods to explore facets of mindfulness. *Assessment* **2006**, *13*, 27-45.
9. Kohls, N.; Sauer, S.; Walach, H. Facets of mindfulness—results of an online study investigating the freiburg mindfulness inventory. *Pers. Individ. Differ.* **2009**, *46*, 224-230.
10. Collins, S.E.; Chawla, N.; Hsu, S.H.; Grow, J.; Otto, J.M.; Marlatt, G.A. Language-based measures of mindfulness: Initial validity and clinical utility. *Psychol. Addict. Behav.* **2009**, *23*, 743-749.
11. Didonna, F. *Clinical Handbook of Mindfulness*; Springer: Berlin, Germany, 2008.

12. Walach, H.; Buchheld, N.; Büttenmüller, V.; Kleinknecht, N.; Schmidt, S. Measuring mindfulness—The freiburg mindfulness inventory (fmi). *Pers. Individ. Differ.* **2006**, *40*, 1543-1555.
13. Sauer, S.; Walach, H.; Kohls, N. Gray's behavioural inhibition system as a mediator of mindfulness towards well-being. *Pers. Individ. Differ.* **2011**, *50*, 506-551.
14. Eisendrath, S.J.; Delucchi, K.; Bitner, R.; Fenimore, P.; Smit, M.; McLane, M. Mindfulness-based cognitive therapy for treatment-resistant depression: A pilot study. *Psychother. Psychosom.* **2008**, *77*, 319-320.
15. Leigh, J.; Bowen, S.; Marlatt, G.A. Spirituality, mindfulness and substance abuse. *Addict. Behav.* **2005**, *30*, 1335-1341.
16. Baer, R.A.; Smith, G.T.; Lykins, E.; Button, D.; Krietemeyer, J.; Sauer, S.; Walsh, E.; Duggan, D.; Williams, J.M.G. Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. *Assessment* **2008**, *15*, 329 -342.
17. Sauer, S.; Walach, H.; Schmidt, S.; Hinterberger, T.; Horan, M.; Kohls, N. Implicit and explicit emotional behavior and mindfulness. *Conscious. Cogn.* **2011**, *20*, 1558-1569.
18. Ayala, R.J.D. *The Theory and Practice of Item Response Theory*; Guilford Press: New York, NY, USA, 2009.
19. Embretson, S.E.; Reise, S.P. *Item Response Theory for Psychologists*; LEA: Mahwah, NJ, USA, 2000.
20. Molenaar, I.W.; Fischer, G.H. *Rasch Models: Foundations, Recent Developments, and Applications*; Springer: Berlin, Germany, 1995.
21. Bond, T.G.; Fox, C.M. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*; LEA: Mahwah, NJ, USA, 2007.
22. Karabatsos, G. The rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *J. Appl. Meas.* **2001**, *2*, 389-423.
23. Karabatsos, G.; Ullrich, J.R. Enumerating and testing conjoint measurement models. *Math. Soc. Sci.* **2002**, *43*, 485-504.
24. Rost, J. *Lehrbuch Testtheorie - Testkonstruktion*; Huber: Bern, Switzerland, 2004.
25. Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores*; Lord, F.M., Novick, M.R., Eds.; Addison-Wesley: Reading, PA, USA, 1968; pp. 395-479.
26. Luce, R.D.; Tukey, J.W. Simultaneous conjoint measurement: A new type of fundamental measurement. *J. Math. Psychol.* **1964**, *1*, 1-27.
27. Krantz, D.H.; Luce, R.D.; Suppes, P.; Tversky, A. *Foundations of Measurement Volume i: Additive and Polynomial Representations*; Dover Publications: New York, NY, USA, 2006.
28. Haley, S.M.; McHorney, C.A.; Ware, J.E. Evaluation of the mos sf-36 physical functioning scale (pf-10): I. Unidimensionality and reproducibility of the rasch item scale. *J. Clin. Epidemiol.* **1994**, *47*, 671-684.
29. Henson, S.; Blandon, J.; Cranfield, J. Difficulty of healthy eating: A rasch model approach. *Soc. Sci. Med.* **2010**, *70*, 1574-1580.
30. Hong, S.; Min, S.-Y. Mixed rasch modeling of the self-rating depression scale. *Educ. Psychol. Meas.* **2007**, *67*, 280-299.

31. Norra, C.; Böcker, M.; Wirtz, M.; Gauggel, S.; Forkmann, T. P01-385 - a short screening for depression (desc) using rasch analysis: Development and validation of the new instrument. *Eur. Psychiatr.* **2010**, *25*, 598-598.
32. Salloway, S.G.; Fischer, W.P. Mindfulness practice: A rasch variable construct innovation. *J. Appl. Meas.* **2007**, *8*, 359-372.
33. Van Dam, N.T.; Earleywine, M.; Borders, A. Measuring mindfulness? An item response theory analysis of the mindful attention awareness scale. *Pers. Individ. Differ.* **2010**, *49*, 805-810.
34. Kohls, N.; Walach, H.; Lewith, G. The impact of positive and negative spiritual experiences on distress and the moderating role of mindfulness. *Arch. Psych. Rel.* **2009**, *31*, 357-374.
35. Perline, R.; Wright, B.D.; Wainer, H. The rasch model as additive conjoint measurement. *Appl. Psychol. Meas.* **1979**, *3*, 237-255.
36. Andrich, D. A rating formulation for ordered response categories. *Psychometrika* **1978**, *43*, 561-573.
37. Masters, G. A rasch model for partial credit scoring. *Psychometrika* **1982**, *47*, 149-174.
38. Linacre, J.M. Sample size and item calibration [or person measure] stability. *Rasch Meas. Transact.* **1994**, *7*, 328-328.
39. Linacre, J.M. *A User's Guide to Winsteps, Ministep Rasch-Model Computer Programs*; Winsteps: Chicago, IL, USA, 2010.
40. Tennant, A.; Conaghan, P.G. 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 Rheum.* **2007**, *57*, 1358-1362.
41. Baer, R.A. Measuring mindfulness. *Contemp. Buddh.* **2011**, *12*, 241-261.
42. Schmitt, N.; Stuits, D.M. Factors defined by negatively keyed items: The result of careless respondents? *Appl. Psychol. Meas.* **1985**, *9*, 367-373.
43. Motl, R.W.; DiStefano, C. Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Struct. Equ. Model.* **2002**, *9*, 562-578.
44. Haigh, E.A.P.; Moore, M.T.; Kashdan, T.B.; Fresco, D.M. Examination of the factor structure and concurrent validity of the langer mindfulness/mindlessness scale. *Assessment* **2011**, *18*, 11-26.