

Rasch Analysis of the Kentucky Inventory of Mindfulness Skills

Oleg N. Medvedev^a, Richard J. Siegert^a, Paula Kersten^b and Christian U. Krägeloh^a

^a Auckland University of Technology, New Zealand, Faculty of Health and Environmental Sciences

^b University of Brighton, UK, School of Health Sciences

Address of corresponding author:

Oleg Medvedev
Department of Psychology
Faculty of Health & Environmental Sciences
Auckland University of Technology – North Shore Campus
Private Bag 92006
Auckland 1142
New Zealand
Tel.: +64 9 921 9999 ext 7661
Fax: +64 9 921 9780
Email: oleg.medvedev@aut.ac.nz

Email address of other authors:

Professor Richard Siegert: richard.siegert@aut.ac.nz

Professor Paula Kersten: p.kersten@brighton.ac.uk

Professor Christian Krägeloh: chris.krageloh@aut.ac.nz

Acknowledgements

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.

Rasch Analysis of the Kentucky Inventory of Mindfulness Skills

Abstract

The current study reports Rasch analysis conducted to enhance the psychometric properties of the Kentucky Inventory of Mindfulness Skills (KIMS), the widely-used 39-item multidimensional measure of four mindfulness traits including: Observing, Describing, Act With Awareness, and Accept Non-Judgementally. While this instrument has generally robust psychometric properties, the ability of its subscales to precisely discriminate between trait levels and the functioning of individual items, have not been rigorously investigated. We subjected the responses of 287 participants to Rasch analysis and found that the Rasch model fits KIMS subscales after minor modifications that involved rescoreing one item, removing misfitting items and combining locally dependent items into subtests. Precision of the KIMS can be improved substantially by using the proposed 34-item version of the instrument together with the ordinal-to-interval conversion tables presented here, without any need to modify the original response format. These findings can be useful for clinicians applying mindfulness-based interventions and researchers investigating neurophysiological and psychological correlates of trait mindfulness.

Keywords: Mindfulness, Measurement, Kentucky Inventory of Mindfulness Skills, Rasch analysis, Psychometrics

Introduction

Mindfulness practice is a non-invasive method for the management of stress, emotional problems and for the improvement of psychological well-being (Baer, 2003; Brown & Ryan, 2003). In the context of psychotherapy, mindfulness can be defined as “the non-judgmental observation of the on-going stream of internal and external stimuli as they arise” (Baer, 2003, p. 125). Considering the most cited definitions, mindfulness refers to paying attention to and being aware of internal and external experiences of the present moment associated with a non-judgmental attitude and may include friendliness, acceptance, kindness, curiosity and allowing (Segal, Williams, & Teasdale, 2013). There has been a surge in the application of mindfulness training for a wide range of psychological and health conditions. Indeed, there is a rapidly growing evidence base for the therapeutic application of mindfulness techniques for alleviating symptoms and enhancing the coping abilities of people suffering from anxiety, stress, depression, emotional instability, substance abuse, post-traumatic stress disorder, borderline personality disorder, psychophysiological disorders, and suicidal/self-harm behavior (Chiesa & Serretti, 2010; Hofmann, Sawyer, Witt, & Oh, 2010; Zoogman, Goldberg, Hoyt, & Miller, 2014). Also, mindfulness-based interventions (MBIs) were reported to enhance psychological well-being (Bennet & Dorje, 2015; Josefsson, Lindwall, & Broberg, 2014) and regulation of emotions (Chambers, Gullone, & Allen, 2009; Lyvers, Makin, Tomas, Thorberg, & Samios, 2014). Trait mindfulness was implicated as an important predictor in different models of psychological well-being (Brown & Kasser, 2005; Pearson, Brown, Bravo, & Witkiewitz, In Press) and hence its baseline levels should be controlled when evaluating outcomes of MBIs (Visted, Vøllestad, Nielsen, & Nielsen, 2015). Consequently, there is a need for well-validated and precise mindfulness measures that accurately reflect psychological changes in people participating in MBIs.

1 The Kentucky Inventory of Mindful Skills (KIMS) was developed as a multi-dimensional self-report
2 measure of mindfulness-related skills introduced in the context of Dialectical Behavior Therapy
3 (DBT) (Baer, Smith, & Allen, 2004). DBT treatment unifies mindfulness of non-judgmental
4 observation derived from Zen Buddhism with Western contemplative traditions (Hayes, Follette, &
5 Linehan, 2004). At the beginning of therapy, the goal is to develop individual skills of observing
6 thoughts, emotions, and external stimuli by describing them. DBT emphasizes acting with awareness
7 as a skill by cultivating it through a series of exercises that develop a routine of focusing attention on
8 activities. Non-judgemental acceptance is also a primary skill that is recognized as part of the
9 therapeutic process. To foster this skill, patients are encouraged to accept their reality and tolerate
10 any unwanted feelings or thoughts without judgement (Linehan, 1993a, 1993b). It was proposed that
11 mindfulness-related facets measured by the KIMS have utility in therapeutic contexts because they
12 allows professionals to separate areas of skill development and, accordingly, assist individuals in
13 strengthening specific skills (Baer et al., 2004).
14

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31 It should be noted that unlike mindfulness-based stress reduction (Kabat-Zinn, 1982, 1990) and
32 mindfulness-based cognitive therapy (Segal, Williams, & Teasdale, 2002), which emphasize the
33 central role of mindfulness in the therapeutic process, DBT includes mindfulness as a sub-
34 component among other treatment tools to increase sensory and perceptual awareness in normal,
35 non-meditative circumstances (Hayes, Strosahl, & Wilson, 1999; Linehan, 1993a). For example, the
36 Describe subscale would not be consistent with the most cited mindfulness definitions used in
37 psychology to design mindfulness measures (Bishop et al., 2006; Kabat-Zinn, 1994; Segal et al.,
38 2013), which limits comparisons of the KIMS with other mindfulness measures. The exception is the
39 Five Facets Mindfulness Questionnaire (FFMQ) (Baer, Smith, Hopkins, Krietemeyer, & Toney,
40 2006), which has some structural similarities with the KIMS and was constructed from the combined
41 items of five mindfulness scales including the KIMS (Baer et al., 2004), the Mindful Attention and
42 Awareness Scale (MAAS) (Brown & Ryan, 2003), the Freiburg Mindfulness Inventory (Buchheld,
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Grossman, & Walach, 2001; Walach, Buchheld, Buttenmuller, Kleinknecht, & Schmidt, 2006), the Southampton Mindfulness Questionnaire (Chadwick et al., 2008), and the Cognitive and Affective Mindfulness Scale (Feldman, Hayes, Kumar, Greeson, & Laurenceau, 2007). However, the FFMQ has five subscales compared to the four of KIMS and includes items that are not specifically designed for the measurement of mindfulness skills utilized in a DBT context.

The KIMS includes 39 items divided into four subscales: Accept Non-Judgementally, Observe, Act With Awareness, and Describe (Baer et al., 2004). Accept Non-Judgementally is a subscale of the KIMS that measures the judging behavior present in individuals, such as self-criticism. Observe is a subscale that measures the degree of attention an individual pays to both external events and internal emotions, sensations, and cognitions. Act With Awareness assesses the individual's ability to be fully attentive to the present moment. Describe is a subscale that measures an individual's predisposition to describe or label their external and internal experiences. The items are presented in a 5-point Likert scale format, with responses ranging from 'Never or very rarely true' = 1 to 'Very often or always true' = 5. Examples of typical items reflecting the four skills include: "I tell myself I shouldn't be feeling the way I'm feeling" (Non-Judgementally), "I notice when my mood changes" (Observe), "I tend to do several things at once" (Act With Awareness), and "I find words to describe my feelings" (Describe). Evidence indicates acceptable internal consistency for the total KIMS scale and all subscales, and good test-retest reliability for all subscales ($r = .81$ to $.86$), with the exception of Observe ($r = .65$). In an exploratory factor analysis, 43% of variance in the data was explained by four factors, which was interpreted as support for the four-factor model of the KIMS (Baer et al., 2004). Confirmatory factor analysis supported the four-factor model but failed to confirm an overarching second-order mindfulness factor (Baer et al., 2004; Baum et al., 2010).

Convergent and divergent validity of the total KIMS and its subscales Non-Judgementally and Act With Awareness were supported by positive correlations with self-compassion, openness, and

1
2 emotional intelligence and negative correlations with mindlessness, neuroticism, and dissociation
3 (Baer et al., 2006). However, these relationships appear less consistent for the Observe and the
4
5 Describe subscales (Baer et al., 2004; Christopher & Gilbert, 2010; Frewen, Evans, Maraj, Dozois, &
6
7 Partridge, 2008). Describing internal and external experiences may have a therapeutic value although
8
9 it does not feature in most psychological definitions of mindfulness (Baer, 2003; Kabat-Zinn, 1994;
10
11 Segal et al., 2013). The main limitations of the KIMS include relatively low correlations (ranging
12
13 from 0.09 to 0.34) between subscales (Baer et al., 2004) and concerns related to the content validity
14
15 (e.g. the Describe subscale) (Park, Reilly-Spong, & Gross, 2013). To date, no reports are available
16
17 about investigations into the psychometric properties of the KIMS using modern item-response
18
19 theory and in particular the Rasch model. The ultimate goal of a Rasch analysis is conversion from
20
21 ordinal-level data to interval level, which increases measurement precision and permits parametric
22
23 statistical analyses without violation of their assumptions (Brogden, 1977; Rasch, 1961). Generally,
24
25 only few reported studies so far have subjected mindfulness measures to Rasch analysis (Goh,
26
27 Marais, & Ireland, 2015; Inchausti, Prieto, & Delgado, 2014; Medvedev et al., 2015; Sauer, Ziegler,
28
29 Danay, Ives, & Kohls, 2013). While overall these studies provided useful diagnostic information of
30
31 the FMI and English and Spanish versions of the MAAS at both individual item and scale levels,
32
33 only one study has published the conversion tables that allow a reader to transform raw scores from
34
35 the ordinal MAAS scale to an interval level data (Medvedev et al., 2015).
36
37
38
39
40
41
42
43
44

45 Rasch analysis can be beneficial to improve the precision of the instrument, given its distinct
46
47 advantages over classical psychometric methods which have been well argued elsewhere (Rasch,
48
49 1960; Wilson, 2005; Wright & Stone, 1979). Rasch analysis involves a unidimensional measurement
50
51 model (Rasch, 1961) and in the case of the KIMS, Rasch analysis will be applied to each of the four
52
53 KIMS subscales individually due to multidimensionality, low correlation between the subscales, and
54
55 lack of support for an overarching mindfulness factor (Baer et al., 2004; Baum et al., 2010).
56
57
58
59
60
61
62
63
64
65

1
2 Rasch analysis investigates several parameters of the model including local independence
3 assumptions, potential item bias, unidimensionality, and stochastic ordering of items and response
4 options in polytomous items (Tennant & Conaghan, 2007). When fit to the Rasch model is achieved,
5 these parameters satisfy the model requirements and participants can be ordered according to their
6 ability or level on the latent trait (in this case one of the mindfulness skills). Similarly, items are
7 ordered by difficulty or level of the specific mindfulness trait assessed by an item. In Rasch analysis,
8 both the items and the participants are ordered on the same log-odds interval scale. Therefore, a
9 participant-item threshold distribution can be presented graphically to illustrate how well the range
10 of item difficulty covers or targets the abilities of the sample. The other benefit of Rasch analysis is
11 identification of item bias or Differential Item Functioning (DIF). DIF is evident if participants with
12 the same level on the latent trait, but from different groups (e.g. females and males), respond
13 differently to an item.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

30 The KIMS (Baer et al., 2004) is a widely-used multidimensional measure of four mindfulness traits
31 with generally accepted psychometric properties. The main purpose of the KIMS was to be used in
32 mindfulness-based treatment and studies, and it is critical to establish precision of its subscales.
33 However, the ability of the KIMS subscales to precisely discriminate between trait levels and the
34 functioning of individual items has not been investigated rigorously. Rasch analysis is a suitable
35 method to investigate the performance of individual items to discriminate on their overarching trait,
36 but to date Rasch analysis has not been used to study the psychometric properties of the KIMS
37 subscales. The aim of this study is to apply Rasch analysis to investigate the psychometric properties
38 of the KIMS and to explore strategies to improve precision and item functioning of its subscales.
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 **Method**

54 *Participants*

55
56
57
58
59
60
61
62
63
64
65

1
2 This study analyzed data from a sample of 287 New Zealand university students (78.7% females,
3 19.2% males, 2.1% missing gender). The mean age was 23.05 (SD=7.64), with ages ranging from 18
4
5 to 59. Ethnicities included 51.3% Caucasians, 8.7% Māori, 7.7% Pasifika, 19.7% Asians, and 12.2%
6
7 of unspecified others. The sample size met recommended optimal sample size estimates for Rasch
8
9 analysis (Linacre, 1994). To investigate DIF, two age categories were created: 18-20 (n=149) and
10
11 21-59 (n=127). There were 43 (15%) individuals regularly engaging in mindfulness practice, as
12
13 opposed to 240 not engaging in regular practice, and 4 individuals with data missing. Therefore, DIF
14
15 was tested for the person factors including gender, ethnic group, age, and engagement in meditation
16
17 and relaxation practices, where meditation practice refers to regularly performing formal meditation
18
19 exercises and relaxation practices refers to regular exercises such as yoga or progressive muscle
20
21 relaxation.
22
23
24

25 26 27 *Procedure*

28
29 Potential participants were approached in lectures and invited to complete the survey and to hand the
30
31 survey back to the researchers or submit it to a locked collection box at their respective faculty.
32
33

34
35 Students completed the questionnaire in class before the lecture or during a break. The authors'
36
37 university ethics committee approved this study.
38
39

40 41 *Measures*

42
43 The Kentucky Inventory of Mindfulness Skills (KIMS) is a 39-item self-report questionnaire
44
45 developed to capture the four mindfulness skills acquired in DBT treatment including Accept Non-
46
47 Judgementally, Observe, Act With Awareness and Describe (Baer et al., 2004). There are 16
48
49 negatively worded items measuring absence of mindfulness skills including items 3, 4, 8, 11, 14, 16,
50
51 18, 20, 22, 24, 27, 28, 31, 32, 35 and 36. These items were reversed coded prior to statistical
52
53 analysis.
54
55
56
57
58
59
60
61
62
63
64
65

Data Analysis

Prior to Rasch analysis, basic psychometric properties including reliability and factor structure of the KIMS were tested to be used for later comparisons with the Rasch results. Descriptive statistics, reliability analysis, and exploratory factor analysis were completed using IBM SPSS v.22, and Rasch analysis was completed using the software RUMM2030 (Andrich, Sheridan, & Luo, 2009). Rasch analysis is a unidimensional measurement model that involves testing of unidimensionality along with other psychometric criteria. Therefore, unidimensionality of the KIMS subscales and their fit to the Rasch model were analyzed separately for each subscale of the KIMS including: Accept Non-Judgementally, Observe, Describe, and Act With Awareness. First, the likelihood-ratio test was conducted on the initial analysis output for each subscale to confirm appropriateness of the unrestricted (Partial-Credit) version of the model. Rasch analyses followed ten main steps described elsewhere (Siegert, Tennant, & Turner-Stokes, 2010):

1. Overall goodness of fit test of the data to the Rasch model.
2. Identifying and rescoreing disordered thresholds.
3. Identifying and removing items not fitting to the Rasch model.
4. Re-test of overall and individual item fit to the Rasch model.
5. Examining Differential Item Functioning (DIF) for personal factors (e.g. gender).
6. Unidimensionality test of the four KIMS subscales.
7. Examining the residual correlation matrix for local dependency.
8. Inspection of the participant-item thresholds distribution plot.
9. Equating test between the final solution for a subscale (steps 1-8) and the original subscale.
10. Post-hoc comparisons between Rasch results and traditional psychometric criteria (e.g. reliability, factor analysis).

Rasch analysis starts with the overall evaluation of the model fit followed by identifying any items

1 displaying disordered thresholds by examining the threshold map. A threshold is disordered when
2 higher ability of an individual on a latent trait (e.g. Observe) is not consistently reflected in a gradual
3
4 increase in ordinal response options for that specific item. Typically, ordering of thresholds is
5
6 achieved by collapsing closest response categories. Overall goodness of fit to the model is usually
7
8 retested after items with disordered thresholds have been rescored. After ordering of thresholds is
9
10 completed, items showing the poorest fit to the model are removed one at a time and the overall
11
12 model fit is again tested.
13
14
15
16
17

18 Rasch analysis follows this iterative process until achieving the following criteria: Both overall and
19
20 individual item fit to the model are acceptable, and unidimensionality is clearly evident. Item-trait
21
22 interaction indexed by an overall and individual item chi-square fit statistic should be not significant
23
24 ($p > 0.05$, Bonferroni adjusted). Also, overall person and item fit-residuals should have a mean close
25
26 to 0.00 and a standard deviation close to 1.00 in the case of an excellent fit. Individual items are
27
28 expected to have fit residuals in the range between -2.50 and +2.50. All items should be locally
29
30 independent, which is tested by examining the residual correlations between items. According to the
31
32 conservative estimation (Marais & Andrich, 2008), the residual correlations between items
33
34 exceeding 0.20 above the mean of all residual correlations are indicative of local dependency. If
35
36 residual correlations between items are above this level, instead of removing locally dependent
37
38 items, these items are combined into a subtest (Wainer & Kiely, 1987), and the overall model fit is
39
40 once again examined. Also, the Rasch model requires no significant differences in DIF (Bonferroni
41
42 adjusted) for age, gender, and other personal factors. If any of the examined items shows DIF for a
43
44 specific personal factor, that item can be split into relevant categories, without the need to remove it
45
46 (Wainer & Kiely, 1987), followed by re-testing of the overall model fit. Dimensionality is examined
47
48 by an independent-samples t test comparing person-locations for two groups of items with the
49
50 highest positive and the highest negative loadings on the first principal component after a latent trait
51
52 (Rasch) factor is removed (Smith, 2002). Unidimensionality was tested for each subscale
53
54
55
56
57
58
59
60
61
62
63
64
65

individually because of multidimensionality and weak relationships between subscales of the KIMS. For instance, if a subscale has 10 items then the items with the highest positive factor loadings above 0.2 will form one set (e.g. 3 items) and equal number of items with the highest negative loadings (e.g. 3 items) will form another set. Then the estimates of each individual on these two sets of items will be compared by paired-samples *t* tests meaning that 287 *t* tests will be conducted for the current sample ($n=287$). Then the percentage of significant *t* tests will be computed together with the +/- 95% binominal confidence interval (CI). If the lower bound of the binominal CI computed for this percentage of significant *t* tests is smaller than 5% (e.g. 4%), then, based on statistical convention of alpha 0.05, we do not have any reason to believe that there is a real difference between the estimates, and unidimensionality is therefore evident. Additionally, reliability of the subscales was assessed by the person separation index (PSI) in Rasch analysis, which measures the ability of a scale to discriminate between groups at different levels of the latent trait. PSI is numerically close to Cronbach's alpha and similarly estimates the error-free proportion of variance of person estimates distribution relative to the total variance including error. Unlike Cronbach's alpha, PSI calculation involves non-linear transformation of the raw scores and can be performed with random missing data, which is an advantage of Rasch analysis.

Results

Exploratory factor analysis (EFA) using the principal axis factoring extracted ten factors with eigenvalues above 1.00. However, the largest amount of variance in the data (44%) was explained by just four factors and supported by the clear cut-off point on the scree-plot, which was consistent with the original report (Baer et al., 2004). For that reason, the number of extracted factors was fixed to four. Applying Varimax rotation with Kaiser normalization yielded the factor loadings presented in Table 1, with items generally following the factor structure of the original study (Baer et al., 2004). The internal consistency of the full 39-item scale was satisfactory with Cronbach's alpha of 0.82. However, some individual item-to-total correlations for the full scale were low, ranging from -0.19

1
2 to 0.50 (mean $r=0.29$), with two items falling below 0.10 (item 8, “I tend to evaluate whether my
3
4 perceptions are right or wrong”, $r=-0.19$, and item 19, “When I do things, I get totally wrapped up in
5
6 them and don’t think about anything else”, $r=0.03$). The low correlations (0.10 to 0.30) found
7
8 between the subscales were consistent with the original validation report (Baer et al., 2004) and
9
10 provide additional evidence for multidimensionality of the KIMS.

11
12 **<PLEASE INSERT TABLE 1 ABOUT HERE>**

13
14 Table 2 shows means and standard deviations for meditators and non-meditators together with
15
16 Cronbach’s alpha coefficients for each subscale of the KIMS. Cronbach’s alpha for the subscales
17
18 was in the acceptable range with an exception of the Act With Awareness subscale ($\alpha=0.65$).
19
20 According to expectations, significantly higher mean values were observed for meditators compared
21
22 to non-meditators, with the exception of Accept Non-Judgmentally subscale as evidenced by
23
24 subsequent t tests (Observe: $t(276) = -4.49, p < .001$; Describe: $t(274) = -2.42, p = .016$; Act With
25
26 Awareness: $t(274) = -2.26, p = .024$).

27
28
29
30
31 **<PLEASE INSERT TABLE 2 ABOUT HERE>**

32 33 *Rasch Analysis for Accept Non-Judgementally Subscale*

34
35 The initial model fit statistics for the 9-item Accept Non-Judgementally subscale are presented in
36
37 Table 3. PSI of 0.88 confirmed satisfactory reliability of the subscale, and none of the items
38
39 displayed disordered thresholds. However, the overall model fit was poor ($\chi^2(36)=157.43, p<0.001$),
40
41 and item 8 displayed an extremely high fit residual of 8.08, well above the 2.50 cut-off point (Table
42
43 1). Table 1 shows the location or difficulty of each item on the Rasch scale in probability units or
44
45 logits.
46
47
48

49
50 Deletion of item 8 resulted in a satisfactory overall model fit ($\chi^2(40)=45.43, p>0.05$) and acceptable
51
52 reliability (PSI=0.89) of the subscale. Following the deletion of item 8, item 24 also exhibited an
53
54 unacceptably high fit-residual of 3.64 and was therefore also removed before the analysis continued.
55
56
57
58
59
60
61
62
63
64
65

1
2 The overall model fit improved after the deletion of item 24 ($\chi^2(35)=34.89$, $p=0.47$), and no other
3
4 misfitting items were identified. At this stage, the residual correlation matrix was examined, and
5
6 local dependencies were found between items 4 and 12, and between items 16 and 32, as evidenced
7
8
9 by residual correlations exceeding the 0.20 limit above the mean of all residual correlations. After
10
11 combining two pairs of locally dependent items into two subtests, the overall good fit to the Rasch
12
13 model was further improved ($\chi^2(25)=20.97$, $p=0.69$, Table 3, ‘Accept Non-Judgementally’, Final
14
15 Analysis). At this stage, all individual items had acceptable fit to the model, and no other locally
16
17
18
19
20 dependent items could be identified.

21
22
23 **<PLEASE INSERT TABLE 3 ABOUT HERE>**

24
25
26 To test unidimensionality, the set of person estimates from the items with the highest positive
27
28 loadings on the first principal component were compared with the set of estimates from the items
29
30 with the highest negative loadings by an independent-samples t test. After calibrating t tests between
31
32 both sets of estimates to the same metric, 20 t test comparisons out of 287 (6.97%) were significant.
33
34 A binominal test was used to calculate the precise amount of acceptable deviations for the current
35
36 sample. Unidimensionality of the final solution was confirmed by the overlap of the 5 % cutoff point
37
38 on the lower bound surrounding t test (Table 3, ‘Accept Non-Judgementally’, Final Analysis). No
39
40 DIFs were found for person factors including gender, ethnic group, age, and engagement in
41
42
43 meditation and relaxation practices.
44
45
46

47
48
49 The person-item threshold distribution plot for the Accept Non-Judgementally subscale (Final
50
51 Analysis) is presented in Figure 1. The plot represents the relationship between distribution of item
52
53 difficulty and person ability on the latent trait (e.g. Accept Non-Judgementally) converted to the
54
55 same metric in logit units. Distribution of person thresholds is close to normal with some signs of
56

1 ceiling and floor effects. However, over 90% of the sample was adequately covered by the items of
2 the modified subscale.

3
4
5
6 **<PLEASE INSERT FIGURE 1 ABOUT HERE>**

7
8 *Rasch Analysis for Observe Subscale*

9
10 Initial analysis conducted for the 12-item Observe subscale yielded acceptable chi-square
11 ($\chi^2(48)=63.02$, $p=0.07$) and reliability (PSI=0.82) values. However, t test comparisons between two
12 sets of estimates with highest and lowest loadings on the first principal component after removing
13 the latent trait component failed to confirm unidimensionality, with 8.71% of significant t tests and
14 lower bound overlap above 5% (Table 4). Also, item 29 displayed disordered thresholds and needed
15 to be rescored before the analysis continued. After collapsing response options “Never or very rarely
16 true” and “Rarely true”, and “Sometimes true” and “Often true”, thresholds of item 29 were precisely
17 ordered. Figure 2 illustrates item response probability curve for item 29 before and after rescoring.

18
19
20
21
22
23
24
25
26
27
28
29
30
31 **<PLEASE INSERT FIGURE 2 ABOUT HERE>**

32 The overall model fit was slightly improved after rescoring item 29 ($\chi^2(48)=62.12$, $p=0.09$), and all
33 individual items showed a good fit to the Rasch model (Table 4, Observe). However,
34 unidimensionality of the subscale was not confirmed. Unidimensionality of a scale can be
35 compromised by locally dependent items, which can be identified by residual correlations between
36 them. Therefore, the residual correlation matrix was examined, and local dependency was found
37 between items 21 and 25, and items 30 and 37, which were then combined into two subtests. This
38 minor modification produced the final solution for this subscale, with overall good fit to the model
39 ($\chi^2(40)=40.50$, $p=0.45$) and acceptable reliability (PSI=0.80) (Table 4). A binominal test to test
40 unidimensionality of the final solution indicated overlap on the lower bound surrounding t test with
41 the 5 % cutoff point, which confirmed unidimensionality (Table 4, ‘Observe’, Final Analysis). No
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2 DIF was noted for personal factors, such as gender, ethnic group, age, and meditation and relaxation
3 practices.

4
5 Figure 1 shows the item-person threshold distribution for the final solution of the Observe subscale.
6
7 Overall, person thresholds are distributed close to normal and well targeted by item thresholds, but
8 there are some signs of a small ceiling effect. In this analysis, a good fit of the Observe subscale to
9 the Rasch model was achieved with the minor modifications of rescoreing one item and creating two
10 subtests, without a need to remove any misfitting items.
11
12
13
14
15

16 17 18 19 *Rasch Analysis for Describe Subscale*

20
21 Initial analysis of the 8-item Describe subscale indicated acceptable reliability (PSI= 0.85) but an
22 overall lack of fit to the model ($\chi^2(32)=54.40$, $p=0.008$) and lack of evidence for unidimensionality
23 (Table 4). At this stage no items displayed unacceptably high fit residuals. Therefore, the residual
24 correlation matrix was examined, indicating local dependency between items 14, 18, and 22, which
25 were then combined into a subtest before analysis continued. After creating the subtest, the chi-
26 square had a lower but still significant value ($\chi^2(24)=39.12$, $p =0.03$). However, at the individual
27 item level, item 6 displayed a high fit residual of 2.81, above the 2.50 cut-off point and was removed,
28 which resulted in a good overall model fit ($\chi^2(20)=22.52$, $p=0.31$) and continued acceptable
29 reliability (PSI=0.78) (Table 4, Describe, Final). At the individual item level, all items showed
30 acceptable fit to the model. Unidimensionality of the final solution was confirmed by the binominal
31 test indicating overlap on the lower bound surrounding t test with the 5 % cutoff point. No
32 significant DIFs were found for personal factors.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

50
51 The person-item threshold distribution for the final analysis of the Describe subscale shows
52 acceptable targeting of the person locations by the item thresholds (Figure 1). However, a slight
53 ceiling effect was apparent indicating some limitation of the subscale in measuring higher personal
54
55
56
57
58
59
60
61
62
63
64
65

abilities on Describe. Thus, satisfactory fit to the model was evident after few modifications that involved combining locally dependent items into a subtest and removal of one non-fitting item.

Rasch Analysis for Act With Awareness Subscale

Initial testing of the 10-item Act With Awareness subscale of the KIMS revealed an overall lack of fit to the model with a significant chi-square for overall person-trait interaction ($\chi^2(40)=67.39$, $p=0.004$) and reliability (PSI) of 0.66 (Table 4, Act With Awareness, Initial). The assumption of unidimensionality was violated as indicated by the binominal test not overlapping the 5 % cutoff point on the lower bound. Consequently, individual item fit statistics were examined, indicating that item 35 had a significantly high fit residual ($\chi^2(4)=15.75$, $p < 0.001$). Item 35 was thus deleted, resulting in acceptable overall model fit, with a non-significant chi-square value ($\chi^2(36)=41.20$, $p=0.25$) but a slight reduction in reliability (PSI=0.61). Additionally, evidence for unidimensionality was lacking as the binominal test indicated no overlap with the 5% cutoff point on the lower bound. To test for local dependency, the residual correlation matrix was examined showing that item 19 had high residual correlations with items 7 and 38 that exceeded the 0.20 cut-off point above the mean of all residual correlations. Item 19 also had the lowest item-to-total correlation of 0.19 (Table 4) and so was removed resulting in the final solution with an acceptable chi-square value ($\chi^2(4)=38.40$, $p=0.202$) and evidence for unidimensionality (Table 4, Act With Awareness, Final). However, the reliability of the Act With Awareness subscale decreased slightly after these modifications (PSI=0.60). At this stage, all individual items were showing good fit to the model, and no DIF was identified for any personal factors.

The bottom panel of Figure 1 shows the item-person thresholds distribution for the final solution of the Act With Awareness subscale, with fairly good coverage of the range of individuals locations by the subscale items thresholds. Thus, the Act With Awareness subscale modified by removing two misfitting items satisfied all but reliability criteria for fitness to the Rasch model.

Ordinal-to-Interval Scale Conversion Table

1
2
3 Table 4 includes conversion scores from an ordinal-to-interval level scale for all four subscales of the
4
5 KIMS. For convenience, all the scores are adjusted to the scoring algorithm of each original
6
7 subscale, and the total ordinal score is calculated by adding the scores of all the individual items
8
9 included in a final version of a subscale after negatively worded items are reverse coded. Also, item
10
11 29 has to be rescored according to the algorithm at the bottom of Table 4, before calculating the total
12
13 score for the Observe subscale. Therefore, ordinal scores are represented on the left-hand side and
14
15 corresponding Rasch interval-level scores on the right-hand side (Table 4). The conversion table
16
17 provided here allows users to increase precision of the KIMS subscales without the need to modify
18
19 the original response format of the scale. These conversions can only be used when there are no
20
21 missing data.
22
23
24
25

26
27
28 **<PLEASE INSERT TABLE 4 ABOUT HERE>**
29
30
31

32
33 **Discussion**
34
35

36 The KIMS (Baer et al., 2004) is a multidimensional measure of mindfulness widely used to assess four
37
38 mindfulness-related skills: Observing, Describing, Act With Awareness, and Accept Non-
39
40 Judgementally. These mindfulness skills are linked to therapeutic outcomes of MBIs and especially
41
42 in the context of DBT (Baer et al., 2004; Dimidjian & Linehan, 2003). Our intention was not to
43
44 assess the overall construct validity of the KIMS as such reports are already available (Baer et al.,
45
46 2004; Park et al., 2013) but rather to perform investigation and fine-tuning for each individual
47
48 subscale of the KIMS and test their structural validity using Rasch analysis. Thus, the aim of the
49
50 current study was to use strategies of Rasch analysis to improve precision and item functioning of
51
52 the KIMS subscales.
53
54
55

56
57
58 Given the fact that the subscales of the KIMS are commonly found only to be loosely related to each
59
60 other (Baum et al., 2010), a finding confirmed in the present study, we focused our analyses on the
61
62
63
64
65

1 subscales. Accuracy of any ordinal scale is limited, but can be improved up to an interval level using
2 the Rasch model (Tennant & Conaghan, 2007). The results of this analysis show that precision and
3
4 item functioning of the KIMS subscales can be improved substantially using the Rasch model.
5 Satisfactory fit to the Rasch model was achieved by a few modifications of the KIMS subscales that
6
7 involved rescoring one item, deleting non-fitting items and combining locally dependent items into
8
9 subtests (Table 4). Generally, the results support structural validity and reliability of the modified
10
11 KIMS subscales to measure mindfulness skills in the sample population. However, initially low
12
13 reliability of Act With Awareness subscale remained at the level of 0.60, indicating that the subscale
14
15 is unable to distinguish between 2 strata with different ability levels (Fisher, 1992), which limits its
16
17 applicability. The Rasch analysis confirms unidimensionality of all four modified KIMS subscales
18
19 meaning that raw scores can be readily transformed into interval level scores using the same metric
20
21 (Table 4) as long as there are no missing data. Clinicians and researchers may use the conversion
22
23 algorithms of Table 4 to transform ordinal data into interval-level scores to investigate precise effects
24
25 of MBIs on specific mindfulness traits.
26
27
28
29
30
31
32
33
34
35
36

37 The modification of the KIMS subscales involved rescoring of only one item (item 29) that displayed
38
39 disordered thresholds in the Observe subscale, which indicates that the response options selected by
40
41 the authors (Baer et al., 2004) were overall appropriate. Also, only five items needed to be removed
42
43 as they did not fit the Rasch model. No items were removed from the Observe subscale, and only
44
45 item 6 (“I can easily put my beliefs, opinions, and expectations into words”) was removed from the
46
47 Describe subscale due to its poor fit. This item might not fit well to the model because it is not
48
49 focused on describing experiences, which is the primary target of the Describe subscale. Also,
50
51 focusing on describing beliefs, opinions, and expectations could be seen as moving away from the
52
53 present moment and thus associated with less mindfulness. Two semantically close non-fitting items
54
55 were deleted from the Accept Non-Judgementally subscale, namely item 8 (“I tend to evaluate
56
57
58
59
60
61
62
63
64
65

1 whether my perceptions are right or wrong”) and item 24 (“I tend to make judgments about how
2 worthwhile or worthless my experiences are”). Item 8 had the lowest item-to-total correlation for
3
4 both the full scale ($r=-0.19$) and its subscale ($r=0.21$), and it is the most difficult item where just few
5 individuals scored high. Item 24 also seems to measure dichotomous judgement but is worded in a
6
7 more adaptive way. These items might not work well with the current sample because they focus on
8
9 extreme dichotomous judgement attitudes perhaps more common among people with borderline
10
11 personality disorder (Linehan, 1993a, 1993b). Finally, two non-fitting items were removed from the
12
13 Act With Awareness subscale: item 19 (“When I do things, I get totally wrapped up in them and
14
15 don’t think about anything else”) and item 35 (“When I’m working on something, part of my mind is
16
17 occupied with other topics, such as what I’ll be doing later, or things I’d rather be doing”). Item 19
18
19 had very low item-to-total correlations for the full scale ($r=0.03$) and its subscale ($r=0.16$) and seems
20
21 to measure one-pointed concentration rather than mindfulness (Brown, Ryan, & Creswell, 2007;
22
23 Olendzki, 2005). Item 35 seems to be a negatively worded counterpart of item 19 and it might also
24
25 lead to bias due to its relatively complex wording. Future studies might examine how rewording of
26
27 these items may improve psychometric properties of the KIMS subscales. However, until such
28
29 investigation has been completed, it is recommended to use the proposed subscales versions together
30
31 with the ordinal-to-interval scoring algorithm.

32
33 Removing non-fitting items resulted in the 34-item KIMS version that included the following
34
35 modified subscales: the 7-item Accept Non-Judgementally, the 7-item Describe and the 8-item Act
36
37 With Awareness. The original 12-item Observe subscale retained all its items after the modification.
38
39 Combining locally dependent items into subtests was beneficial in achieving the satisfactory model
40
41 fit for the subscales Observing, Describing, and Accept Non-Judgementally, without the need to
42
43 discard more items. The psychometric properties of the modified KIMS subscales are, therefore,
44
45 improved substantially with the exception of Act With Awareness subscale, which had initially low
46
47 internal consistency that could not be improved by the current modifications.
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Unlike in the Rasch analysis of the MAAS (Medvedev et al., 2015), where uniform rescaling of all
2 items was conducted to correct disordered thresholds, the KIMS items displayed no disordered
3
4 thresholds, with an exception of item 29, which supports psychometric properties of the KIMS.
5
6 Similarly to the MAAS, no more than two items were removed per subscale to achieve a satisfactory
7
8 fit to the Rasch model, thus providing support for overall good structural validity of both scales.
9
10 Also, four items (6, 8, 24 and 35) out of the five non-fitting items identified through Rasch analysis
11
12 of the KIMS were included in the FFMQ constructed through factor analysis of the available
13
14 mindfulness questionnaires (Baer et al., 2006), and it might be worthwhile to investigate functioning
15
16 of those items also in the FFMQ using Rasch analysis. In addition, consistent with the earlier reports
17
18 (Park et al., 2013), significantly higher mean scores were found for meditators compared to non-
19
20 meditators for all but Accept Non-Judgmentally subscale of KIMS, which supports construct validity
21
22 of these subscales.
23
24

25
26
27 The following limitations should be noted. The study was conducted with a single sample of
28
29 university students and should be replicated with more diverse samples including clinical and general
30
31 populations. The analyses might also have been affected by degree of missing data as well as
32
33 disproportional distribution of gender, age, and formal meditation experience in the sample.
34
35
36

37
38 Although the sample reflects New Zealand's diversity of ethnic groups (Statistics New Zealand,
39
40 2013), no efforts were made to purposively sample underrepresented groups. Even though a
41
42 satisfactory fit to the Rasch model was achieved for all subscales of the KIMS, the reliability of Act
43
44 With Awareness subscale could not be improved, and item 29 required rescaling before computing a
45
46 total score of the Observe subscale. However, if you have complete data, conversion from ordinal to
47
48 an interval level scale can be conducted simply by adding responses on each modified subscale
49
50 version and selecting a corresponding interval score in the right column (Table 4). The benefits of
51
52 this conversion certainly outweigh inconvenience, and the authors can be contacted if assistance with
53
54 data conversion is necessary. In addition, an Excel spreadsheet will be provided online to simplify
55
56
57
58
59
60
61
62
63
64
65

1 data conversion.
2
3
4

5 Mindfulness skills have emerged as important contributors of therapeutic outcomes and their
6 accurate assessment represents an on-going challenge. The current study reported Rasch analysis
7 conducted to advance psychometric properties of the widely-used KIMS, a multidimensional
8 measure of four mindfulness traits. We demonstrated that the KIMS subscales are structurally (or
9 internally) valid after modifications that involved rescoreing item 29, removing misfitting items 6, 8,
10 19, 24, and 35, as well as combining locally dependent items into subtests. Precision of the KIMS
11 can be improved substantially by using the proposed 34-item version of the instrument together with
12 the ordinal-to-interval conversion table presented here (Table 4), without any need to modify the
13 original response format. These findings will be of interest for clinicians applying mindfulness-based
14 interventions and researchers investigating neurophysiological and psychological correlates of trait
15 mindfulness.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

33 **Compliance with Ethical Standards**

34 We declare that we have no conflicts of interest in connection with this paper. The study was
35 conducted in compliance with the guidelines of the authors' university ethics committee.
36
37
38
39
40
41

42 **References**

- 43 Andrich, D., Sheridan, B., & Luo, G. (2009). *RUMM 2030*. Perth: RUMM Laboratory.
44
45 Baer, R. (2003). Mindfulness training as a clinical intervention: a conceptual and empirical
46 review. *Clinical Psychology: Science and Practice*, 10(2), 125-142.
47
48 Baer, R., Smith, G., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment
49 methods to explore facets of mindfulness. *Assessment*, 13(1), 27-45.
50
51 Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: the
52 Kentucky Inventory of Mindfulness Skills. *Assessment*, 11(3), 191-206.
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Baum, C., Kuyken, W., Bohus, M., Heidenreich, T., Michalak, J., & Steil, R. (2010). The Psychometric Properties of the Kentucky Inventory of the Mindfulness Skills in Clinical Populations. *Assessment, 17*(2), 220-229.
- 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
- Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., ... ,& Devins, G. (2006). Mindfulness: a proposed operational definition. *Clinical Psychology: Science and Practice, 11*(3), 230-241.
- Brogden, H. E. (1977). The Rasch model, the law of comparative judgment and additive conjoint measurement. *Psychometrika, 42*(4), 631-634.
- 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 (PMI). *Journal for Meditation and Meditation Research, 1*, 11-34.
- 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.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Chambers, R., Gullone, E., & Allen, N. B. (2009). Mindfulness emotion regulation: an integrative review. *Clinical Psychology Review, 29*(6), 560-572.
- 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., & 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.
- Dimidjian, S., & Linehan, M. M. (2003). Defining an agenda for future research on the clinical application of mindfulness practice. *Clinical Psychology: Science and Practice, 10*(2), 166-171.
- 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.
- Fisher, W. (1992). Reliability statistics. *Rasch Measurement Transactions, 6*(3), 238.
- 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.
- Goh, H. E., Marais, I., & Ireland, M. J. (2015). A Rasch model analysis of the Mindful Attention Awareness Scale. *Assesment*. doi:DOI: 10.1177/1073191115607043
- Hayes, S. C., Follette, V. M., & Linehan, M. M. (Eds.). (2004). *Mindfulness and acceptance: expanding cognitive-behavioral tradition*. New York: The Guilford Press.
- Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and commitment therapy: an experiential approach to behavior change*. New York: Guilford Press.

- 1
2 Hofmann, S. G., Sawyer, A. T., Witt, A. A., & Oh, D. (2010). The effect of mindfulness-based
3 therapy on anxiety and depression: a meta-analytic review. *Journal of Consulting and*
4 *Clinical Psychology, 78*(2), 169-183.
- 5
6
7 Inchausti, F., Prieto, G., & Delgado, A. R. (2014). Rasch analysis of the Spanish version of the
8
9 Mindful Attention Awareness Scale (MAAS) in a clinical sample. *Revista de Psiquiatría y*
10 *Salud Mental, 7*(1), 32-41.
- 11
12
13
14
15 Josefsson, T., Lindwall, M., & Broberg, A. G. (2014). The effects of a short-term mindfulness
16
17 based intervention on self-reported mindfulness, decentering, executive attention,
18
19 psychological health, and coping style: examining unique mindfulness effects and
20
21 mediators. *Mindfulness, 5*(1), 18-35.
- 22
23
24
25 Kabat-Zinn, J. (1982). An outpatient program in behavioural medicine for chronic pain patients
26
27 based on the practice of mindfulness meditation: theoretical considerations and
28
29 preliminary results. *General Hospital Psychiatry, 4*(1), 33-47.
- 30
31
32 Kabat-Zinn, J. (1990). *Full catastrophe living: using the wisdom of your body and mind to face*
33
34 *stress, pain and illness*. New York, USA: Delacorte.
- 35
36
37 Kabat-Zinn, J. (1994). *Wherever you go, there you are: mindfulness meditation in everyday life*.
38
39 New York, USA: Hyperion.
- 40
41
42 Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement*
43
44 *Transactions, 7*, 328.
- 45
46
47 Linehan, M. M. (1993a). *Cognitive-behavioural treatment of borderline personality disorder*. New
48
49 York, USA: Guilford Press.
- 50
51
52 Linehan, M. M. (1993b). *Skills training manual for treating borderline personality disorder*. New
53
54 York, USA: Guilford Press.
- 55
56
57 Lyvers, M., Makin, C., Toms, E., Thorberg, F. A., & Samios, C. (2014). Trait mindfulness in
58
59 relation to emotional self-regulation and executive function. *Mindfulness, 5*(6), 619-625.
- 60
61
62
63
64
65

- 1
2 Marais, I., & Andrich, D. (2008). Effects of varying magnitude and patterns of response
3 dependence in the unidimensional Rasch model. *Journal of Applied Measurement*, 9(2),
4 105-124.
5
6
7 Medvedev, O. N., Siegert, R. J., Feng, X. J., Billington, D. R., Jang, J. Y., & Krägeloh, C. U. (2015).
8 Measuring trait mindfulness: how to improve the precision of the Mindful Attention
9 Awareness Scale using a Rasch model. *Mindfulness*. doi:10.1007/s12671-015-0454-z
10
11
12
13
14 Olendzki, A. (Ed.). (2005). *The roots of mindfulness*. New York, USA: Guilford.
15
16
17 Park, T., Reilly-Spong, M., & Gross, C. R. (2013). Mindfulness: a systematic review of
18 instruments to measure an emergent patient-reported outcome (PRO). *Quality of Life*
19 *Research*, 22(10), 2639-2659.
20
21
22
23
24 Pearson, M. R., Brown, D. B., Bravo, A. J., & Witkiewitz, K. (2015). Staying in the moment and
25 finding purpose: the associations of trait mindfulness, decentering, and purpose in life
26 with depressive symptoms, anxiety symptoms, and alcohol-related problems.
27
28
29
30
31
32
33
34
35
36
37
38
39 Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test*. Copenhagen:
40 Danish Institute for Educational Research.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *University of California Press*. Symposium conducted at the meeting of the Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability., Berkeley, California.
- Sauer, S., Ziegler, M., Danay, E., Ives, J., & Kohls, N. (2013). Specific objectivity of mindfulness - a Rasch analysis of the Freiburg Mindfulness Inventory. *Mindfulness*, 4(1), 45-54.
- Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2002). *Mindfulness based cognitive therapy for depression*. New York, USA: Guilford Press.
- Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2013). *Mindfulness-Based Cognitive Therapy for depression* (2nd ed.). New York: Guilford Press.

- 1
2 Siegert, R. J., Tennant, A., & Turner-Stokes, L. (2010). Rasch analysis of the Beck Depression
3 Inventory-II in a neurological rehabilitation sample. *Disability and Rehabilitation*, 32(1),
4
5 8-17.
6
7 Statistics New Zealand. (2013). *Census ethnic group profiles*. Wellington: New Zealand
8
9 Government.
10
11 Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is
12
13 it and why use it? When should it be applied, and what should one look for in a Rasch
14
15 paper? *Arthritis & Rheumatism*, 57(8), 1358-1362.
16
17
18 Visted, E., Vøllestad, J., Nielsen, M. B., & Nielsen, G. H. (2015). The impact of group-based
19
20 mindfulness training on self-reported mindfulness: a systematic review and
21
22 meta-analysis. *Mindfulness*, 6(3), 501-522.
23
24
25 Wainer, H., & Kiely, G. (1987). Item clusters and computerized adaptive testing: a case for
26
27 testlets. *Journal of Educational Measurement*, 24(3), 185-201.
28
29
30 Walach, H., Buchheld, N., Buttenmüller, V., Kleinknecht, N., & Schmidt, S. (2006). Measuring
31
32 mindfulness - the Freiburg Mindfulness Inventory (FMI). *Personality and Individual*
33
34 *Differences*, 40(8), 1543-1555.
35
36
37
38 Wilson, M. (2005). *Constructing measures*. Mahwah, NJ: LEA.
39
40
41 Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, USA: MESA Press.
42
43
44 Zoogman, S., Goldberg, S. B., Hoyt, W. T., & Miller, L. (2015). Mindfulness interventions with
45
46 youth: a meta-analysis. *Mindfulness*, 6(2), 290-302.
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

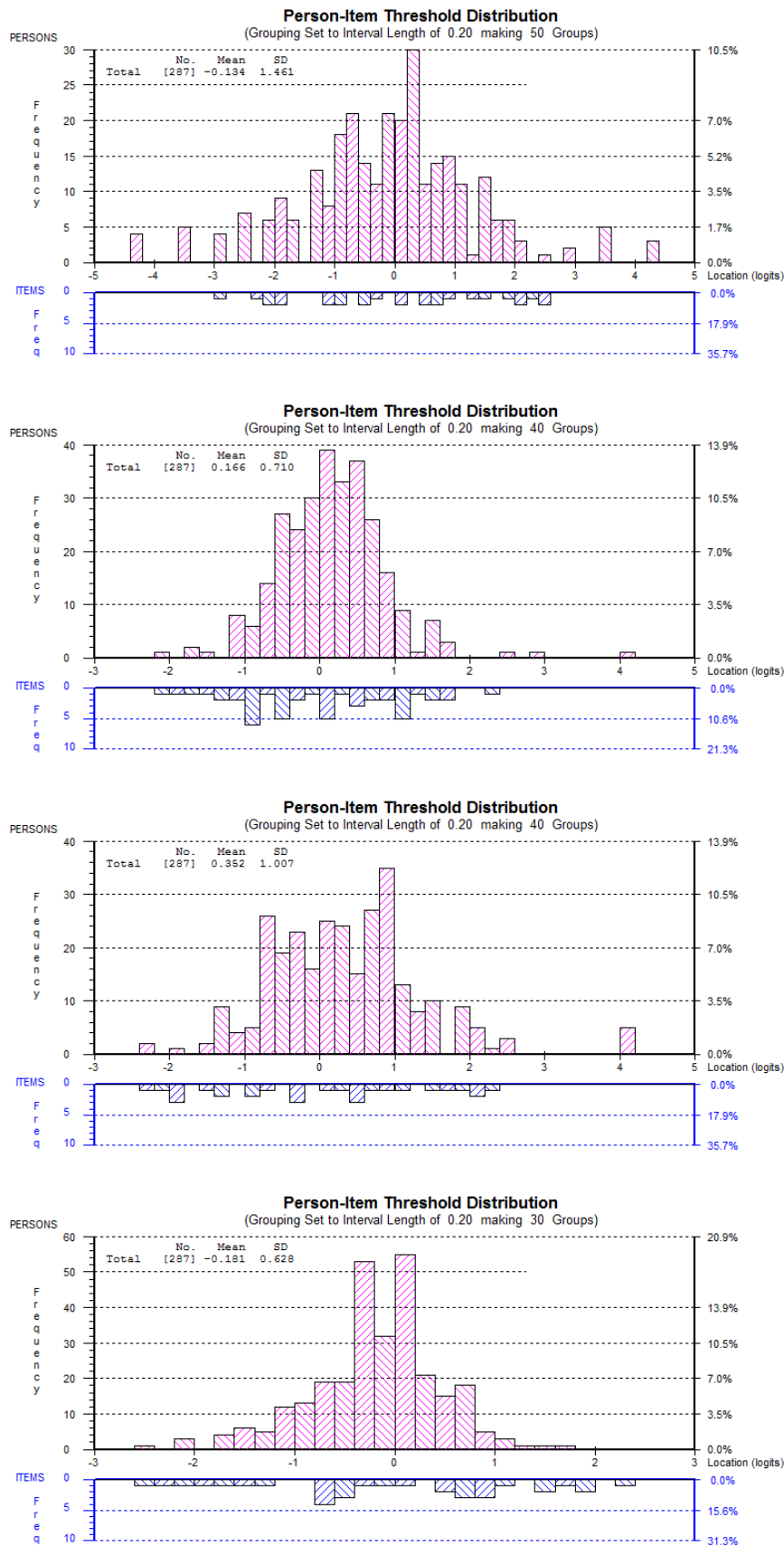
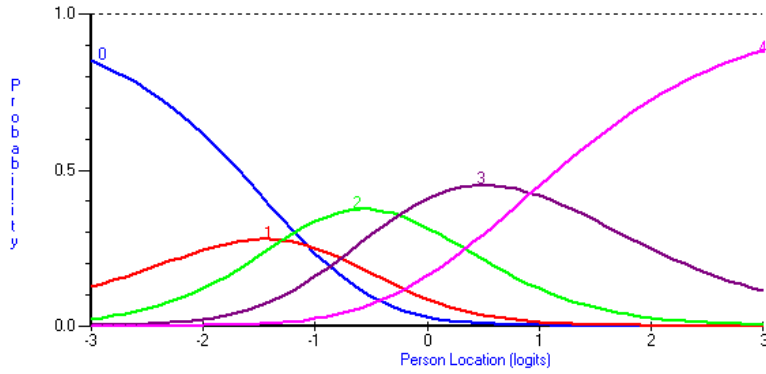


Figure 1. Person-item threshold distribution for modified KIMS subscales from top to bottom including *Accept*, *Observe*, *Describe* and *Act* ($n=287$).

I0029 I notice the smells and aromas Locn = -0.422 Spread = 0.352 FitRes = -0.119 ChiSq[Pr] = 0.733 F[Pr] = 0.771



I0029 I notice the smells and aromas Locn = -0.202 Spread = 1.562 FitRes = 0.081 ChiSq[Pr] = 0.151 F[Pr] = 0.121

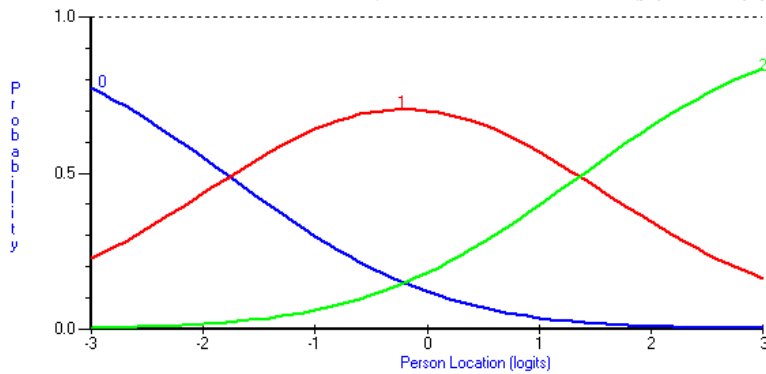


Figure 2. Item category probability curves illustrating disordered thresholds for KIMS item 29 before rescaling (top panel) and orderly thresholds after rescaling (bottom panel).

Table 1. *Initial item location, item fit residual, corrected item-to total correlation and factor loadings for Accept Non-Judgmentally (Accept), Observe, Describe and Act With Awareness (Act) subscale items of KIMS.*

N	Subscale/ Item	Item Location	Item Fit Residual	Item-Total	Factor Loading
<u>Accept</u>					
4	I criticize myself for having irrational or inappropriate emotions. ^R	0.29	-1.21	0.68	0.77
8	I tend to evaluate whether my perceptions are right or wrong. ^R	0.46	8.08	0.21	0.27
12	I tell myself that I shouldn't be feeling the way I'm feeling. ^R	0.09	-0.50	0.67	0.76
16	I believe some of my thoughts are abnormal or bad and I shouldn't think that way. ^R	-0.20	-0.86	0.68	0.76
20	I make judgments about whether my thoughts are good or bad. ^R	0.14	-1.27	0.70	0.77
24	I tend to make judgments about how worthwhile or worthless my experiences are. ^R	0.02	2.16	0.56	0.62
28	I tell myself that I shouldn't be thinking the way I'm thinking. ^R	-0.17	-2.37	0.74	0.83
32	I think some of my emotions are bad or inappropriate and I shouldn't feel them. ^R	-0.42	-2.47	0.73	0.82
36	I disapprove of myself when I have irrational ideas. ^R	-0.21	-0.45	0.68	0.74
<u>Observe</u>					
1	I notice changes in my body, such as whether my breathing slows down or speeds up.	-0.22	1.10	0.42	0.49
5	I pay attention to whether my muscles are tense or relaxed.	0.62	0.07	0.49	0.53
9	When I'm walking, I deliberately notice the sensations of my body moving.	0.72	0.60	0.47	0.60
13	When I take a shower or a bath, I stay alert to the sensations of water on my body.	0.17	1.23	0.45	0.56
17	I notice how foods and drinks affect my thoughts, bodily sensations, and emotions.	0.19	0.87	0.43	0.50
21	I pay attention to sensations, such as the wind in my hair or sun on my face.	0.11	-1.37	0.58	0.71
25	I pay attention to sounds, such as clocks ticking, birds chirping, or cars passing.	0.07	1.06	0.45	0.61
29	I notice the smells and aromas of things.	-0.42	-0.12	0.49	0.63
30	I intentionally stay aware of my feelings.	-0.10	-0.63	0.54	0.57
33	I notice visual elements in art or nature, such as colors, shapes, textures, or patterns of light and shadow.	-0.22	1.53	0.43	0.54
37	I pay attention to how my emotions affect my thoughts and behavior.	-0.43	0.12	0.48	0.49
39	I notice when my moods begin to change.	-0.50	1.75	0.37	0.37
<u>Describe</u>					
2	I'm good at finding the words to describe my feelings.	-0.04	-1.33	0.65	0.76
6	I can easily put my beliefs, opinions, and expectations into words.	-0.41	-1.34	0.68	0.73
10	I'm good at thinking of words to express my perceptions, such as how things taste, smell or sound	0.18	1.36	0.56	0.55
14	It's hard for me to find the words to describe what I'm thinking. ^R	-0.01	-2.24	0.69	0.78
18	I have trouble thinking of the right words to express how I feel about things. ^R	-0.02	-2.16	0.72	0.79
22	When I have a sensation in my body, it's difficult for me to describe it because I can't find the right words. ^R	-0.42	2.25	0.44	0.59
26	Even when I'm feeling terribly upset, I can find a way to put it into words.	0.37	3.64	0.51	0.60
34	My natural tendency is to put my experiences into words.	0.36	3.10	0.48	0.56
<u>Act</u>					
3	When I do things, my mind wanders off and I'm easily distracted. ^R	0.22	-0.77	0.43	0.54
7	When I'm doing something, I'm only focused on what I'm doing, nothing else.	-0.12	-0.15	0.40	0.55
11	I drive on "automatic pilot" without paying attention to what I'm doing. ^R	-0.48	1.81	0.24	0.34
15	When I'm reading, I focus all my attention on what I'm reading.	-0.60	1.23	0.28	0.47
19	When I do things, I get totally wrapped up in them and don't think about anything else.	-0.20	1.56	0.16	0.44
23	I don't pay attention to what I'm doing because I'm daydreaming, worrying, or otherwise distracted. ^R	-0.12	-0.27	0.39	0.37
27	When I'm doing chores, such as cleaning or laundry, I tend to daydream or think of other things. ^R	0.59	0.93	0.24	0.42
31	I tend to do several things at once rather than focusing on one thing at a time. ^R	0.21	0.42	0.31	0.49
35	When I'm working on something, part of my mind is occupied with other topics, such as what I'll be doing later, or things I'd rather be doing. ^R	0.52	-1.64	0.49	0.62
38	I get completely absorbed in what I'm doing, so that all my attention is focused on it.	-0.02	0.70	0.26	0.47

Note: ^R reverse-scored item.

Table 2. Descriptive statistics for the Accept Non-Judgmentally (Accept), Observe, Describe and Act With Awareness (Act) subscales of KIMS including means and standard deviations (SD) for meditators and non-meditators, and Cronbach's alpha coefficients.

Subscale	Meditators (n=42)		Non-Meditators (n=221)		Cronbach's alpha
	Mean	SD	Mean	SD	
<u>Accept</u>	25.40	7.74	25.70	7.37	0.88
<u>Observe</u>	43.19	6.80	37.76	7.31	0.82
<u>Describe</u>	26.90	4.91	26.44	5.77	0.85
<u>Act</u>	29.29	5.60	27.42	4.63	0.65

Table 3. Summary of fit statistics for the initial and the final Rasch analyses of the four KIMS subscales Accept Non-Judgmentally (Accept), Observe, Describe and Act With Awareness (Act).

Analyses	Item fit residual		Person fit residual		Goodness of fit		PSI	Independent t-test %LB ^a
	Value / SD		Value / SD		χ^2 (df)	p		
<u>Accept</u>								
Initial	0.12	3.27	-0.44	1.54	157.43 (36)	< 0.001	0.88	6.20
Final	0.31	1.21	-0.53	1.38	20.97 (25)	0.690	0.85	4.40
<u>Observe</u>								
Initial	0.52	0.93	-0.29	1.40	63.02 (48)	0.070	0.82	7.20
Final	0.55	0.62	-0.28	1.30	40.50 (40)	0.450	0.80	4.80
<u>Describe</u>								
Initial	0.41	2.44	-0.50	1.60	54.37 (32)	0.008	0.85	7.60
Final	0.62	1.62	-0.50	1.36	22.52 (20)	0.310	0.78	3.40
<u>Act</u>								
Initial	0.38	1.09	-0.38	1.44	67.39 (40)	0.004	0.67	8.30
Final	0.45	0.85	-0.38	1.36	38.40 (32)	0.202	0.60	4.40

Note: ^aLB = lower bound of the 95-% confidence interval.

Table 4. *Converting from ordinal to interval level scores for the subscales Accept Non-Judgmentally (Accept), Observe, Describe and Act With Awareness (Act) of the 34-item version of the KIMS.*

<u>Accept Scores</u>		<u>Observe Scores</u>				<u>Describe Scores</u>		<u>Act Scores</u>	
Ordinal	Interval	Ordinal	Interval	Ordinal	Interval	Ordinal	Interval	Ordinal	Interval
7	7.00	10	10.00	43	37.67	7	7.00	8	8.00
8	9.72	11	14.49	44	38.22	8	9.84	9	11.19
9	11.65	12	17.46	45	38.80	9	11.72	10	13.43
10	13.02	13	19.43	46	39.40	10	12.96	11	15.01
11	14.12	14	20.92	47	40.04	11	13.90	12	16.25
12	15.06	15	22.11	48	40.74	12	14.70	13	17.29
13	15.90	16	23.12	49	41.48	13	15.41	14	18.21
14	16.66	17	24.00	50	42.31	14	16.07	15	19.02
15	17.38	18	24.79	51	43.25	15	16.72	16	19.77
16	18.06	19	25.50	52	44.32	16	17.36	17	20.46
17	18.71	20	26.17	53	45.59	17	17.99	18	21.11
18	19.34	21	26.78	54	47.17	18	18.61	19	21.73
19	19.96	22	27.37	55	49.24	19	19.24	20	22.32
20	20.57	23	27.92	56	52.36	20	19.86	21	22.89
21	21.18	24	28.45	57	57.00	21	20.48	22	23.45
22	21.79	25	28.97			22	21.10	23	23.99
23	22.41	26	29.47			23	21.72	24	24.53
24	23.03	27	29.97			24	22.34	25	25.06
25	23.66	28	30.44			25	22.96	26	25.60
26	24.30	29	30.92			26	23.58	27	26.13
27	24.96	30	31.40			27	24.21	28	26.68
28	25.66	31	31.87			28	24.85	29	27.24
29	26.39	32	32.33			29	25.53	30	27.81
30	27.19	33	32.80			30	26.29	31	28.41
31	28.08	34	33.27			31	27.17	32	29.05
32	29.13	35	33.73			32	28.26	33	29.73
33	30.44	36	34.20			33	29.71	34	30.47
34	32.32	37	34.68			34	31.85	35	31.30
35	35.00	38	35.16			35	35.00	36	32.25
		39	35.64					37	33.39
		40	36.13					38	34.84
		41	36.63					39	36.95
		42	37.14					40	40.00

Note: Item 29 from the *Observe* subscale needs to be rescored before calculating the ordinal scores for this subscale as follows: 1=1, 2=1, 3=2, 4=2, 5=3. Negatively worded items 3, 4, 8, 11, 14, 16, 18, 20, 22, 24, 27, 28, 31, 32, 35 and 36 have to be reversed coded prior calculating the total score. For the *Accept* subscale, add items 4, 12, 16, 20, 28, 32, and 36. For the *Observe* subscale, add items 1, 5, 9, 13, 17, 21, 25, 29, 30, 33, 37, and 39. For the *Describe* subscale, add items 2, 6, 10, 14, 18, 22, 26, and 34. For the *Act* subscale, add items 3, 7, 11, 15, 19, 23, 27, 31, and 38. For each subscale sum score, find the equivalent interval-level score in the above conversion table. This table cannot be used for respondents with missing data.