ORIGINAL PAPER

Validation of the Comprehensive Inventory of Mindfulness Experiences (CHIME) in English Using Rasch Methodology

Samuel Wilkinson[1] · Leticia Ribeiro[2] · Christian U. Krägeloh[3] · Claudia Bergomi[4] · Marie Parsons[5] · Alex Siegling[6] · Wolfgang Tschacher[7] · Zeno Kupper[7] · Oleg N. Medvedev[1]

Abstract

Objectives Although mindfulness has been studied for multiple decades, psychometric research has yet to agree upon the optimal way to measure the mindfulness construct. Prior research has identified eight distinct aspects of mindfulness that were not adequately captured by any of the available measures. Hence, the Comprehensive Inventory of Mindfulness Experiences (CHIME) was developed. The CHIME contains 37 items and was originally developed in the German language. The CHIME has demonstrated excellent psychometric properties in both German and Dutch, but so far, no English version has been validated. The purpose of the present study was to investigate the psychometric characteristics of the translated English-language CHIME scale using Rasch methodology.

Method The current study utilized Partial Credit Rasch analysis to evaluate the psychometric characteristics of the English CHIME. The sample included responses from 620 participants from the general population residing in the USA. The validity of the English CHIME was examined by correlating its scores with various measures of mindfulness and psychological functioning.

Results Initial Rasch analysis of the English CHIME showed poor model fit, local dependency, and evidence against the assumption of unidimensionality. Several minor modifications, that involved creating super-items, were required to fit the Rasch model (χ^2(45)=31.99, p=0.93). This model displayed evidence of unidimensionality, invariance across personal factors, and a high reliability (PSI=0.92). Ordinal-interval transformation tables were produced, which increase the English CHIME's precision of measurement. The English CHIME's external validity was established by moderate–high correlations with other measures of mindfulness and various measures of psychological functioning.

Conclusions The results of this study provide evidence for the validity of the English CHIME scale, which can be used to assess the overarching construct of mindfulness.

Keywords Mindfulness · Rasch analysis · Comprehensive Inventory of Mindfulness Experiences · Validation · Assessment · Psychometrics

Mindfulness can be defined as "the awareness that emerges through paying attention on purpose, in the present moment, and nonjudgmentally to the unfolding of experience moment by moment" (Kabat-Zinn, 2003, p.10). Interventions based on mindfulness have been shown to reduce both the physical and psychological symptoms of various psychopathologies, such as depression, anxiety, stress, borderline personality disorder, substance abuse, suicidal/self-harm behavior, and post-traumatic stress disorder (Krägeloh et al., 2019).

Additionally, mindfulness-based interventions (MBIs) have been shown to increase the well-being of participants (Bennett & Dorjee, 2016) as well as improve their emotional regulation (de Vibe et al., 2018). As mindfulness training has many positive effects, it is crucial for researchers and clinicians to be able to accurately measure the overarching mindfulness construct and discriminate precisely between individual mindfulness levels.

At present, the most commonly used multifaceted measure of mindfulness is the Five Facet Mindfulness Questionnaire (FFMQ; Baer et al., 2006). The FFMQ was constructed by applying factor analysis to the Kentucky Inventory of Mindfulness Skills (KIMS; Baer et al., 2004), the Mindful

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

Extended author information available on the last page of the article

Attention and Awareness Scale (MAAS; Brown & Ryan, 2003), and other available mindfulness measures. The FFMQ includes five subscales: Describing, Observing, Non-judging, Non-reacting to inner experience, and Acting with awareness. However, despite the well-established efficacy of the FFMQ, an analysis of mindfulness measures conducted by Bergomi et al. (2013) revealed that there were aspects of mindfulness that were not adequately captured by the FFMQ or any other available mindfulness assessment. Hence, to address this issue and create a complete measure of mindfulness, the Comprehensive Inventory of Mindfulness Experiences (CHIME) was developed (Bergomi et al., 2014). The CHIME includes all the components of mindfulness highlighted by Bergomi et al. (2013) and is rooted within the relevant theoretical frameworks (Krägeloh et al., 2019). The CHIME comprised eight subscales, which measure awareness of internal experiences, awareness of external experiences, acting with awareness, accepting non-judgemental attitude, nonreactive decentring, openness to experience, awareness of thoughts' relativity, and insightful understanding.

Since its development, the CHIME has been shown to have excellent psychometric properties. Specifically, the CHIME has been validated using participants from a Mindfulness-Based Stress Reduction (MBSR) intervention in addition to participants from a community sample (Bergomi et al., 2014). The CHIME demonstrated high internal consistency (α ranging from 0.70 to 0.90) in addition to high reliability (test re-test reliability, with r ranging from 0.70 to 0.90). Confirmatory factor analysis (CFA) was run on a different sample, which confirmed that the 8-factor structure was appropriate. Additionally, each CHIME item demonstrated invariance of measurement across the personal factors included in the study, indicating that there was no reliable difference in the participant's ability to understand the CHIME items. The external validity of the scale was established by a high positive correlation with the FFMQ ($r=0.85$), as well as moderate negative correlations with depression (-0.46), anxiety (-0.39), and stress (-0.40) (Bergomi et al., 2014).

The CHIME scale has two main advantages over other existing mindfulness measures. Firstly, it was developed with a strong theoretical grounding in traditional conceptualizations of mindfulness (Bergomi et al., 2014; Krägeloh et al., 2019). This contrasts with the FFMQ, which was constructed by applying factor analysis to other available mindfulness measures (Baer et al., 2006). Subsequently, most of the FFMQ's items are derived from the MAAS and the KIMS; hence, the FFMQ inherits the methodological flaws inherent in these scales. Specifically, the MAAS has been heavily criticized in mindfulness literature as not being a measure of mindfulness, but rather a measure of mindlessness/inattention (Bergomi et al., 2013). Additionally, the

KIMS was developed using a conceptualization of mindfulness as it is outlined in Dialectical Behavior Therapy (DBT; Linehan, 1993), which is an intervention primarily used to treat symptoms of borderline personality disorder. To address these issues, the CHIME was developed using a conceptualization of mindfulness as it is outlined in Eastern spiritual traditions. Secondly, the CHIME incorporates a wider range of characteristics, such as awareness of internal and external experiences, acting with awareness, accepting and non-judgmental attitude, nonreactive decentring, openness to experience, awareness of thoughts' relativity, and insightful understanding, that define and operationalize mindfulness more comprehensively than previous measures (Bergomi et al., 2014). Most of the prior mindfulness measures incorporate the mindfulness factors of awareness attention and non-judgmental attitude. However, in traditional conceptualizations, mindfulness is described as a path that develops an understanding of reality which can be used to alleviate suffering and enhance well-being (Harvey, 2013). Hence, it has been argued that a complete measure of mindfulness must incorporate these wisdom factors in addition to the non-judgment and attitude factors. These wisdom factors are incorporated into the CHIME as insightful understanding and awareness of thoughts' relativity.

The scale was originally developed in the German language and has been validated using both traditional methods and Rasch analysis (Bergomi et al., 2014; Medvedev et al., 2019). These analyses confirmed that the scale has excellent psychometric properties. Additionally, a Dutch translation of the CHIME, as well as a short version, has been developed and validated using classical test theory methods (Cladder-Micus et al., 2019). However, at present, there is only an English-language CHIME version adapted for children and adolescents (Johnson et al., 2016), whereas no English version for adults has been developed and validated to date. Even though the facets of the CHIME are based on relevant theoretical frameworks and translated using appropriate methodology, it cannot be assumed that facets would have the same psychometric properties as the validated German version. Such assumption is weak due to being subject to sampling error and variations between populations. Hence, further investigation is needed to evaluate the psychometric characteristics of the CHIME after it has been translated into English.

The narrative of developing an ordinal measure of mindfulness with the same properties as height or temperature is highly desirable while debatable in the current literature. First, height, temperature, and related measures are considered ratio-level scales in the Stevens classic system for variables (due to a known intercept for the scale representing the absence of the variable). This is rarely the case for psychological variables, including mindfulness. Second, considering psychometric variables like mindfulness as scales

(interval or ratio) is highly debatable, with relevant contributions arguing that these variables are ordinal variables (Beaujean et al., 2018). Therefore, Rasch analysis is especially suitable for evaluating and improving ordinal scales properties due to several key advantages over classical test theory methods (Rasch, 1961; Tennant & Conaghan, 2007).

The Rasch model accounts for both the underlying ability of the person and the difficulty of each item (Bond & Fox, 2007), allowing interval scores to be derived from ordinal questionnaires. This ordinal-interval conversion increases the precision of measurement and allows for parametric statistics to be conducted (Brogden, 1977; Rasch, 1961). Sandham et al. (2019) used the process of squeezing vitamin C from different fruits as a metaphor to explain how the Rasch model works. In this analogy, different fruits, such as apples, blueberries, and bananas, represent different items on a scale. Each fruit (item) contains differing levels of vitamin C (the latent trait being measured) and therefore contribute with different amounts of vitamin C to the overall (trait) smoothie being produced. In the same way, different items of a scale contribute different amounts of the latent trait (which in the present study is mindfulness ability) to the overall score. Rasch analysis allows for the measurement of a latent construct (vitamin C) to the extent that it is contained within each item (fruit) while also filtering out other constructs (other vitamins and minerals), manifesting as fit residuals in the model. The increased precision of measurement is demonstratable by comparing the accuracy of the original ordinal-level scores with their Rasch-converted interval equivalents (Norquist et al., 2004).

In fact, the Rasch model (Rasch, 1960) was proposed 8 years before the first introduction of the item response theory (IRT) in 1968, and uses the same computational algorithms as the one-parameter IRT model (Lord & Novick, 1968; Rasch, 1960). This basic Rasch/IRT model accounts for both a person's ability and item difficulty and complies with the fundamental measurement principles outlined by Thurstone (1931). These principles require equal units of measurement across the continuum of a scale, scale invariance by personal factors, and evidence of unidimensionality in measuring the overarching latent trait as operationalized by scale items and facets. The key difference between multidimensional IRT models (e.g., 2 or more parameter models) and Rasch models is that the latter strictly adhere to the principles of fundamental measurement, whereas two or more parameter IRT models use additional parameters that potentially inflate model fit by adjusting the model for the sample data (Hobart & Cano, 2009). For this reason, Rasch model is more suitable for transforming ordinal responses into interval-level scale.

While latent variable techniques such as CFA can also be used to estimate the ability of persons and difficulty of items using intercepts and thresholds, these techniques do not provide accurate conversion of ordinal scores to interval-level data, which is a unique advantage of one-parameter IRT and Rasch models (Hobart & Cano, 2009). Another benefit of Rasch modelling is reduced measurement error, where participant parameter estimations usually have a lower standard deviation. The lower standard deviation also allows for more precise assessment of change in individual scores, making the measurement more accurate and reliable.

As opposed to item parcelling used in traditional methods such as CFA and exploratory factor analysis (EFA), the super-item models (or testlets) used in Rasch analysis are generally based on residual correlations and aim to resolve the local dependency caused by spurious correlations or method effects (Oyler et al., 2022). Items are considered to be locally dependent if they have residual correlations above 0.20, which is identified by examining a residual correlation matrix. Local dependency affects the overall fit of the model and may lead to spurious correlations between items, which interferes with unidimensionality (Christensen et al., 2016; Wainer & Kiely, 1987). When items share variance that is not directly related to the central construct that is being measured, they are usually locally dependent, for example, facets of the FFMQ (Medvedev et al., 2017). Hence, using super-items helps to solve the problem of local dependency between the items and achieve an acceptable fit of the Rasch model, which increases reliability. It should be noted that multidimensional scales cannot achieve an acceptable fit to the unidimensional Rasch model, and meeting strict criteria of the Rasch model is taken as an evidence for measuring one overarching latent trait regardless how the items/super-items are combined within the model (Balalla et al., 2019; Mitchell-Parker et al., 2017). In more traditional techniques such as CFA, the psychometric properties of analyzed scales remain unmodified, whereas Rasch analysis can transform the ordinal data into true interval-level scores, similar to those used to measure physical phenomena such as temperature, weight, or height. In Rasch analysis, all of the super-item modifications are included and accounted for in the conversion algorithms (Medvedev et al., 2019).

As is evident from the current literature, the application of Rasch analysis is providing a growing contribution to the field of mindfulness assessment (Krägeloh et al., 2019; Medvedev & Krägeloh, 2022). For example, Rasch analysis is increasingly being used to explore the psychometric characteristics of mindfulness measures (Goh et al., 2017; Medvedev et al., 2016; Sauer et al., 2013). In 2013, Sauer et al. utilized Rasch analysis, analyzing the properties of the Freiburg Mindfulness Inventory (FMI-14, Walach et al., 2006). After discarding one item, they achieved a Rasch model fit. However, a better fit was obtained after the two dimensions of "presence" and "acceptance" had the Rasch model applied to them individually. On a similar note, both Medvedev, Siegert, Feng, et al. (2016) and Goh

et al. (2017) applied Rasch analysis to assess the characteristics of the MAAS (Brown & Ryan, 2003). Interestingly, their approaches to fit the Rasch model were quite different. Goh et al. (2017) removed 5 items that were misfitting, whereas Medvedev, Siegert, Feng, et al. (2016) only discarded 2. Medvedev, Siegert, Feng, et al. (2016) were able to reduce the number of items they discarded by using super-items to resolve issues of local dependency. In contrast, Medvedev et al. (2016) conducted Rasch analysis on the KIMS (Baer et al., 2004) but were unable to find a fit for the overall scale, even after they had discarded 5 items that were misfitting. Medvedev, Siegert, Feng, et al. (2016), Medvedev, Siegert, Kersten et al. (2016) ended up resolving this issue and achieved an adequate fit by applying the Rasch model to each of the KIMS subscales individually. Rasch analysis was also used to evaluate the psychometrics of the FFMQ (Medvedev et al., 2017). To fit the Rasch model, two misfitting items of the FFMQ were discarded. Additionally, each of the individual FFMQ subscales was summed to form super-items. After these modifications, the best model fit was obtained. Finally, although it is not solely a measure of mindfulness, the Self-Compassion Scale (SCS; Neff, 2003) was also scrutinized using Rasch methodology (Finaulahi et al., 2021). Although no SCS items were significantly misfitting the Rasch model, the individual items were found to be locally dependent, which affected the overall model fit. Once again, this problem was solved by summing the SCS items to form four higher-order super-items.

By utilizing modern Rasch methodology, the present study aimed to evaluate the psychometric characteristics of the CHIME, after it was translated into English. A product of this analysis is to develop ordinal-interval conversion algorithms, increasing the CHIME's accuracy of measurement, which will aid in future mindfulness research. To establish construct validity of the measure, the present study correlated the CHIME scores with other measures that are expected to be related to the construct of mindfulness. If the CHIME successfully measures the construct of mindfulness, then it should be positively correlated with other validated measures of mindfulness (i.e., the FFMQ). As mindfulness has been shown to promote well-being (Bennett & Dorjee, 2016), mindfulness measures should be positively correlated with measures of well-being (Satisfaction with Life Scale and Positive Affect) and negatively correlated with measures of ill-being (Negative Affect, Depression Anxiety Stress Scales). Informed by previous studies utilizing Rasch analysis, we hypothesized that, if the individual items and subscales of the English CHIME adequately represent the underlying mindfulness construct, then a Rasch model fit will be obtained if the subscales are treated as super-items. Additionally, positive relationships are expected between the CHIME and the FFMQ, as well as measures of well-being. Negative relationships are expected to be found between the CHIME and measures of ill-being.

Method

Participants

Out of the 760 English-speaking participants from the general population residing in the USA who responded to the survey, 140 did not complete all the study measures and were therefore not included in the subsequent analyses. The 620 participants sample selected for the study was identified to minimize the probability of type-1 and type-2 errors. Typically, overfitting occurs with larger sample sizes (e.g., $n=1000$) while a lower sample size ($n<250$) may be insufficient to calibrate items (Hagell & Westergren, 2016). The sample comprised 301 (48.5%) males, and 238 (38.4%) participants indicated that they meditated or engaged in some form of contemplative practice (such as Yoga or Tai-Chi). In this sample, 503 participants (81.1%) were White American, 69 (11.1%) were African American, 26 (4.2%) were Asian American, and 22 (3.6%) identified as other. All participants were proficient in English and were living in the USA at the time of data collection. The ages of the participants spanned between 18 and 70 years old ($M=41.65$, $SD=13.07$). To determine whether there was any differential item functioning (DIF), participant data was sorted into different age categories. The categories utilized were 18–34, 35–49, and 50–70.

Procedure

Data were collected through an online survey using Qualtrics Research Services in March 2018 and targeted participants of both sexes (male/female split of 50/50). The average survey completion time was estimated by the work of five volunteers who filled out the survey and reported the length of their work. An average completion time of 15 min was used as a parameter for data collection. The online data collection lasted up to 10 days, and each participant was compensated with US$5 after completing the survey. Each participant had their IP address recorded, preventing the same participant from filling out the survey multiple times. The IP address collection confirmed that the survey was distributed throughout the country, guaranteeing good sample representation. This research was approved by the author's institutional ethics review board. All participants who were involved in the study provided informed consent to participate in the research.

Measures

The Comprehensive Inventory of Mindfulness Experiences (CHIME) is a self-report questionnaire containing 37 items. It includes 8 subscales that measure different aspects of mindfulness: awareness of internal experiences, awareness of external experiences, acting with awareness, accepting non-judgmental attitude, nonreactive decentring, openness to experience, awareness of thoughts' relativity, and insightful understanding (Bergomi et al., 2014). The measure utilizes a 6-point Likert scale format, ranging from *almost never*=1 to *almost always*=6, and negatively worded items (7, 10, 17, 19, 22, 26, 30, 33, and 36) require reverse coding before subscale scores can be calculated. The total score is computed by summing the answers from each individual CHIME item. Higher scores on the CHIME correspond to higher levels of the underlying mindfulness construct. A routine check of reliability was conducted before the Rasch analysis. The overall scale had a strong internal consistency reflected by both Cronbach's alpha and McDonald's Omega of 0.95, suggesting that it is a good candidate for a unidimensional Rasch model.

The translation of the original German CHIME scale into English was completed in multiple stages, adhering to the recommendations outline by Hambleton, (2005) and the guidelines laid out by the International Test Commission (2017). The original CHIME authors (German native speakers who also spoke English fluently) translated the CHIME items into English and then the translated items were reviewed by two English speakers who were familiar with meditation. The translated items were then given to a professional translator, who translated the items back into German. The back-translated items were compared with the original CHIME items to assess whether the translation was successful. Considering the validity of cultural and language differences between the two populations, the CHIME items were further adapted before running the validation study. Twenty graduate psychology students originally from the USA (8 males, 12 females) rated the CHIME items according to their readability and understandability. Participants were also asked to rewrite the items according to their understanding. This small sample survey was then subjected to a qualitative analysis in which a group of six mindfulness researchers and practitioners analyzed item by item, adapting them as needed, according to the participants' feedback.

The FFMQ (Baer et al., 2006) contains 39 items, measuring 5 separate facets: Describing, Observing, Non-judging, Non-reacting to inner experience, and Acting with awareness. Answers are scored on a 5-point Likert scale, from *never*=1 to *always true*=5. The measure contains 19 items that are negatively worded (3, 5, 8, 10, 12, 13, 14, 16, 17, 18, 22, 23, 25, 28, 30, 34, 35, 38, and 39), which requires

reverse coding before the total and subscale scores can be calculated.

The SCS is a measure which contains 26 items and assesses a person's underlying amount of self-compassion (Neff, 2003). The measure includes 6 subscales: Self-Kindness, Self-Judgement, Common Humanity, Isolation, Mindfulness, and Over-Identification. Questions are answered on a 5-point Likert scale, from *almost never*=1 to *almost always*=5. Half of the items are negatively worded (1, 2, 4, 6, 8, 11, 13, 16, 18, 20, 21, 24, and 25) and require reverse coding before total and subscale scores can be calculated. There are multiple versions of the SCS and the abbreviated version containing 12 items, the Self-Compassion Scale Short Form (SCS-SF; Raes et al., 2011), was used in this study.

The Satisfaction with Life Scale (SWLS) is a brief inventory including five items to measure a person's self-reported life satisfaction (Diener et al., 1985). Items are scored on a 7-point Likert scale, from *strongly disagree*=1 to *strongly agree*=7.

The Depression Anxiety Stress Scales (DASS) is a self-report measure that assesses the negative emotion facets of depression, anxiety, and stress (Lovibond & Lovibond, 1995). The measure includes 42 items, 14 for each subscale. Scoring of the items takes place on a 4-point Likert scale, ranging from *Did not apply to me at all*=0 to *Applied to me very much or most of the time*=3. In addition to the 42-item questionnaire, a shorter version of the scale (the DASS-21) containing 21 items is also available (Antony et al., 1998). Both versions of the scale have been shown to have high internal consistencies and robust psychometric properties. The DASS-21 was utilized in the present study.

The Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) is a short list of adjectives that describing different emotions and feelings. Participants are instructed to indicate the degree to which they have felt the emotions/feelings over the past week, and answers are recorded on a 5-point Likert scale. Scores range from *Very slightly or not at all*=1 to *Extremely*=5. Once the survey is complete, answers from all the positive emotion adjectives are added to create the Positive Affect scale, and answers from all the negative emotion adjectives are added to create the Negative Affect Scale.

Data Analyses

Prior to Rasch analyses, we have examined psychometric properties and reliability of the CHIME facets using IBM SPSS v27. Rasch analysis was conducted using RUMM2030 software (Andrich et al., 2009). The Rasch analysis was performed following procedure outlined elsewhere (Medvedev et al., 2019). A likelihood-ratio test was computed on the initial analysis output for the CHIME scale prior to the main

analysis. The likelihood-ratio test supported the suitability of the unrestricted partial credit version of the Rasch model ($p<0.001$) (Masters, 1982). Rasch analysis was conducted for the full scale, where all the subscales were treated as super-items, following the methodology of Lundgren-Nilsson et al. (2013).

Unlike classical test theory methods, the model fit in Rasch analysis indicates that the scale complies with the fundamental principles of measurement, such as invariance across personal factors, unidimensionality, and the same measurement units across scale continuum (Oyler et al., 2022). Therefore, obtaining the best model fit in Rasch analysis is desirable as this will minimize potential deviations from the interval scale parameters (Hobart & Cano, 2009). The ideal Rasch model will have a mean close to 0.00 and a standard deviation close to 1.00 for the overall and person fit residuals (see Balalla et al., 2019, for an overview of these criteria). Individual item fit residuals are expected to range from −2.50 to +2.50. Trait-item interaction, reflected by an overall and individual Chi-square fit statistic, should be non-significant ($p<0.05$, Bonferroni adjusted). The residuals correlation matrix should display no evidence of local dependency between individual items. A correlation magnitude of 0.20 relative to the mean residual correlation indicates local dependency (Christensen et al., 2016), an issue which can be resolved by combining the dependent items into super-items (Lundgren-Nilsson et al., 2013; Wainer & Kiely, 1987). In the Rasch model, personal factors should not produce any significant DIF. The person separation index (PSI) reflects how precisely subjects have been spread out along the measurement construct defined by the items and is used to test the reliability of subscales in Rasch analysis. PSI is interpreted in a similar way to Cronbach's alpha (Tennant & Conaghan, 2007).

In Rasch analysis, dimensionality is typically investigated by using independent-samples t-tests to compare person estimates for two item groups. In this method, the highest positive and negative factor loadings are on the first principal component of the residuals after the latent factor has been removed (Smith, 2002). If the percentage of significant t-test comparisons does not exceed 5%, or if the 5% cutoff point is overlapped by the lower bound of a binomial confidence interval (computed for the number of significant t-tests),

then the scale is considered unidimensional (Tennant & Pallant, 2006).

Results

To ensure that the English CHIME version is suitable for application of the unidimensional Rasch model, we have conducted the overall unidimensionality assessment using parallel analyses (Timmerman & Lorenzo-Seva, 2011), which supported the overarching mindfulness construct operationalized by English CHIME items. Mean of item residual absolute loadings (MIREAL = 0.28) was below 0.30, while unidimensional congruence (0.87) and explained common variance (0.77) were approaching benchmarks set up for strict unidimensionality. Similarly, all facets of the CHIME demonstrated unidimensionality as evidenced by McDonald's omega (ω=0.70–0.91) and high loadings on each single facet factor (0.42–0.86) with exception of Accepting nonjudgmental attitude (ω=0.56) and Item 12 from the Acting with awareness facets (<0.40) that were retained due to their importance for the overall construct validity of the scale.

Table 1 shows the fit statistics of the Rasch model from the main analysis. The English CHIME scale displayed poor model fit in the initial analysis, with a significant Chi-square indicating that there was a deviation from the expectations of the Rasch model ($\chi^2(185)$=1111.39, $p<0.001$). Although the reliability of the scale was strong (PSI=0.95), there was evidence against the assumption of unidimensionality (Table 1). Table 2 shows the fit statistics for each individual item. Notably, several items were misfitting, with fit residuals outside the acceptable range of −2.50 to +2.50. The misfitting items included Items 1 ("When my mood changes, I notice it right away"), 10 ("I break or spill things because I am not paying attention or am thinking of something else"), 11 ("I see my mistakes and difficulties without judging myself"), 17 ("In everyday life, I get distracted by many memories, images, or daydreams"), 19 ("I try to stay busy to avoid specific thoughts or feelings from coming to mind"), 26 ("When I read, I have to reread paragraphs because I was thinking of something else"), and 36 ("I resent my own mistakes and weaknesses").

Table 1 The overall Rasch model fit statistics for the English CHIME version, including initial and final analyses (n=620)

Analyses	Item fit residual		Person fit residual		Goodness of fit		PSI	Unidimensionality Independent t-test	
	Value	SD	Value	SD	χ^2 (df)	P		%	Lower bound
Initial	0.93	3.41	−1.23	3.78	1111.39 (185)	<0.01	0.95	25.97	24.25 (NO)
8 super-items	0.02	3.00	−0.68	1.57	170.25 (72)	<0.01	0.90	0.81	−0.91 (YES)
Final	0.03	1.76	−0.61	1.30	31.99 (45)	0.93	0.92	6.61	4.90 (YES)

Table 2 Initial Rasch model fit statistics for individual items including item location, fit residual, and Chi-square (*n*=620)

No	Item Content	Location	Fit Resid	Chi Sq
1	When my mood changes, I notice it right away.	−0.37	3.37*	14.28
2	During both ups and downs of life, I am kind to myself.	0.02	8.17*	10.98
3	In everyday life I notice when my negative attitudes toward a situation make things worse.	−0.22	0.11	8.87
4	It is clear to me that my evaluation of situations and people can easily change.	−0.32	−0.21	10.39
5	When I am sitting or lying down, I notice the sensations in my body.	0.18	1.69	2.38
6	I am able to smile when I notice myself seeing things as more complicated than they actually are.	0.20	−0.85	12.48
7	I am hard on myself when I make a mistake.[R]	0.04	7.60*	121.91
8	When I have distressing thoughts or images, I am able to feel calm soon afterward.	0.24	−1.11	14.93
9	I notice the details in nature, such as colors, shapes, and textures.	−0.39	−0.91	15.16
10	I break or spill things because I am not paying attention or I am thinking of something else.[R]	0.65	7.85*	133.50
11	I see my mistakes and difficulties without judging myself.	0.24	4.09*	7.03
12	It is easy for me to stay focused on what I am doing.	0.02	2.07	13.47
13	When I have distressing thoughts or images, I am able to notice them without having to react.	0.11	−1.23	13.34
14	When I talk to other people, I notice what feelings I am experiencing.	−0.28	−1.43	32.81
15	When I have been needlessly hard on myself, I can see it with some humor.	0.22	−1.28	11.17
16	In difficult or triggering situations, I can pause for a moment without reacting immediately.	0.12	−1.30	17.60
17	In everyday life, I get distracted by many memories, images, or daydreams.[R]	0.19	5.18*	63.65
18	When I ride in a car, bus, or train, I am aware of the surroundings, such as the landscape.	−0.35	−0.70	10.21
19	I try to stay busy to avoid specific thoughts or feelings from coming to mind.[R]	0.18	4.33*	41.16
20	When caught in thoughts and emotions, I am able to "step back" and quickly notice the thought or image without being taken over by it.	0.16	−2.04	24.97
21	I pay attention to sensations, such as the wind in my hair or sunshine on my face.	−0.30	−2.26	35.01
22	I try to distract myself when I feel unpleasant emotions.[R]	0.09	1.95	10.82
23	In everyday life, I realize my thoughts are not always facts.	−0.12	−1.33	10.68
24	I am able to smile to myself when I notice I have made a big deal out of a small problem.	0.13	−1.85	15.59
25	I am able to notice my thoughts and feelings without getting tangled up in them.	0.06	−2.50	24.62
26	When I read, I have to reread paragraphs because I was thinking of something else.[R]	0.20	5.24*	75.77
27	I notice sounds in my environment, such as birds chirping or cars passing.	−0.55	−1.85	28.16
28	I notice my thoughts and feelings and can also "step back" and observe them from a distance.	0.16	−1.35	20.07
29	I clearly notice changes in my body, such as quicker or slower breathing.	0.05	−1.63	20.83
30	I do not like it when I am angry or fearful, and try to get rid of these feelings.[R]	−0.06	2.52	10.73
31	In everyday life, I am aware that my view on things is not always based on facts.	0.02	−0.35	3.12
32	Even when I make a big mistake, I treat myself with kindness and understanding.	0.19	−0.79	10.39
33	When I am in pain, I try to avoid the sensations as much as possible.[R]	−0.03	0.23	2.08
34	I am aware of how I am feeling at any given time.	−0.29	−1.06	9.56
35	I am aware that even my strongly held opinions may change over time.	−0.32	−2.11	20.26
36	I resent my own mistakes and weaknesses.[R]	0.19	9.90*	209.39
37	I am able to notice when I needlessly make life more difficult for myself.	−0.05	−1.89	24.04

[R]Reverse-scored item

*Significant misfit (*p*<0.05)

The residual correlation matrix indicated that there were high residual correlations (above 0.20) between items representing the 8 mindfulness facets. Hence, the individual English CHIME items were combined to form 8 super-items, which were then treated as individual items in the Rasch analysis, a method that is now becoming common practice (Lundgren-Nilsson et al., 2013; Medvedev et al., 2018). After creation of the 8 super-items, the fit improved considerably, the reliability remained high (PSI=0.90), and strict unidimensionality was confirmed (Table 1, 8 super-items). However, the Chi-square remained significant (χ^2(72)=170.25, *p*<0.001), indicating deviation from the Rasch model.

Examination of the correlation matrix between the 8 super-items revealed local dependency between 6 out of the 8 super-items (above 0.20). To address this issue, the

locally dependent super-items were reconfigured to form 3 new super-items as follows: 2 (*awareness of external experiences*) was combined with 4 (*accepting non-judgmental attitude*), 3 (*acting with awareness*) was combined with 8 (*insightful understanding*), and 5 (*nonreactive decentring*) was combined with 6 (*openness to experience*). After creating these higher-order super-items, the excellent fit was obtained ($\chi^2(45)=31.99$, $p=0.928$). This model displayed high reliability (PSI=0.92), and unidimensionality was confirmed (Table 1, Final). DIF was also examined for the best-fitting model, and the scale was found invariant across personal factors, including meditation practice, gender, and age. All the individual super-items displayed excellent fit to the overall Rasch model (Table 3).

The distribution of person-item thresholds of the English CHIME is illustrated in Fig. 1. The plot demonstrates excellent targeting of the sample by item thresholds ($M=0.31$, $SD=0.50$), with person mean slightly above item mean. There were no detectable ceiling or floor effects, with person abilities perfectly covered by the scale items. Overall, Fig. 1 indicates that there is a good combination of both easy and difficult items in the English CHIME scale.

Table 4 shows the conversion algorithms that allow for ordinal scores to be converted into interval scores for English CHIME scale. The conversion algorithms are based on the final analysis using 5 super-items. In the conversion table, the ordinal scores are presented in the first column, logit unit interval scores are presented in the second column, and in the third column, the interval scores are presented in the original scale metric. To use the conversion table, (1) reverse code the negatively worded items, (2) compute the total score by summing the individual scores of items, and (3) find the equivalent interval level scores in logits in the second column and the original scale metric in the third column. Note that these conversions are not able to be performed for participants who have responses that are missing.

To compare the original ordinal English CHIME scores with the Rasch-transformed interval scores, a paired-samples *t*-test was run. The data used for *t*-tests comparisons met the common assumptions of normality with skewness and kurtosis values within the acceptable range from −2.00 to +2.00. The *t*-test revealed that the difference between ordinal scores ($M=145.95$, $SD=29.16$) and the interval scores ($M=136.26$, $SD=19.59$) was significant, $t(619)=-18.93$, $p<0.001$, $d=-0.76$. Note that the standard deviation of the interval scores was noticeably smaller than the ordinal scores, providing additional evidence that the measurement error was significantly reduced by using super-items. Independent-samples *t*-tests were then conducted comparing meditators and non-meditators scores, using the original ordinal scores and scores that had undergone the Rasch interval transformation. Meditators' ordinal scores ($M=153.66$, $SD=29.62$) were found to be significantly higher compared to non-meditators' ordinal scores ($M=146.18$, $SD=29.62$),

Table 3 Individual super-items Rasch fit statistics for the final analysis ($n=620$)

Final super-items	Item Location	Fit Residual	Chi-Square
Super-item 7 (awareness of thoughts' relativity)	−0.06	1.22	1.69
Super-item 2 (awareness of external experiences) + 4 (accepting nonjudgmental attitude)	−0.04	−1.77	2.75
Super-item 3 (acting with awareness) +8 (insightful understanding)	0.12	−0.11	3.72
Super-item 5 (nonreactive decentering) + 6 (openness to experience)	0.15	−1.70	1.02
Super-item 1 (awareness of internal experiences)	−0.16	2.20	4.70

Fig. 1 Person-item threshold distribution of the English CHIME version (final analysis). It indicates that mindfulness abilities of the sample (above) are well covered by the range of CHIME items' difficulties (below) with sample abilities resemble a normal distribution
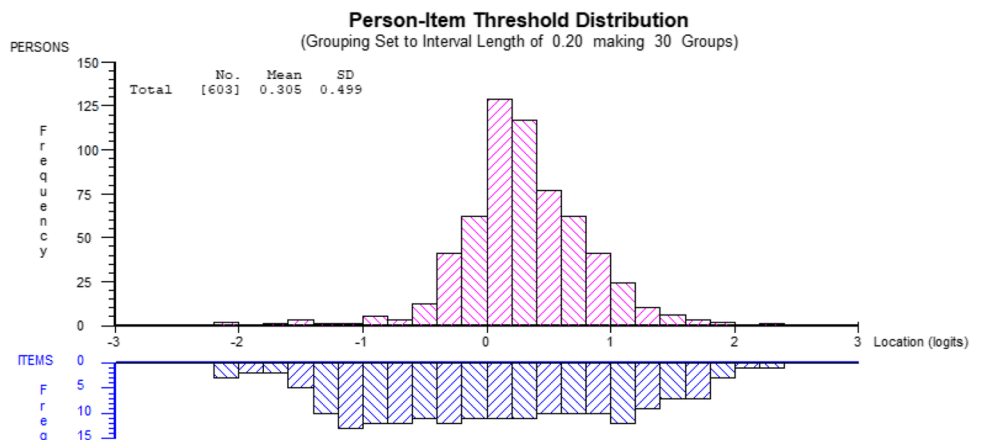
Table 4 Ordinal-to-interval conversion table for the English CHIME scale

Ordinal scores	Interval Logits	Interval Scale	Ordinal scores	Interval Logits	Interval Scale	Ordinal scores	Interval Logits	Interval Scale	Ordinal scores	Interval Logits	Interval Scale	Ordinal scores	Interval Logits	Interval Scale	Ordinal scores	Interval Logits	Interval Scale
37	−3.62	37.00	68	−1.08	101.09	99	−0.55	114.50	130	−0.01	128.16	161	0.58	142.96	192	1.23	159.32
38	−3.01	52.43	69	−1.06	101.60	100	−0.53	114.93	131	0.01	128.62	162	0.60	143.46	193	1.25	159.92
39	−2.64	61.83	70	−1.04	102.08	101	−0.51	115.36	132	0.03	129.07	163	0.62	143.97	194	1.28	160.50
40	−2.41	67.57	71	−1.02	102.56	102	−0.50	115.79	133	0.05	129.53	164	0.64	144.47	195	1.30	161.11
41	−2.25	71.63	72	−1.00	103.04	103	−0.48	116.22	134	0.07	130.00	165	0.66	144.98	196	1.33	161.71
42	−2.13	74.76	73	−0.99	103.49	104	−0.46	116.65	135	0.09	130.46	166	0.68	145.48	197	1.35	162.34
43	−2.02	77.33	74	−0.97	103.97	105	−0.45	117.07	136	0.10	130.91	167	0.70	145.98	198	1.38	162.97
44	−1.94	79.49	75	−0.95	104.42	106	−0.43	117.50	137	0.12	131.39	168	0.72	146.49	199	1.40	163.60
45	−1.86	81.38	76	−0.93	104.85	107	−0.41	117.96	138	0.14	131.84	169	0.74	146.99	200	1.43	164.28
46	−1.80	83.05	77	−0.91	105.30	108	−0.39	118.38	139	0.16	132.32	170	0.76	147.50	201	1.45	164.94
47	−1.74	84.54	78	−0.90	105.73	109	−0.38	118.81	140	0.18	132.80	171	0.78	148.00	202	1.48	165.64
48	−1.68	85.87	79	−0.88	106.16	110	−0.36	119.24	141	0.20	133.26	172	0.80	148.53	203	1.51	166.37
49	−1.63	87.13	80	−0.86	106.59	111	−0.34	119.70	142	0.22	133.73	173	0.82	149.03	204	1.54	167.13
50	−1.59	88.27	81	−0.85	106.99	112	−0.33	120.12	143	0.23	134.21	174	0.84	149.56	205	1.57	167.91
51	−1.55	89.32	82	−0.83	107.42	113	−0.31	120.55	144	0.25	134.69	175	0.86	150.07	206	1.60	168.74
52	−1.51	90.31	83	−0.81	107.85	114	−0.29	121.01	145	0.27	135.17	176	0.88	150.60	207	1.64	169.60
53	−1.47	91.24	84	−0.80	108.25	115	−0.27	121.43	146	0.29	135.65	177	0.90	151.10	208	1.67	170.51
54	−1.44	92.12	85	−0.78	108.68	116	−0.26	121.89	147	0.31	136.13	178	0.93	151.63	209	1.71	171.49
55	−1.40	92.93	86	−0.76	109.08	117	−0.24	122.32	148	0.33	136.61	179	0.95	152.16	210	1.76	172.55
56	−1.37	93.71	87	−0.75	109.49	118	−0.22	122.77	149	0.35	137.09	180	0.97	152.69	211	1.80	173.71
57	−1.34	94.47	88	−0.73	109.92	119	−0.20	123.20	150	0.37	137.57	181	0.99	153.22	212	1.85	174.97
58	−1.31	95.20	89	−0.71	110.32	120	−0.19	123.65	151	0.39	138.04	182	1.01	153.77	213	1.91	176.35
59	−1.29	95.88	90	−0.70	110.75	121	−0.17	124.11	152	0.41	138.55	183	1.03	154.30	214	1.97	177.92
60	−1.26	96.53	91	−0.68	111.15	122	−0.15	124.53	153	0.43	139.03	184	1.05	154.83	215	2.04	179.73
61	−1.24	97.16	92	−0.66	111.58	123	−0.13	124.99	154	0.44	139.51	185	1.07	155.38	216	2.12	181.82
62	−1.21	97.77	93	−0.65	111.98	124	−0.11	125.44	155	0.46	140.01	186	1.10	155.94	217	2.22	184.32
63	−1.19	98.37	94	−0.63	112.41	125	−0.10	125.90	156	0.48	140.49	187	1.12	156.49	218	2.34	187.39
64	−1.17	98.95	95	−0.62	112.81	126	−0.08	126.35	157	0.50	140.99	188	1.14	157.05	219	2.50	191.43
65	−1.14	99.51	96	−0.60	113.24	127	−0.06	126.80	158	0.52	141.47	189	1.16	157.60	220	2.73	197.15
66	−1.12	100.04	97	−0.58	113.67	128	−0.04	127.26	159	0.54	141.98	190	1.19	158.18	221	3.10	206.55
67	−1.10	100.57	98	−0.56	114.10	129	−0.02	127.71	160	0.56	142.48	191	1.21	158.76	222	3.72	222.00

$t(314)=2.00$, $p=0.047$, $d=0.26$. Similarly, meditators' interval scores ($M=141.01$, $SD=19.23$) were significantly higher compared to non-meditators' interval scores ($M=136.08$, $SD=18.04$), $t(314)=2.00$, $p=0.047$, $d=0.26$. These results provided evidence that the English CHIME accurately differentiates between mindfulness levels, as people who practice meditation tend to score higher on the scale than those who do not practice meditation.

Table 5 shows the correlations between the English CHIME total interval scores and other relevant psychometric measures. All the additional measures included in Table 5, excluding the SWLS, were converted into interval scores using the available Rasch transformation algorithms. All scales data met the standard assumptions of normality for Pearson's correlation, with skewness and kurtosis values within the acceptable range from -2.00 to $+2.00$. All correlations were in the expected directions, except for the relationship between the English CHIME total score and DASS-Anxiety, which remained non-significant. The external validity of the English CHIME was demonstrated by high positive correlations with both the FFMQ and SCS, and a moderate positive correlation with positive affect. As expected, the English CHIME correlated negatively with stress, depression, and negative affect.

Discussion

Using modern Rasch methodology, the present research validated the English CHIME measure and developed ordinal-to-interval transformation tables to be used in future mindfulness research. The English CHIME instrument displayed evidence of internal structural validity, unidimensionality, high reliability, and external validity. Several items in the initial analysis were misfitting the Rasch model. This issue was addressed by summing the individual items to form super-items. The utilization of super-items effectively

reduced the measurement error due to spurious correlations between items and possible method effects (Finaulahi et al., 2021), resulting in the best fit. After the best fitting model was achieved, ordinal-to-interval transformation algorithms were developed, which may be used to measure mindfulness and the effects of MBIs more precisely in future research.

Rasch transformation is highly desirable in the case of multifaceted measures because it reduces the error of measurement, while also accounting for the different contributions of items and subscales to the overarching construct (i.e., mindfulness as measured by the English CHIME). A major limitation of ordinal measures is that they do not consider the difficulty (location) of items when calculating the total scores. Not considering item difficulty decreases the accuracy of assessment and increases the amount of measurement error (Bond & Fox, 2007; Norquist et al., 2004). As a result of the Rasch transformation, ordinal English CHIME scores may be converted into interval scores by using the transformation algorithms displayed in Table 4, allowing for parametric statistics to be conducted. This ordinal-to-interval transformation is achieved by accounting for unwanted measurement error unrelated to the overarching mindfulness construct, increasing the accuracy of the English CHIME, without modification to the original scale format. This conversion can be conducted for the full scale if there are no missing data.

The Rasch transformation algorithm allows for the measurement of mindfulness to the extent by which it is reflected by each individual item or super-item, while also filtering out the irrelevant influences associated with other constructs (e.g., personality) along with methodological errors. For example, Item 10 ("I break or spill things because I am not paying attention or I am thinking of something else") appears to express mindfulness to a larger extent, as reflected by a positive location of 0.65, than Item 27 ("I notice sounds in my environment, such as birds chirping or cars passing"), which has a negative location of -0.55. The same is true at

Table 5 Correlations between Rasch-transformed CHIME total scores and other psychological measures ($n=208$)

Scales	CHIME-R	FFMQ-R	SCS-R	SWLS	PA-R	NA-R	Depression-R	Anxiety-R
FFMQ-R	0.66**							
SCS-R	0.61**	0.70**						
SWLS	0.37**	0.33**	0.45**					
PA-R	0.56**	0.42**	0.45**	0.47**				
NA-R	−0.20**	−0.46**	−0.42**	−0.25**	−0.07			
Depression-R	−0.23**	−0.48**	−0.49**	−0.39**	−0.18**	0.72**		
Anxiety-R	−0.01	−0.33**	−0.29**	−0.14**	0.05	0.68**	0.77**	
Stress-R	−0.15**	−0.41**	−0.42**	−0.25**	−0.05	0.72**	0.82**	0.82**

R denotes scales that have undergone Rasch transformation; *CHIME*, Comprehensive Inventory of Mindfulness Experiences; *FFMQ*, Five-Facet Mindfulness Questionnaire; *SCS*, Self-Compassion Scale; *SWLS*, Satisfaction with Life Scale; *PA*, PANAS Positive Affect; *NA*, PANAS Negative Affect; *Depression*, DASS Depression; *Anxiety*, DASS Anxiety; *Stress*, DASS Stress

** $p< 0.01$; * $p< 0.05$

super-item level. Using the vitamin C example by Sandham et al. (2019), vitamin C, a super-item, could be compared to a smoothie made up of fruits which have differing levels of vitamin C. The super-item (smoothy) thresholds are estimated by using the combined responses of all the included items (fruits). This example is useful to illustrate that the super-items and their combinations used in the current study refine the measurement of the overarching mindfulness construct (as captured by the English CHIME items) and have no implications for either the factor structure of the English CHIME or the use of its individual subscales. A reader should be mindful that, by using individual subscales, they are only able to measure a specific facet that is relevant to mindfulness.

The initial exploration of the English CHIME scale revealed that the baseline model showed promising psychometric properties. However, the significant Chi-square showed a Rasch model misfit, with several items exhibiting issues of local dependency. Hence, the individual English CHIME items were summed to form 8 super-items, following the methodology previously utilized by Lundgren-Nilsson et al. (2013). A significant advantage of this approach is that the issue of local dependency can be dealt with without the deletion of any items. Deleting items can affect the construct validity of a scale and therefore should be avoided whenever possible. Even though the English CHIME retained its excellent properties after these modifications, the Chi-square remained significant. After some investigation, it was determined that six out of the 8 super-items showed issues of local dependency. Hence, the model was adjusted, and instead the English CHIME items were summed to form 5 super-items. After this adjustment, the Chi-square became non-significant, and the best Rasch model fit was obtained.

Compared with other mindfulness measures, the English CHIME required only marginal modification to fit the Rasch model. This is in contrast with the Rasch exploration of the KIMS (Medvedev, Siegert, Kersten, et al., 2016), where the fit was only achieved at subscale level after the deletion of five items. On a similar note, the FFMQ required the deletion of two items to fit the Rasch model, with the five facets summed to form super-items (Medvedev et al., 2017). The English CHIME, on the other hand, did not require the deletion of any items to obtain the best fit. This result is consistent with the Rasch analysis conducted on the original CHIME scale in German, which also did not require the deletion of any items (Medvedev et al., 2019). However, in the German scale, the adequate fit was obtained by forming 8 super-items, whereas in the present study, the adequate fit was achieved after forming 5 super-items. This discrepancy is likely an artifact of translating the scale from German to English. Although both the German and English scales displayed exceptional psychometric properties, the results of the present research indicate that the CHIME items may

present with a different pattern of local dependency when translated into different languages. However, in terms of its overall structure and scoring as a unidimensional profile, there is no difference between the German and English CHIME versions.

In following the original validation study (Medvedev et al., 2019), the English CHIME showed correlations in the expected directions with the other psychometric measures, apart from the correlation with DASS-Anxiety, which remained non-significant. Unsurprisingly, the English CHIME showed convergent validity with the FFMQ. Although the magnitude of the correlation was high (0.66), indicating that the scales reflect the overlapping construct, the English CHIME explains a unique amount of variance not covered by the FFMQ. The English CHIME was found to be positively correlated with negative affect, stress, and depression. These results are consistent with the substantial literature findings, which demonstrate higher mindfulness levels are linked with better psychological functioning (Fumero et al., 2020; Morton et al., 2020; Thomas et al., 2020). The non-significant correlation between the English CHIME total score and DASS-Anxiety is unexpected and may be explained by the relatively low anxiety levels in the current ample. However, this will require further investigation to verify whether the effect is consistent across different samples.

A significant difference was found between the English CHIME total scores of meditators and non-meditators. This effect was consistent when examining the original ordinal data and the Rasch transformed interval data. Although the overall effect sizes were similar, the result does not undermine the benefits of conducting the ordinal-to-interval conversion because the transformation increases the precision of measurement. Furthermore, this result demonstrates that the English CHIME maintains its excellent psychometric properties, even without the ordinal-interval conversion.

No differences were found in the functioning of the English CHIME items across the personal factors of meditation practice, gender, and age. Additionally, the person-item plots show that the English CHIME items discriminate near perfectly between different levels of meditation experience, meaning that the scale functions just as well for meditators and non-meditators. These results are identical to those found in the original validation of the CHIME scale, and taken together, provide strong evidence that the CHIME maintains its reliability and validity across different populations.

Limitations and Future Directions

The primary limitation of the present research stems from the use of online survey data collection. Although the use of online data collection is becoming more prevalent, it

can present research with unique challenges. In the present research, participants voluntarily signed up to participate in the study. Even though the proportion of population engaging in mindfulness practice is growing, the sample may still be affected by self-selection bias, and therefore been skewed towards individuals who were familiar with the concept of mindfulness. This effect is evidenced by the relatively high number of participants who indicated that they practiced mindfulness or engaged in some form of contemplative practice (38.4%). The higher-than-average level of mindfulness in this sample could distort the correlations between other psychometric measures, an effect which may explain the non-significant relationship between the English CHIME total score and DASS-Anxiety. Future research could address this issue by selecting participants based on different sampling characteristics. In addition, we could not provide any solid evidence on the nature of the correlations, and they remain assumed as a residual phenomenon observed in research based on self-reported measures. When multiple constructs are measured using common methods (e.g., multiple-item scales presented within the same survey), this may result in spurious correlations due to the assessment tools rather than to the constructs being measured (Podsakoff et al., 2012).

The present study included participants from the general population; and therefore, future research should seek to replicate this research in different population types, such as people diagnosed with affective conditions or other psychological disorders. Both samples in this study consisted largely of participants who identified themselves as White American. Future research could use more diverse samples to generalize this study's findings to other cultural and ethnic groups. Nonetheless, Rasch analysis is less dependent on sample characteristics compared with other traditional methods (e.g., factor analysis) (Tennant & Conaghan, 2007). Additionally, the sample utilized in the present study was large enough to ensure the robustness and generalizability of these results. For its German-language version, the CHIME has been shown to be stable over time, with little evidence that items are interpreted differently before and after a mindfulness intervention (Krägeloh et al., 2018). Such evidence is important as it demonstrates that the questionnaire is evaluated using the same standard at both measurement times. Future work will need to confirm the absence of such a response shift in the English CHIME version.

Perhaps the most interesting distinction between the validation of the original German scale and the translated English scale is the different super-item structure required to obtain the best Rasch model fit. In the original German validation (Medvedev et al., 2019), adequate fit was confirmed when the individual CHIME items were summed to form 8 super-items, which is in line with the subscales of the original measure. However, in the present study, the

fit was achieved after summing the items to form 5 super-items. Taken alone, this result suggests that the CHIME may be affected by spurious residual correlations that are different in English compared to German. The question then is, do the linguistic conceptualization and understanding of mindfulness itself differ significantly across different languages and cultures? Further research is required to address this question. Replications of the psychometric analyses for all of the available languages of the CHIME are also necessary to ensure that results are generalizable and not due to overfitting (Nosek et al., 2022). Although some of the differences found between the German- and English-language versions are relatively minor, future work could explore to what extent such differences may be replicable when the two versions are directly compared in a pooled psychometric analysis.

The facets of the CHIME represent the relevant aspects of the overarching mindfulness construct as developed by Bergomi et al. (2014) and are theoretically supported in Bergomi et al. (2013). The relevance of these facets to the overarching mindfulness construct was demonstrated by Rasch analysis of the German CHIME version by Medvedev et al. (2018). This study aimed to validate the CHIME as a measure of the overarching mindfulness construct that can be measured using one single interval level score, like physical phenomena such as height, weight, temperature, and speed. We acknowledge that measuring individual facets included in the CHIME might be beneficial for specific purposes; however, none of these facets would represent a valid construct of mindfulness by themselves. Therefore, in this study, we focused on developing an overarching mindfulness scale that produces a single interval level mindfulness assessment score.

As our understanding of the benefits of mindfulness continues to grow, it is becoming increasingly valuable for researchers and clinicians to be able to measure the mindfulness construct accurately. The present study utilized Rasch analysis methods to explore the psychometric characteristics of the English CHIME scale. The results demonstrate that the CHIME maintains its structural validity and reliability after being translated from German to English. However, a slightly different super-item structure was used for the English CHIME compared to the German CHIME, which may be related to error variances associated with linguistic differences or other unknown method effects and requires further investigation. Therefore, replications of the psychometric analyses for all of the available languages of the CHIME are also necessary to ensure that results are generalizable and not due to overfitting (Nosek et al., 2022). Although some of the differences found between the German- and English-language versions are relatively minor, future work could explore to what extent

such differences may be replicable when the two versions are directly compared in a pooled psychometric analysis.

Declarations

Ethical Approval This study was approved by the Pacific University Institutional Research Board (protocol number 131-17), which follows internationally recognized ethical standards.

Informed Consent All participants provided informed consent to participate in the study.

Conflict of Interest We declare no conflicts of interest in connection with this paper.

References

Andrich, D., Sheridan, B., & Luo, G. (2009). *RUMM 2030*. RUMM Laboratory.

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10*(2), 176–181. https://doi.org/10.1037/1040-3590.10.2.176

Baer, R., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment, 11*(3), 191–206. https://doi.org/10.1177/1073191104268029

Baer, R., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment, 13*(1), 27–45. https://doi.org/10.1177/1073191105283504

Balalla, S. K., Medvedev, O. N., Siegert, R. J., & Krägeloh, C. U. (2019). Validation of the WHOQOL-BREF and shorter versions using Rasch analysis in traumatic brain injury and orthopedic populations. *Archives of Physical Medicine and Rehabilitation, 100*(10), 1853–1862. https://doi.org/10.1016/j.apmr.2019.05.029

Beaujean, A. A., Benson, N. F., McGill, R. J., & Dombrowski, S. C. (2018). A misuse of IQ scores: Using the dual discrepancy/consistency model for identifying specific learning disabilities. *Journal of Intelligence, 6*(3), 36. https://doi.org/10.3390/jintelligence6030036

Bennett, K., & Dorjee, D. (2016). The impact of a mindfulness-based stress reduction course (MBSR) on well-being and academic attainment of sixth-form students. *Mindfulness, 7*(1), 105–114. https://doi.org/10.1007/s12671-015-0430-7

Bergomi, C., Tschacher, W., & Kupper, Z. (2013). The assessment of mindfulness with self-report measures: existing scales and open issues. *Mindfulness, 4*(3), 191–202. https://doi.org/10.1007/s12671-012-0110-9

Bergomi, C., Tschacher, W., & Kupper, Z. (2014). Konstruktion und erste Validierung eines Fragebogens zur umfassenden Erfassung von Achtsamkeit [Construction and Initial Validation of a Questionnaire for the Comprehensive Investigation of Mindfulness]. *Diagnostica, 60*, 111–125. https://doi.org/10.1026/0012-1924/a000109

Bond, T., & G., & Fox, C. M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences* (2nd ed.). Routledge.

Brogden, H. E. (1977). The Rasch model, the law of comparative judgment and additive conjoint measurement. *Psychometrika, 42*(4), 631–634. https://doi.org/10.1007/BF02295985

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822–884. https://doi.org/10.1037/0022-3514.84.4.822

Christensen, K. B., Makransky, G., & Horton, M. (2016). Critical values for Yen's Q3: identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*(3), 178–194. https://doi.org/10.1177/0146621616677520

Cladder-Micus, M. B., Verweij, H., van Ravesteijn, H., Van der Gucht, K., Raes, F., & Speckens, A. E. (2019). Validation of the Dutch Comprehensive Inventory of Mindfulness Experiences (CHIME) and development of a short form (CHIME-SF). *Mindfulness, 10*(9), 1893–1904. https://doi.org/10.1007/s12671-019-01125-7

de Vibe, M., Solhaug, I., Rosenvinge, J. H., Tyssen, R., Hanley, A., & Garland, E. (2018). Six-year positive effects of a mindfulness-based intervention on mindfulness, coping and well-being in medical and psychology students; results from a randomized controlled trial. *PloS ONE, 13*(4), e0196053. https://doi.org/10.1371/journal.pone.0196053

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Finaulahi, K. P., Sumich, A., Heym, N., & Medvedev, O. N. (2021). Investigating psychometric properties of the Self-Compassion Scale using Rasch methodology. *Mindfulness, 12*(3), 730–740. https://doi.org/10.1007/s12671-020-01539-8

Fumero, A., Peñate, W., Oyanadel, C., & Porter, B. (2020). The effectiveness of mindfulness-based interventions on anxiety disorders. A systematic meta-review. *European Journal of Investigation in Health, Psychology and Education, 10*(3), 704–719. https://doi.org/10.3390/ejihpe10030052

Goh, H. E., Marais, I., & Ireland, M. J. (2017). A Rasch model analysis of the Mindful Attention Awareness Scale. *Assessment, 24*(3), 387–398. https://doi.org/10.1177/1073191115607043

Hagell, P., & Westergren, A. (2016). Sample size and statistical conclusions from tests of fit to the Rasch model according to the Rasch unidimensional measurement model (Rumm) program in health outcome measurement. *Journal of Applied Measurement, 17*(4), 416–431.

Hobart, J., & Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technology Assessment, 13*(12), 1–200. https://doi.org/10.3310/hta13120

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. Spielberger (Eds.), Adapting educational and psychological tests for cross-cultural assessment (pp. 3-38). Lawrence Erlbaum Publishers.

Harvey, P. (2013). *Dukkha*, non-self, and the teaching on the four "noble truths". In S. M. Emmanuel (Ed.), *A companion to Buddhist philosophy* (pp. 26–45). Wiley. https://doi.org/10.1002/9781118324004.ch2

Johnson, C., Burke, C., Brinkman, S., & Wade, T. (2016). Development and validation of a multifactor scale in youth: The Comprehensive Inventory of Mindfulness Experiences – Adolescents (CHIME-A). *Psychological Assessment, 29*(3), 264–281. https://doi.org/10.1037/pas0000342

Kabat-Zinn, J. (2003). Mindfulness based interventions in clinical psychology. *Science and Practice, 10*(2), 144–156. https://doi.org/10.1093/clipsy.bpg016

Krägeloh, C. U., Bergomi, C., Siegert, R. J., & Medvedev, O. N. (2018). Response shift after a mindfulness-based intervention: measurement invariance testing of the Comprehensive Inventory of Mindfulness Experiences. *Mindfulness, 9*(1), 212–220. https://doi.org/10.1007/s12671-017-0764-4

Krägeloh, C. U., Henning, M. A., Medvedev, O. N., Feng, X. J., Moir, F., Billington, R., & Siegert, R. J. (2019). *Mindfulness-based intervention research: characteristics, approaches, and developments*. Routledge.

Linehan, M. M. (1993). *Skills training manual for treating borderline personality disorder*. New York: Guilford Press.

Lord, F. M., & Novick, M. R. (1968). *Statistical theory of mental test scores*. Addison-Wesley.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy, 33*(3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-u

Lundgren-Nilsson, Å., Jonsdottir, I. H., Ahlborg, G., & Tennant, A. (2013). Construct validity of the Psychological General Well Being Index (PGWBI) in a sample of patients undergoing treatment for stress-related exhaustion: a Rasch analysis. *Health and Quality of Life Outcomes, 11*, 2. https://doi.org/10.1186/1477-7525-11-2

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. https://doi.org/10.1007/BF02296272

Medvedev, O. N., Bergomi, C., Röthlin, P., & Krägeloh, C. U. (2019). Assessing the psychometric properties of the Comprehensive Inventory of Mindfulness Experiences (CHIME) using Rasch analysis. *European Journal of Psychological Assessment, 35*(5), 650–657. https://doi.org/10.1027/1015-5759/a000453

Medvedev, O. N., & Krägeloh, C. U. (2022). Rasch measurement model. In O. N. Medvedev, C. U. Krägeloh, R. J. Siegert, & N. N. Singh (Eds.), *Handbook of assessment in mindfulness research*. Springer. https://doi.org/10.1007/978-3-030-77644-2_4-1

Medvedev, O. N., Siegert, R. J., Feng, X. J., Billington, D. R., Jang, J. Y., & Krägeloh, C. U. (2016). Measuring trait mindfulness: how to improve the precision of the Mindful Attention Awareness Scale using a Rasch model. *Mindfulness, 7*(2), 384–395. https://doi.org/10.1007/s12671-015-0454-z

Medvedev, O. N., Siegert, R. J., Kersten, P., & Krägeloh, C. U. (2016). Rasch analysis of the Kentucky Inventory of Mindfulness Skills. *Mindfulness, 7*(2), 466–478. https://doi.org/10.1007/s12671-015-0475-7

Medvedev, O. N., Turner-Stokes, L., Ashford, S., & Siegert, R. J. (2018). Rasch analysis of the UK Functional Assessment Measure in patients with complex disability after stroke. *Journal of rehabilitation medicine, 50*(5), 420–428. https://doi.org/10.2340/16501977-2580

Medvedev, O. N., Siegert, R. J., Kersten, P., & Krägeloh, C. U. (2017). Improving the Precision of the Five Facet Mindfulness Questionnaire using a Rasch approach. *Mindfulness, 8*(4), 995–1008. https://doi.org/10.1007/s12671-016-0676-8

Mitchell-Parker, K., Medvedev, O. N., & Siegert, R. J. (2017). Rasch analysis of the Frost Multidimensional Perfectionism Scale. *Australian Journal of Psychology, 70*(3), 258–268. https://doi.org/10.1111/ajpy.12192

Morton, M. L., Helminen, E. C., & Felver, J. C. (2020). A systematic review of mindfulness interventions on psychophysiological responses to acute stress. *Mindfulness, 11*(9), 2039–2054. https://doi.org/10.1007/s12671-020-01386-7

Neff, K. D. (2003). The development and validation of a scale to measure self-compassion. *Self and Identity, 2*(3), 223–250. https://doi.org/10.1080/15298860390209035

Norquist, J. M., Fitzpatrick, R., Dawson, J., & Jenkinson, C. (2004). Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Medical Care, 42*(1), I25–I36. https://doi.org/10.1097/01.mlr.0000103530.13056.88

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology, 73*, 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Oyler, D. L., Pratscher, S. D., Leng, P., Medvedev, O. N., & Bettencourt, B. (2022). Over-identifying with social roles: Selfing Scale development and validation. *Mindfulness, 13*(8), 1957–1970. https://doi.org/10.1007/s12671-022-01931-6

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Raes, F., Pommier, E., Neff, K. D., & Van Gucht, D. (2011). Construction and factorial validation of a short form of the self-compassion scale. *Clinical Psychology & Psychotherapy, 18*(3), 250–255. https://doi.org/10.1002/cpp.702

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test*. Danish Institute for Educational Research.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Symposium conducted at the meeting of the Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press.

Sandham, M. H., Medvedev, O. N., Hedgecock, E., Higginson, I. J., & Siegert, R. J. (2019). A Rasch analysis of the Integrated palliative Care Outcome Scale. *Journal of Pain and Symptom Management, 57*(2), 290–296. https://doi.org/10.1016/j.jpainsymman.2018.11.019

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. https://doi.org/10.1007/s12671-012-0145-y

Smith, E. V. (2002). Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205–231.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism, 57*(8), 1358–1362. https://doi.org/10.1002/art.23108y

Tennant, A., & Pallant, J. F. (2006). Unidimensionality matters! (a tale of two Smiths?). *Rasch Measurement Transactions, 20*, 1048–1051.

Thomas, R., Chur-Hansen, A., & Turner, M. (2020). A systematic review of studies on the use of Mindfulness-Based Cognitive Therapy for the treatment of anxiety and depression in older people. *Mindfulness, 11*(7), 1599–1609. https://doi.org/10.1007/s12671-020-01336-3

Thurstone, L. L. (1931). The measurement of social attitudes. *Abnormal and Social Psychology, 27*, 249–269.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*, 209–220. https://doi.org/10.1037/a0023353

Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: a case for testlets. *Journal of Educational Measurement, 24*(3), 185–201.

Walach, H., Buchheld, N., Buttenmüller, V., Kleinknecht, N., & Schmidt, S. (2006). Measuring mindfulness—the Freiburg Mindfulness Inventory (FMI). *Personality and Individual Differences, 40*(8), 1543–1555. https://doi.org/10.1016/j.paid.2005.11.025

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037//0022-3514.54.6.1063

Authors and Affiliations

Samuel Wilkinson[1] · Leticia Ribeiro[2] · Christian U. Krägeloh[3] · Claudia Bergomi[4] · Marie Parsons[5] · Alex Siegling[6] · Wolfgang Tschacher[7] · Zeno Kupper[7] · Oleg N. Medvedev[1]

Samuel Wilkinson
samwilkinson959@gmail.com

Leticia Ribeiro
lribeiro@iastate.edu

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

Marie Parsons
parsone4@miamioh.edu

Alex Siegling
alexbsiegling@gmail.com

Wolfgang Tschacher
wolfgang.tschacher@upd.unibe.ch

Zeno Kupper
zeno.kupper@upd.unibe.ch

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

[2] Iowa State University, Student Counselling Center, IA, Ames, USA

[3] Auckland University of Technology, School of Clinical Sciences, Auckland, New Zealand

[4] Bern, Switzerland

[5] Boston University, Department of Psychological and Brain Sciences, Boston, USA

[6] University College London, Division of Psychology and Language Sciences, London, UK

[7] University of Bern, University Hospital of Psychiatry and Psychotherapy, Bern, Switzerland