Mindfulness (2022) 13:3017–3027
https://doi.org/10.1007/s12671-022-02006-2

ORIGINAL PAPER

Enhancing the Assessment of Gratitude in Mindfulness Research: a Rasch Analysis of the 6-Item Gratitude Questionnaire

Emerson Bartholomew[1] · **Navad Iqbal**[2] · **Oleg Medvedev**[3]

Accepted: 8 October 2022 / Published online: 18 October 2022

Abstract

Objectives The 6-item gratitude questionnaire (GQ-6) is a widely used measure reported to be reliable and valid under traditional and Rasch investigations. However, recent investigations were inconsistent regarding the item structure of the GQ-6, with three investigations concluding that item 6 be removed. Previous Rasch analyses also did not produce interval conversion tables, a key benefit of this method which provides a means of improving scale accuracy and aligning the scale with the fundamental principles of measurement.

Methods A Partial Credit Rasch Model was used to evaluate psychometric properties of the GQ-6 using a combined sample of 663 respondents from the USA ($n = 345$) and India ($n = 318$).

Results Initial analysis showed significant scale dysfunction, with overall and individual item misfit, local dependency, disordered response thresholds, instances of differential item functioning by age and nationality, along with poor reliability. Through the use of recent advances in the Rasch methodology, locally dependent items were combined into two super-items and the best fit to the Rasch model was obtained with evidence of strict unidimensionality, invariance, and strong reliability. Results indicate the GQ-6 is suitable for individual and group assessment, while also permitting the creation of an ordinal-to-interval conversion algorithm which has been provided here.

Conclusions This study confirmed the robust psychometric properties of the GQ-6 after minor modifications and provides a means for clinicians and researchers to improve the accuracy of this widely used measure in mindfulness research and other relevant studies without modification of its original response format.

Keywords Rasch Analysis · Gratitude · Measurement · Validity · Reliability

Gratitude is a construct comprising appreciation and thankfulness in daily living, as well as in the level of grateful response to a perceived beneficial event (Emmons, 2004; McCullough et al., 2002). Often referred to as the sister of mindfulness (Rosenzweig, 2013), gratitude has shown a consistent positive association with mindfulness (Schutte et al., 2021; Seear & Vella-Brodrick, 2013) and is increasingly recognized as an important component in enhancing its benefits. Gratitude has been shown to serve as a mediator in the relationship of mindfulness to both life satisfaction, mood, and psychological well-being (Chen et al., 2017; Swickert et al., 2019; Voci et al., 2019). Several studies posit that mindfulness is a foundational pre-requisite to unlocking the benefits of gratitude, and several of these works utilized the six-item, short-form gratitude questionnaire (GQ-6; McCullough et al., 2002) as the primary assessment tool (Jankowski & Sandage, 2014; Ivtzan et al., 2016; Reeves et al., 2021; Voci et al., 2019; Yang et al., 2021). This gratitude-mindfulness connection has significant measurement implications as both variables are increasingly being used as predictors in investigations of physical health measures.

Gratitude has been implicated in improvements to chronic conditions such as diabetes (Schache et al., 2019b), inflammatory bowel disease (Sirois & Wood, 2017), HIV (Moskowitz et al., 2017), fibromyalgia (Toussaint et al., 2017), arthritis, and chronic obstructive pulmonary disease

✉ Emerson Bartholomew
mbar321@aucklanduni.ac.nz

1 Department of Psychological Medicine, Faculty of Medical and Health Sciences, University of Auckland, 85 Park Road, Grafton, Private Bag 92019, Auckland, New Zealand

2 Department of Psychology, Jamia Millia Islamia Central University, New Delhi, India

3 School of Psychology, University of Waikato, Hamilton, New Zealand

(Eaton et al., 2013). Specific physical health measures include decreases in ambulatory blood pressure, eating disorder behaviour, re-admission rates for those hospitalized with acute cardiac events (Huffman et al., 2016; Redwine et al., 2016), improvements in sleep quality, hours spent exercising, and positive changes in heart rate variability and inflammatory markers (Bai et al., 2019; Digdon & Koble, 2011; Emmons & McCullough, 2003; Heckendorf et al., 2019; Jackowska et al., 2016; Mills et al., 2015; Redwine et al., 2016; Schache et al., 2019b; Wolfe & Patterson, 2017).

Whether physical health benefits from gratitude are moderated by levels of mindfulness, or gratitude is a pre-requisite for many mindfulness benefits, in either case it becomes critical to improve the measurement accuracy of gratitude. Given the currently ordinal level of gratitude measures, scores collected using scales such as the GQ-6 are unsuitable for addition/subtraction, analysis by parametric statistics, or in comparison to interval/ratio level data such as the aforementioned physical health outcomes.

The GQ-6 is a unique and convenient measure of gratitude, and shorter tests such as this have advantages in a higher willingness to respond as well as provide more thoughtful answers (Fan & Yan, 2010). Several previous studies have examined the GQ-6 for its psychometric properties and have found evidence of good reliability and validity (Kashdan et al., 2009; Wood et al., 2009). Cross-cultural investigations assessing the GQ-6 have also spanned more than eight languages and unique cultural contexts, as well as targeting several specific populations i.e. firefighters, athletes, high school students, and undergraduates. These previous validation studies ratify the reliability and validity of the GQ-6 under a classical test theory (CTT) approach (i.e. confirmatory factor analysis), but to date, there have been few Rasch investigations. One such Rasch study carried out by Valdez and Chu (2020) examined the scale using a student sample from a 'collectivist' society of the Philippines and tested for a correlation of gratitude to academic motivation. The investigation found a significant ceiling effect and reinforced a previous conclusion by Chen et al. (2009) that recommended a 5-item solution with the removal of the sixth item based on a non-significant factor loading. These findings were supported in a study by Froh et al. (2011) working under a CTT approach, which also removed item 6, citing low factor loadings and interviews with their adolescent sample who expressed it was vague. However, removal of scale items can have negative impacts on reliability and content validity, especially in the case of already short scales such as the GQ-6. Recent advances in Rasch methodology have made possible the rectification of scale dysfunction without item removal, and this solution will be explored to see if psychometric properties remain satisfactory (Medvedev et al., 2018a, 2018b).

A recent investigation of the GQ-6 in an Indian sample endorsed the scales' validity and reliability, but was not able to test for measurement invariance between another sample such as from the USA (Garg et al., 2021). Measurement invariance between cultures is vital for meaningful cross-cultural comparisons of scores, and without this property, prevalence figures and comparisons could be misleading to readers.

Polytomous Rasch analysis is a commonly used modern test theory technique based on principles outlined by Rasch (1960), and it may be conducted using either the rating scale model (Andrich, 1978) or the partial credit model (Masters, 1982). Polytomous Rasch analysis has increasingly become the gold standard for statistical analysis in fields such as education, agriculture, psychology, and medical and health science for its effectiveness in evaluating and improving latent trait measures (Kreiner & Christensen, 2007). The main advantage the Rasch approach has over traditional analyses is the transformation of ordinal scores into a linear measure, which improves accuracy of measurement. This approach also examines several key features in addition to standard validity and reliability, such as unidimensionality, appropriate sample targeting, local dependency, and differential item functioning (DIF) (Tennant & Conaghan, 2007).

Our study aimed to examine the GQ-6 from a modern test theory perspective using Rasch analysis. We hypothesize that our approach would highlight key aspects of scale dysfunction and that these issues can be rectified without removing items. Retaining items is advantageous in that it preserves both construct validity and standard scale administration. The USA and India are the largest English-speaking countries in the world and while India has a larger and more culturally diverse population compared to the USA, the USA is also a multicultural country. Therefore, a cross-cultural sample from the USA and India was used to ensure generalizability of the results and to test for differential item functioning (DIF) between cultures.

Method

Participants

Data were collected in two samples. Firstly, 500 US-based respondents were recruited via Amazon's Mechanical Turk service from the 19th of July 2021 to the 29th of July 2021. This sample was reduced from 500 to 345 for quality optimization purposes based on participant completion times; participants who took less than 55 s to complete the test were excluded. We assumed that responding to a 6-item scale should take a minimum of 10 s per item (60 s in total); therefore, all responses with less than 10 s per item (e.g. 54 s) were considered as invalid. Sample 1 ($n = 345$) consisted of 206 (59.8%) males, 138 (40.0%) females, and 1 (0.2%) other. The mean age was 36 years ($SD = 12.46$) and ages ranged from 18 to 64. Sample 1 respondents were incentivized by payment of 80c USD for their response. The second sample (sample 2) was collected from 318 respondents from India via Google Forms (82 non-responders were left out). There were no participants who completed the survey for less than 55 s in this sample. Data were collected between the 20th of January 2021 and the 20th of February 2021. Sample 2 was comprised of 137 (43.1%) males and 181 (56.9%) females, and the age of the participants varied from 20 to 50 years ($M = 28.97$, $SD = 5.97$). No incentives were given to the subjects from sample 2 to participate in the study.

Sample size recommendations for Rasch analysis (Hagell & Westergren, 2016; Linacre, 1994) were met for both samples while the total sample exceeded recommendations. Data were split into four age groups (18–30, 31–45, 46–55, 56 +) for later testing of potential differential item functioning (DIF). Ethics approval was obtained from the authors' local ethics committee, and all participants provided informed consent to participate in the study.

Procedure

Participants were invited to take part via Amazon Mechanical Turk in the US sample, and by email or Whatsapp message in the sample from India. Each respondent reviewed a participant information sheet and was made aware that responses given could be included in data analysis and publication in scholarly journals. Participants were asked to indicate whether they consented to participate, and those who did not were not included in the final dataset.

Measures

This study used the 6-item gratitude questionnaire developed by McCullough et al. (2002), a Likert-type self-report measure ranging from 1 = strongly disagree to 7 = strongly agree, with items 3 and 6 requiring reverse coding before computing the total score. The total score is the sum of individual items scores, with higher scores corresponding to higher levels of gratitude. The scale reliability Cronbach's alpha (α) was 0.71 and McDonalds omega (ω) did not converge with US sample 1, indicating the need for Rasch analysis to improve psychometric properties of the scale, while $\alpha = 0.82$ and $\omega = 0.88$ were obtained with Indian sample 2.

Data Analyses

Our analysis utilized two forms of software, IBM SPSS v26 for descriptive statistics and RUMM2030 (Andrich et al., 2009) for Rasch analysis. There are two Rasch models available for use with polytomous data, the rating scale model (RS; Andrich et al., 2009) and the partial credit model (PCM; Masters, 1982). To determine which was appropriate for our use here, we conducted an Andersen's likelihood ratio test (LRT; Andersen, 1973). Our findings from the LRT indicated that an unrestricted PCM Rasch model approach would be most appropriate for our data. This investigation has focused on several key features of importance based on guidelines outlined by Wright et al. (1994) and Tennant and Conaghan (2007). Rasch analysis is an iterative process aimed at the achievement of a satisfactory fit to the Rasch model in terms of both an overall and individual item fit along with unidimensionality. Unidimensionality is tested via Smith's test and is confirmed if the number of significant comparisons using *t* test (or the lower bound of the binominal confidence interval around the percentage of significant *t* tests) is below 5% (Medvedev et al., 2016; Smith, 2002; Tennant & Pallant, 2006). Model fit requires a non-significant item-trait interaction tested using chi-square, and item fit requires residuals between -2.50 and $+2.50$. Scale items must also be locally independent to prevent spurious correlations, and if found above a level of 0.20 should be considered for the creation of super-items (Christensen et al., 2016; Medvedev et al., 2018a, 2018b; Wainer & Kiely, 1987). Sample targeting is acceptable when the average sample location is ± 0.50 when compared to a mean item difficulty of 0. Satisfactory fit to the Rasch model allows for the creation of a linear conversion algorithm enabling the measurement of gratitude on a finer, interval level (Medvedev et al., 2020). In Rasch approaches, reliability is tested via Person Separation Index (PSI), with guidelines by Tennant and Conaghan (2007) outlining 0.70 as a minimum threshold for group and 0.80 for individual use. Differential item functioning is another source of potential dysfunction and will be evaluated based on age, gender, and sample. Lastly, response category thresholds may be disordered and will be checked via item characteristic curves (ICCs) and if found may be addressed with scale adjustment (collapsing of disordered categories).

Results

Results of Andersen's likelihood ratio test were significant ($p < 0.01$), suggesting the use of the unrestricted partial credit version of the Rasch model (Masters, 1982) due to inconsistency in response option thresholds between items required for the RS model (Andrich et al., 2009). Initial analysis displayed in Table 1 showed the overall model fit statistics poorly fitting the Rasch model, indicated by a significant item-trait interaction chi-square ($p < 0.01$) suggesting that measurement properties of individual items are

Table 1 Summary of the Rasch model fit statistics for the initial and final analysis of the 6-item gratitude scale ($n = 643$)

Analyses	Item fit residual		Person fit residual		Goodness of fit		PSI	Unidimensionality
	Mean	SD	Mean	SD	χ^2 (df)	*p*		Sig. *t* test in %
Initial	0.30	3.37	−0.22	0.85	514.75 (48)	<0.01	.59	2.11 (yes)
Final	0.28	0.14	−0.56	0.88	20.69 (16)	.191	.81	3.32 (yes)

PSI Person Separation Index without extremes

not consistent across different trait levels. Reliability was below acceptable levels for both group (> 0.70) and individual (> 0.80) assessment with a PSI value below 0.60 (Table 1). Initial overall item fit statistics showed a large standard deviation (3.37) clearly reflecting deviation from the model expectations. Table 2 presents a closer view of initial item fit statistics, showing significant misfits for five out of six items.

The scale showed disordered thresholds for all items, and can be seen for items 1 and 2 in Fig. 1, representative of the disordering patterns for all other items, where categories 1, 3, and 4 are never modal in item 1, and categories 2 and 3 are never modal for item 2. Item 6 "Long amounts of time can go by before I feel grateful to something or someone" displayed significant uniform DIF by age ($F(2, 641) = 27.22$, $p < 0.001$), with older people scoring lower on this item compared to younger people which was also confirmed by visual examination of the plot and subsequent sign test (Balalla et al., 2020). Items 2, 4, 5, and 6 showed significant uniform DIF by sample, with Indian participants scoring higher on item 2 ($F(1, 641) = 138.85$, $p < 0.001$), 4 ($F(1, 641) = 487.47$, $p < 0.001$), and 6 ($F(1, 641) = 610.22$, $p < 0.001$), and lower on item 5 ($F(1, 641) = 184.42$, $p < 0.001$). This uniform DIF was also confirmed by visual examination of the plots and sign tests.

Significant DIF by gender was observed only for item 5 "As I get older I find myself more able to appreciate the people, events, and situations that have been a part of my life history" ($F(1, 641) = 13.76$, $p < 0.001$). However, visual examination and subsequent sign-test comparison demonstrated that sex differences were not statistically consistent across class intervals ($p > 0.05$.).

Tests for unidimensionality (principal components analysis of residuals and equating *t* tests) showed strong evidence for strict unidimensionality (Table 1). Examination of the residual correlation matrix showed concerning correlations between items 1 and 5, as well as between items 2 and 4. These correlations were larger than the maximum permissible magnitude of 0.20 with reference to the mean of all residual correlations as outlined by Christensen et al. (2016).

Several options were considered in seeking to correct these issues and improve the scale. However, to retain conceptual importance, removing items would be considered as a last resort given that this widely used scale has only 6 items. Therefore, to address local dependency, two super-items were created by combining items 1, 5, and 6 (super-item 1) and items 2, 3, and 4 (super-item 2) using a well-established methodology (Lundgren-Nilsson et al., 2013; Medvedev et al., 2018a, 2018b). The best model fit was achieved after the creation of super-items, as evidenced by a non-significant item-trait interaction chi-square value of 20.69 ($p = 0.191$), indicating no significant deviation of the modified scale from the fundamental principles of

Table 2 Initial items fit statistics of 6-item gratitude questionnaire ($n = 350$)

No	Item content	Location	Fit Resid	Chi Sq
	Initial analysis			
1	I have so much in life to be thankful for	−0.01	−4.17*	103.80*
2	If I had to list everything that I felt grateful for, it would be a very long list	−0.08	−1.78	89.77*
3	When I look at the world, I don't see much to be grateful for	0.06	2.30	91.07*
4	I am grateful to a wide variety of people	−0.29	−1.28	62.96*
5	As I get older, I find myself more able to appreciate the people, events, and situations that have been part of my life history	0.39	1.45	6.42
6	Long amounts of time can go by before I feel grateful to something or someone	−0.08	5.28*	160.72*
	Final analysis of the best fit: 2 super-items			
	Super-item 1 (1 + 5 + 6)	−0.03	0.18	12.32
	Super-item 2 (2 + 3 + 4)	0.03	−0.38	8.37

Fit Resid fit residual, *Chi Sq* chi-square

*Significant item misfit to the Rasch model $p < 0.05$

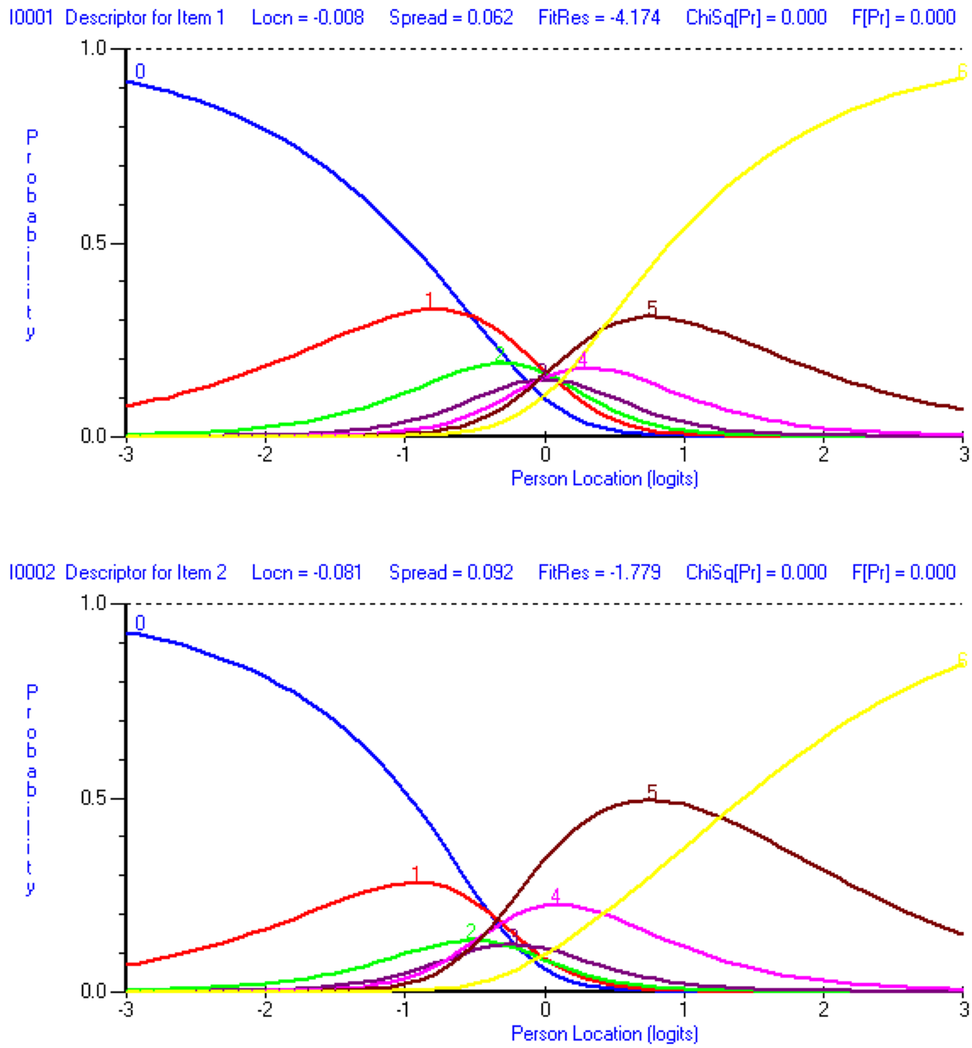

Fig. 1 Item characteristic curve (ICC) for item 1 (above) and 2 (below) before rescoring

measurement defined by the Rasch model. Reliability was also improved markedly with a PSI of 0.81, indicating the appropriateness of the modified GQ-6 for assessment at a group level. Disordered thresholds showed improvement with no categories working inappropriately, as illustrated in Fig. 2. Both super-items demonstrated good model fit and no significant DIF by personal factors verified by visual examination of the plots and sign tests (Table 2).

Figure 3 shows the person-item threshold distribution of the GQ-6 after modification and demonstrates that item thresholds cover over 95% of sample abilities with no significant floor or ceiling effects. The distribution of persons is slightly skewed to the right, with a person mean of 0.93 indicating the overall high level of gratitude of the current sample.

Our modifications allowed the scale to correct several issues without removing items, keeping the original scale and its scoring method intact. Conversion tables have been included to assist clinicians or researchers in converting their data from an ordinal to interval level, but those wishing to do so must remember to reverse code required items and not to include any respondents with missing data (Table 3).

Discussion

This study aimed to investigate the characteristics of the commonly utilized 6-item gratitude scale using Rasch analysis. Our initial results showed scale dysfunction across several metrics, and scale modifications were undertaken to rectify these issues resulting in a satisfactory fit to the unidimensional Rasch model. Acceptable fit permitted the development of a conversion algorithm to transform ordinal raw GQ-6 responses into interval-level data without the need to modify the original scale format, enhancing measurement accuracy.

Initial analysis showed several areas of scale dysfunction; however, one positive sign was the evidence for strict unidimensionality. Aspects of scale dysfunction

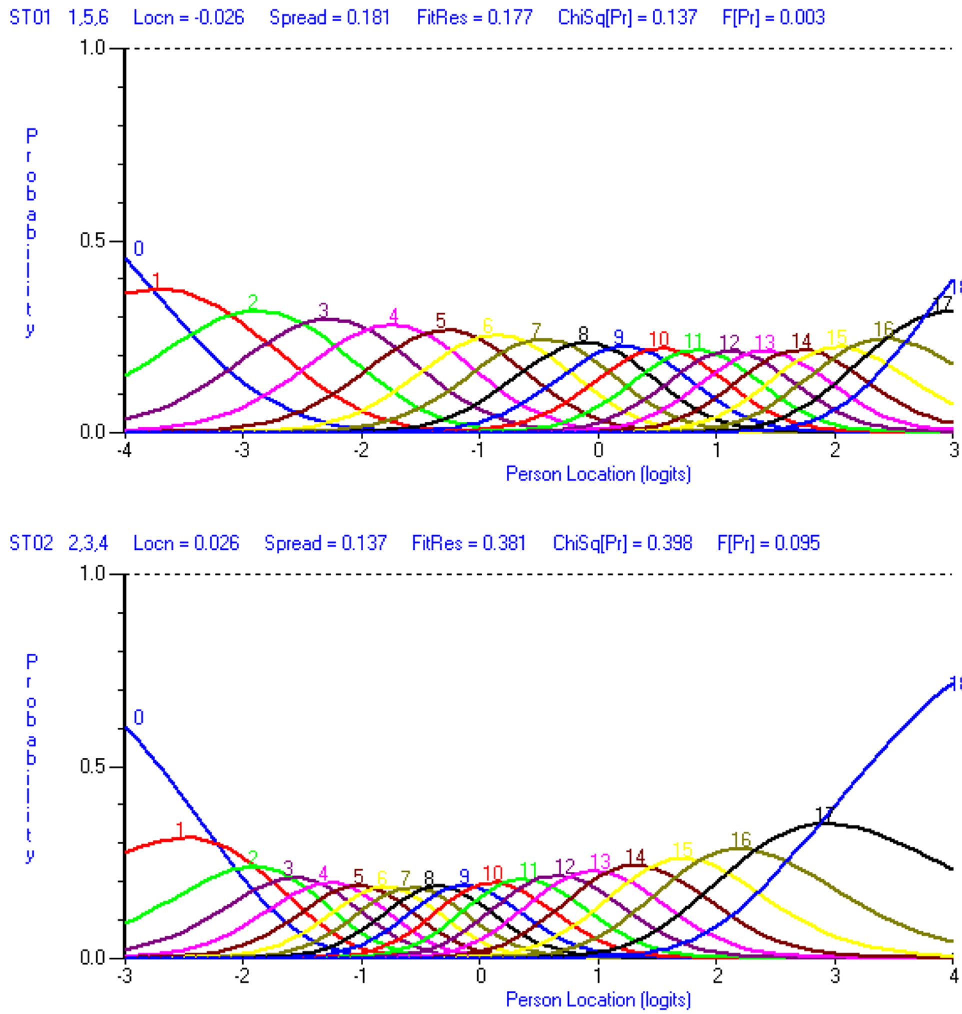

Fig. 2 Item characteristic curve (ICC) for super-items 1 (above) and 2 (below)

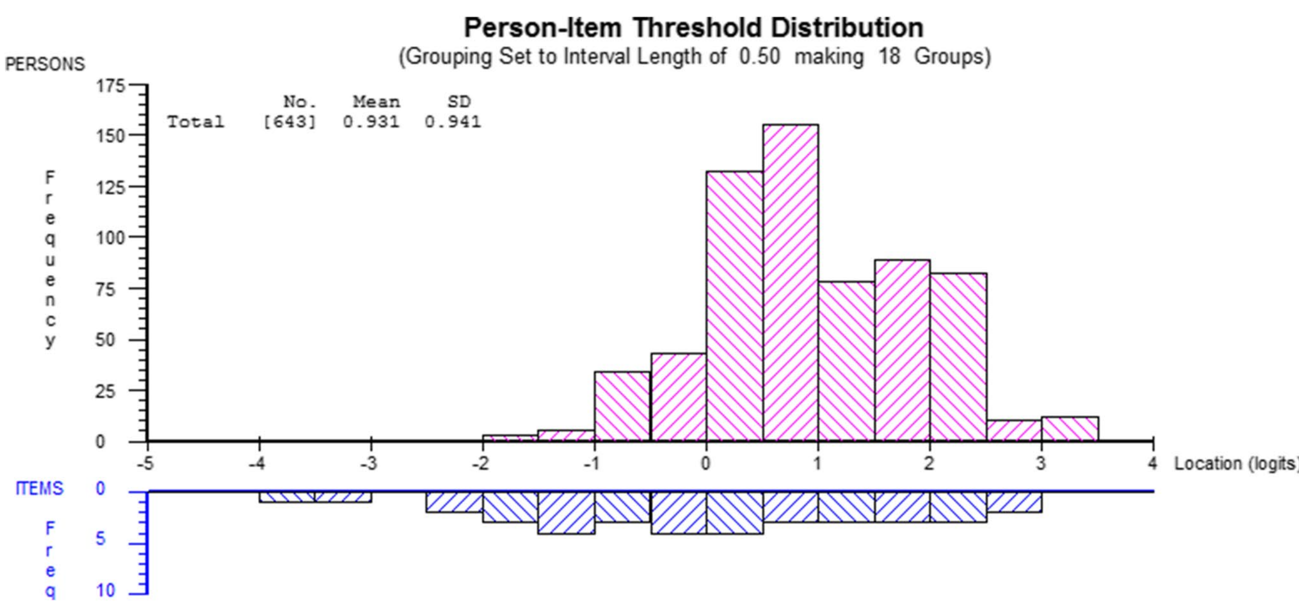

Fig. 3 Person-item threshold distribution for the GQ-6 analysis of the best fit ($n = 643$)

Table 3 Converting from ordinal- to interval-level scores for the 6-item gratitude questionnaire

Ordinal	Interval	
Scores	Logits	Scale
6	−4.38	6.00
7	−3.59	9.45
8	−3.01	11.97
9	−2.59	13.81
10	−2.26	15.26
11	−1.99	16.42
12	−1.77	17.38
13	−1.58	18.22
14	−1.41	18.98
15	−1.25	19.68
16	−1.09	20.35
17	−0.95	20.99
18	−0.81	21.61
19	−0.67	22.22
20	−0.53	22.82
21	−0.39	23.41
22	−0.26	24.01
23	−0.12	24.61
24	0.02	25.21
25	0.16	25.82
26	0.30	26.44
27	0.44	27.07
28	0.59	27.71
29	0.74	28.35
30	0.89	29.01
31	1.04	29.68
32	1.20	30.37
33	1.36	31.06
34	1.52	31.77
35	1.69	32.50
36	1.86	33.26
37	2.04	34.07
38	2.25	34.96
39	2.48	36.00
40	2.78	37.30
41	3.22	39.21
42	3.86	42.00

This conversion table can only be used for respondents with no missing data

were evidenced in the overall poor model fit, unacceptable levels of reliability (0.59), large overall item standard deviation ($SD = 3.37$), significant individual item misfit, disordered thresholds in three items, DIF issues by age and sample, and lastly, local dependency in items 1 and 5, as well as between items 2 and 4. To remedy the host of scale issues, a super-item strategy was used to target the local dependency at the heart of the scale dysfunction while avoiding the removal of items on an already shortened scale (Lundgren-Nilsson et al., 2013; Medvedev et al., 2020). Removal of further items reduces construct validity and conceptual meaning, degrading the quality of results obtained; hence, we aimed to only use item removal as a last resort (Finaulahi et al., 2021; Hopkins et al., 2021; Medvedev et al., 2018a, 2018b).

Two super-items were created by combining items 1, 5, and 6 (super-item 1) and items 2, 3, and 4 (super-item 2) using methodology that focuses on retaining conceptual meaning (Lundgren-Nilsson et al., 2013; Medvedev et al., 2018a, 2018b; Sandham et al., 2019). This modification resulted in a non-significant item-trait interaction chi-square value of 20.69 ($p = 0.191$), indicating that the measurement properties of individual items were consistent across different trait levels. Reliability also improved markedly with a PSI of 0.81, indicating the appropriateness of the modified GQ-6 for assessment at both individual and group levels (Tennant & Conaghan, 2007). Although not directly comparable, our reliability results are consistent with the findings of a recent meta-analysis of various gratitude measures, which reported a Cronbach's alpha score of 0.81 for the GQ-6 (Card, 2019).

After minor modifications, disordered thresholds also showed improvement with categories no longer working inappropriately, as illustrated in Fig. 2. Both super-items demonstrated good model fit, and no significant DIF issues were found by either age or sample. Sample targeting of the scale via the person-item threshold distribution after modification demonstrated that item thresholds cover over 95% of sample abilities with no significant floor or ceiling effects. These results differ from previous findings by Chen et al. (2009) in their report of a ceiling effect in their GQ-6 data supporting benefits of Rasch analysis and application of super-item strategy in the current study. The distribution of persons is slightly skewed to the right, with a person mean of 0.93 indicating a high overall level of gratitude of the current sample. Modifications amended several issues without removing items, keeping the original scale and its raw scoring method intact. Super-item creation is an effective method for rectifying scale dysfunction as items showing local dependency share a common error variance unrelated to trait gratitude. This common error variance caused unwanted spurious correlations which in turn impact negatively on reliability. Due to improvements made by scale modification, we were able to create a conversion algorithm to assist researchers and clinicians in converting their data to an interval level without changes to the original scale or its administration format. However, this still requires reverse coding of items 3 and 6, and respondents with missing data cannot be converted.

Super-item methodology afforded a major advantage in our investigation as the primary goal of this study was to rectify scale dysfunction without removing items as was done by previous investigations under both CTT (Froh et al., 2011) and Rasch approaches (Chen et al., 2009; Valdez & Chu, 2020). The removal of item 6 is undesirable as it negatively impacts reliability (PSI of 0.67 without item 6), content validity, and overall model fit (larger and significant item-trait chi-square). Additionally, it changes the original scale format which makes implementation and comparisons between findings difficult as many recent studies did not adhere to recommendations to use GQ-5. It is more reasonable to retain all items to gain higher reliability, content validity, overall model fit, and original scale administration.

Conversion of scores using this algorithm resulted in improvements to the precision and usefulness of the GQ-6, most notably in the ability to use scale data in comparison to interval data such as physical measurements involving electroencephalography (EEG), or galvanic skin response (GSR) (Medvedev et al., 2020). Several previous gratitude interventions have used interval physiological measures such as blood pressure, cortisol (Jackowska et al., 2016), forced expiratory volume (Cook et al., 2018), glycaemic control (Schache et al., 2019b), heart rate variability (Jackowska et al., 2016; Redwine et al., 2016), and biomarkers of inflammation (Moieni et al., 2019). These investigations would have benefited from an interval level gratitude measure for comparisons, analysis via parametric statistics, and finer distinctions between trait levels aside from only knowing who has higher or lower levels—an Achilles' heel of ordinal scales. Mindfulness is also frequently compared with interval-level health data, and several mindfulness scales have already been converted by a Rasch approach (Medvedev et al., 2016, 2018c, 2020). Enhancing the measurement accuracy of a short and therefore clinically useable scale such as the GQ-6 allows future research to explore the relationship between mindfulness and gratitude as they relate to health benefits.

The current study represents an advancement to research on gratitude interventions as a means of improving physical and mental health outcomes, as ordinal scales such as the GQ-6 are inherently low resolution. That is, we may be able to tell that person A is higher in gratitude than person B, but we cannot tell by how much. Therefore, resolving the lack of precision attributable to ordinal measures aligns a scale with the fundamental principles of measurement (specifically concatenability and invariance), satisfying the assumptions of parametric tests (Rasch, 1960; Stevens, 1946) and facilitating comparisons of latent traits to relevant linear measures, i.e. heart rate, skin conductance, blood pressure, respiratory rate, and salivary cortisol (Allen & Yen, 2002; Stucki et al., 1996). Improvements such as this mark an important step for mindfulness research in beginning a deeper line of inquiry into the foundational relationship between mindfulness and gratitude, in addition to how both of these factors come together to influence physical health.

Achieving synergistic benefits by combining gratitude and mindfulness approaches holds theoretical promise. Several investigations have shown mindfulness to be effective in reducing pain in clinical populations, and some evidence for its influence on neuro-immune markers of inflammation (Creswell et al., 2019). Gratitude has also accumulated a rapid evidence base indicating its importance to both physical health in general and patients in particular (Bartlett & Arpin, 2019; Heckendorf et al., 2019; Moieni et al., 2019; Schnitker & Richardson, 2019). The primary mechanism by which gratitude is thought to influence physical health outcomes is in reducing stress and increasing behavioural and biological restorative pathways (See Boehm and Kubzansky's (2012) model of positive psychological well-being adapted in Schache et al., 2019a). Synergistic benefits may be particularly relevant in chronic health interventions where both mindfulness and gratitude practices have been frequently deployed. By enhancing measurement accuracy, deeper investigations into the fundamental link between gratitude and mindfulness are possible—in the hopes that further benefits to human well-being can be uncovered.

Limitations and Future Directions

There are several limitations to the current study. Firstly, the Rasch transformation scores were not tested with an independent sample. Future work can be done to prove the robustness of the method. Furthermore, this work was conducted with a non-clinical sample and therefore is not generalizable to those with chronic illnesses referenced in this study, i.e. HIV, fibromyalgia, arthritis, COPD. The samples taken were reached through computer and mobile devices that required internet or cell service and therefore may not be representative of those with lower socio-economic status who were not accessible in this way. Future research should look to expand to a multi-national sample beyond the USA and India, as well as to clinical populations for replication of these findings, testing for item bias by illness, ethnicity, and nationality. Using incentives for the US sample while offering no incentives for the Indian sample represents a limitation of this study. However, given the low value of incentive (80c USD), it is unlikely that differences in incentive could have a significant impact on our results.

Author Contribution EB: planned the study, collected the data, conducted the statistical analyses, and wrote the manuscript; NI: collaborated in the data collection and the editing of the manuscript; ONM: conducted the statistical analyses and collaborated with the writing and

editing of the manuscript. All of the authors approved the final version of the manuscript for submission.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability Data will be stored for the purpose of verification for 10 years on a secure cloud drive facilitated by the University of Auckland and is available from the lead author upon request.

We declare that this manuscript is original and has not been previously published or submitted elsewhere. As the sole authors, we report no financial relationships or funding sources.

Declarations

Ethics Approval The study was approved by the New Zealand Ethics Committee on May 12th 2021 Ref# 2021_22, which follows internationally recognized ethical principles.

Informed Consent All participants involved in this study provided their informed consent.

Conflict of Interest The authors declare no competing interests.

References

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Waveland Press.

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38*(1), 123–140. https://doi.org/10.1007/bf02291180

Andrich, D. A. (1978). Rating formulation for ordered response categories. *Psychometrika, 43*, 561–573. https://doi.org/10.1007/BF02293814

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

Bai, C. F., Cui, N. X., Xu, X., Mi, G. L., Sun, J. W., Shao, D., Li, J., Jiang, Y. Z., Yang, Q. Q., Zhang, X., & Cao, F. L. (2019). Effectiveness of two guided self-administered interventions for psychological distress among women with infertility: A three-armed, randomized controlled trial. *Human Reproduction, 34*(7), 1235–1248. https://doi.org/10.1093/humrep/dez066

Balalla, S., Krägeloh, C. U., Medvedev, O. N., & Siegert, R. J. (2020). Is the Rivermead post-concussion symptoms questionnaire a reliable and valid measure to assess long-term symptoms in traumatic brain injury and orthopedic injury patients? A novel investigation using Rasch analysis. *Neurotrauma Reports, 1*(1), 63–72.

Bartlett, M. Y., & Arpin, S. N. (2019). Gratitude and loneliness: Enhancing health and well-being in older adults. *Research on Aging, 41*(8), 772–793. https://doi.org/10.1177/0164027519845354

Boehm, J. K., & Kubzansky, L. D. (2012). The heart's content: The association between positive psychological well-being and cardiovascular health. *Psychological Bulletin, 138*, 655–691. https://doi.org/10.1037/a0027448

Card, N. A. (2019). Meta-analyses of the reliabilities of four measures of gratitude. *The Journal of Positive Psychology, 14*(5), 576–586. https://doi.org/10.1080/17439760.2018.1497690

Chen, L. H., Chen, M. Y., Kee, Y. H., & Tsai, Y. M. (2009). Validation of the Gratitude Questionnaire (GQ) in Taiwanese undergraduate students. *Journal of Happiness Studies, 10*, 655–664. https://doi.org/10.1007/s10902-008-9112-7

Chen, L. H., Wu, C. H., & Chang, J. H. (2017). Gratitude and athletes' life satisfaction: The moderating role of mindfulness. *Journal of Happiness Studies, 18*(4), 1147–1159.

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

Cook, K. A., Woessner, K. M., & White, A. A. (2018). Happy asthma: Improved asthma control with a gratitude journal. *The Journal of Allergy and Clinical Immunology, 6*(6), 2154–2156. https://doi.org/10.1016/j.jaip.2018.04.021

Creswell, J. D., Lindsay, E. K., Villalba, D. K., & Chin, B. (2019). Mindfulness training and physical health: Mechanisms and outcomes. *Psychosomatic Medicine, 81*(3), 224–232. https://doi.org/10.1097/PSY.0000000000000675

Digdon, N., & Koble, A. (2011). Effects of constructive worry, imagery distraction, and gratitude interventions on sleep quality: A pilot trial. *Applied Psychology, 3*(2), 193–206. https://doi.org/10.1111/j.1758-0854.2011.01049.x

Eaton, R., Bradley, G., & Morrissey, S. (2013). Positive predispositions, quality of life and chronic illness. *Psychology, Health & Medicine, 19*(4), 473–489. https://doi.org/10.1080/13548506.2013.824593

Emmons, R. A., & McCullough, M. E. (2003). Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology, 84*(2), 377–389. https://doi.org/10.1037/0022-3514.84.2.377

Emmons, R. A. (2004). The psychology of gratitude. An introduction. In R. A. Emmons & M. E. McCullough (Eds.), *The psychology of gratitude* (pp. 3–16). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195150100.001.0001

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior, 26*(2), 132–139. https://doi.org/10.1016/j.chb.2009.10.015

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

Froh, J., Fan, J., Emmons, R., Bono, G., Huebner, E., & Watkins, P. (2011). Measuring gratitude in youth. *Psychological Assessment, 23*(2), 311–324. https://doi.org/10.1037/a0021590

Garg, N., Katiyar, N., & Mehak,. (2021). Gratitude questionnaire (GQ-6): Exploring psychometric properties in India. *Journal of Religion and Health, 60*(5), 3716–3731. https://doi.org/10.1007/s10943-021-01419-y

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.

Heckendorf, H., Lehr, D., Ebert, D. D., & Freund, H. (2019). Efficacy of an internet and app-based gratitude intervention in reducing repetitive negative thinking and mechanisms of change in the interventions effect on anxiety and depression: Results from a

randomized controlled trial. *Behaviour Research and Therapy, 119*, 103415. https://doi.org/10.1016/j.brat.2019.103415
Hopkins, E. G., Lyndon, M. P., Henning, M. A., & Medvedev, O. N. (2021). Applying Rasch analysis to evaluate and enhance the Academic Motivation Scale. *Australian Journal of Psychology, 73*(3), 348–356. https://doi.org/10.1080/00049530.2021.1904794
Huffman, J. C., Beale, E. E., Celano, C. M., Beach, S. R., Belcher, A. M., Moore, S. V., & Januzzi, J. L. (2016). Effects of optimism and gratitude on physical activity, biomarkers, and readmissions after an acute coronary syndrome. *Circulation: Cardiovascular Quality and Outcomes, 9*(1), 55–63. https://doi.org/10.1161/circoutcomes.115.002184
Ivtzan, I., Young, T., Martman, J., Jeffrey, A., Lomas, T., Hart, R., & Eiroa-Orosa, F. J. (2016). Integrating mindfulness into positive psychology: A randomised controlled trial of an online positive mindfulness program. *Mindfulness, 7*(6), 1396–1407.
Jackowska, M., Brown, J., Ronaldson, A., & Steptoe, A. (2016). The impact of a brief gratitude intervention on subjective well-being, biology and sleep. *Journal of Health Psychology, 21*(10), 2207–2217. https://doi.org/10.1177/1359105315572455
Jankowski, P. J., & Sandage, S. J. (2014). Meditative prayer and intercultural competence: Empirical test of a differentiation-based model. *Mindfulness, 5*(4), 360–372.
Kashdan, T. B., Mishra, A., Breen, W. E., & Froh, J. J. (2009). Gender differences in gratitude: Examining appraisals, narratives, the willingness to express emotions, and changes in psychological needs. *Journal of Personality, 77*, 691–730. https://doi.org/10.1111/j.1467-6494.2009.00562.x
Kreiner, S., & Christensen, K. B. (2007). Validity and objectivity in health-related scales: Analysis by graphical loglinear Rasch models. In von Davier, M., & Carstensen, C. H. (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 329–346). Springer-Verlag. https://doi.org/10.1007/978-0-387-49839-3_21
Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7*(4), 328.
Lundgren-Nilsson, Å., Jonsdottir, I. H., Ahlborg, G., Jr., & 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*(1), 1–9. 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
McCullough, M. E., Emmons, R. A., & Tsang, J. A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology, 82*(1), 112–127. https://doi.org/10.1037/0022-3514.82.1.112
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., Krägeloh, C. U., Titkova, E. A., & Siegert, R. J. (2018). Rasch analysis and ordinal-to-interval conversion tables for the depression, anxiety and stress scale. *Journal of Health Psychology, 25*(10–11), 1374–1383. https://doi.org/10.1177/1359105318755261
Medvedev, O. N., Titkova, E. A., Siegert, R. J., Hwang, Y. S., & Krägeloh, C. U. (2018). Evaluating short versions of the Five Facet Mindfulness Questionnaire using Rasch analysis. *Mindfulness, 9*(5), 1411–1422. https://doi.org/10.1007/s12671-017-0881-0
Medvedev, O. N., Turner-Stokes, L., Ashford, S., & Siegert, R. J. (2018c). 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/10.2340/16501977-2324
Medvedev, O. N., Pratscher, S. D., & Bettencourt, A. (2020). Psychometric evaluation of the Interpersonal Mindfulness Scale using Rasch analysis. *Mindfulness, 11*(8), 2007–2015. https://doi.org/10.1007/s12671-020-01415-5
Mills, P. J., Redwine, L. S., Wilson, K., Pung, M. A., Chinh, K., Greenberg, B. H., & Chopra, D. (2015). The role of gratitude in spiritual well-being in asymptomatic heart failure patients. *Spirituality in Clinical Practice, 2*(1), 5–17. https://doi.org/10.1037/scp0000050
Moieni, M., Irwin, M. R., Haltom, K., Jevtic, I., Meyer, M. L., Breen, E. C., Cole, S. W., & Eisenberger, N. I. (2019). Exploring the role of gratitude and support-giving on inflammatory outcomes. *Emotion, 19*(6), 939–949. https://doi.org/10.1037/emo0000472
Moskowitz, J. T., Carrico, A. W., Duncan, L. G., Cohn, M. A., Cheung, E. O., Batchelder, A., & Folkman, S. (2017). Randomized controlled trial of a positive affect intervention for people newly diagnosed with HIV. *Journal of Consulting and Clinical Psychology, 85*(5), 409. https://doi.org/10.1037/ccp0000188
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.
Redwine, L. S., Henry, B. L., Pung, M. A., Wilson, K., Chinh, K., Knight, B., & Mills, P. J. (2016). Pilot randomized study of a gratitude journaling intervention on heart rate variability and inflammatory biomarkers in patients with stage B heart failure. *Psychosomatic Medicine, 78*(6), 667–676. https://doi.org/10.1097/psy.0000000000000316
Reeves, M., Vogt, F., & Marks, E. (2021). Dispositional mindfulness, gratitude and self-compassion: Factors affecting tinnitus distress. *Mindfulness, 12*(4), 1002–1008. https://doi.org/10.1007/s12671-020-01569-2
Rosenzweig, D. (2013). The sisters of mindfulness. *Journal of Clinical Psychology, 69*, 793–804. https://doi.org/10.1002/jclp.22015
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
Schache, K., Consedine, N., Hofman, P., & Serlachius, A. (2019a). Gratitude - more than just a platitude? The science behind gratitude and health. *British Journal of Health Psychology, 24*(1), 1–9. https://doi.org/10.1111/bjhp.12348
Schache, K. R., Hofman, P. L., & Serlachius, A. S. (2019b). A pilot randomized controlled trial of a gratitude intervention for adolescents with type 1 diabetes. *Diabetic Medicine, 37*(8), 1352–1356. https://doi.org/10.1111/dme.14078
Schnitker, S. A., & Richardson, K. L. (2019). Framing gratitude journaling as prayer amplifies its hedonic and eudaimonic well-being, but not health, benefits. *The Journal of Positive Psychology, 14*(4), 427–439. https://doi.org/10.1080/17439760.2018.1460690
Schutte, N. S., Keng, S. L., & Cheung, M. W. L. (2021). Emotional intelligence mediates the connection between mindfulness and gratitude: A meta-analytic structural equation modelling study. *Mindfulness, 12*(11), 2613–2623. https://doi.org/10.1007/s12671-021-01725-2
Seear, K. H., & Vella-Brodrick, D. A. (2013). Efficacy of positive psychology interventions to increase well-being: Examining the role of dispositional mindfulness. *Social Indicators Research, 114*, 1125–1141. https://doi.org/10.1007/s11205-012-0193-7
Sirois, F. M., & Wood, A. M. (2017). Gratitude uniquely predicts lower depression in chronic illness populations: A longitudinal study of inflammatory bowel disease and arthritis. *Health Psychology, 36*(2), 122–132. https://doi.org/10.1037/hea0000436
Smith Jr. E. V. (2002). Understanding Rasch measurement: Detecting and evaluating the impact of multi-dimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205–231. https://pubmed.ncbi.nlm.nih.gov/12011501/

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677–680. https://doi.org/10.1126/science.103.2684.677

Stucki, G., Daltroy, L., Katz, J., Johannesson, M., & Liang, M. (1996). Interpretation of change scores in ordinal clinical scales and health status measures: The whole may not equal the sum of the parts. *Journal of Clinical Epidemiology, 49*(7), 711–717. https://doi.org/10.1016/0895-4356(96)00016-9

Swickert, R., Bailey, E., Hittner, J., Spector, A., Benson-Townsend, B., & Silver, N. C. (2019). The mediational roles of gratitude and perceived support in explaining the relationship between mindfulness and mood. *Journal of Happiness Studies, 20*(3), 815–828. https://doi.org/10.1007/s10902-017-9952-0

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.23108

Tennant, A., & Pallant, J. (2006). Unidimensionality matters. Rasch Measurement. *Transactions, 20*, 1048–1051.

Toussaint, L., Sirois, F., Hirsch, J., Weber, A., Vajda, C., Schelling, J., Kohls, N., & Offenbacher, M. (2017). Gratitude mediates quality of life differences between fibromyalgia patients and healthy controls. *Quality of Life Research, 26*(9), 2449–2457. https://doi.org/10.1007/s11136-017-1604-7

ValdezChu, J. P. M. S. K. W. (2020). Examining the psychometric validity of the five-item gratitude questionnaire: An item response theory approach. *Journal of Psychoeducational Assessment, 38*(4), 529–536. https://doi.org/10.1177/0734282918816542

Voci, A., Veneziani, C. A., & Fuochi, G. (2019). Relating mindfulness, heartfulness, and psychological well-being: The role of self-compassion and gratitude. *Mindfulness, 10*(2), 339–351. https://doi.org/10.1007/s12671-018-0978-0

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

Wolfe, W. L., & Patterson, K. (2017). Comparison of a gratitude-based and cognitive restructuring intervention for body dissatisfaction and dysfunctional eating behavior in college women. *Eating Disorders, 25*(4), 330–344. https://doi.org/10.1080/10640266.2017.1279908

Wood, A. M., Joseph, S., & Maltby, J. (2009). Gratitude predicts psychological well-being above the Big Five facets. *Personality and Individual Differences, 46*, 443–447. https://doi.org/10.1016/j.paid.2008.11.012

Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

Yang, Y., Kong, X., Guo, Z., & Kou, Y. (2021). Can self-compassion promote gratitude and prosocial behavior in adolescents? A 3-year longitudinal study from China. *Mindfulness, 12*(6), 1377–1386. https://doi.org/10.1007/s12671-021-01605-9