



# Cross-Cultural Adaptation and Application of the One-Parameter Item Response Model to the Santa Clara Brief Compassion Scale (SCBCS)

Peter Adu<sup>1,7</sup> · Tosin Popoola<sup>2</sup> · Naved Iqbal<sup>3</sup> · Anja Roemer<sup>4</sup> · Sunny Collings<sup>1</sup> · Clive Aspin<sup>1</sup> · Oleg N. Medvedev<sup>5</sup> · Colin R. Simpson<sup>1,6</sup>

Accepted: 12 August 2025  
© The Author(s) 2025

## Abstract

**Objectives** International research has consistently demonstrated the positive impact of compassion towards others on both physical and mental well-being, with significant implications for mindfulness practice. Based on this evidence, we aimed to adapt the Santa Clara Brief Compassion Scale (SCBCS) into German while simultaneously conducting a cross-cultural validation and enhancing its measurement precision using Rasch methodology across samples from Germany, Ghana, India, and New Zealand.

**Method** We applied the unrestricted Partial Credit Model to analyze data from a randomly selected subsample of 500 participants, drawn from a total convenience sample of 1822 individuals recruited from the general populations of Germany, Ghana, India, and New Zealand.

**Results** Our initial analysis of the SCBCS showed significant misfit to the Rasch model ( $\chi^2(30) = 58.48, p < 0.001$ ), which was successfully addressed by testlet creation resulting in satisfactory model fit ( $\chi^2(24) = 24.80, p = 0.09$ ). This included strict unidimensionality, strong reliability (Person Separation Index = 0.81), and invariance across personal factors, such as country, educational levels, sex, and age. We then developed an algorithm for transforming ordinal scores to interval-level data to enhance the accuracy of the SCBCS. The scale demonstrated sound divergent and convergent validity.

**Conclusions** Our study has validated both the German and English versions of the SCBCS using Rasch methodology. The precision of measuring compassion towards others using the two versions of the SCBCS can be further enhanced by applying the ordinal-to-interval transformation tables developed in this paper.

**Keywords** Compassion · Rasch methodology · Validation · Reliability · Psychometrics

✉ Peter Adu  
adupete@staff.vuw.ac.nz

Tosin Popoola  
Tosin.Popoola@newcastle.edu.au

Naved Iqbal  
niqbal@jmi.ac.in

Anja Roemer  
A.Roemer@massey.ac.nz

Sunny Collings  
sunny.collings@vuw.ac.nz

Clive Aspin  
clive.aspin@vuw.ac.nz

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

Colin R. Simpson  
colin.simpson@vuw.ac.nz

- 1 School of Health, Wellington Faculty of Health, Victoria University of Wellington, P. O. Box 600, Wellington 6140, New Zealand
- 2 College of Health, Medicine and Wellbeing-School of Nursing and Midwifery, The University of Newcastle, Callaghan, Australia
- 3 Jamia Millia Islamia, New Delhi, India
- 4 School of Psychology, Massey University, Palmerston North, New Zealand
- 5 School of Psychological and Social Sciences, University of Waikato, Hamilton, New Zealand
- 6 Usher Institute, The University of Edinburgh, Edinburgh, UK
- 7 Division of Social and Cultural Psychiatry, Department of Psychiatry, McGill University, Montréal, Canada

Previous research on compassion has shown its positive impact on physical and mental well-being (Kim et al., 2020; Medvedev et al., 2021). Compassion involves an attitude aiming at alleviating suffering and promoting well-being (Perez-Bret et al., 2016). Generally, compassion encompasses five key elements: recognition of suffering, understanding its universality, feeling sympathy and empathy, tolerating the distress associated with witnessing suffering, and the willingness to take action to alleviate suffering (Gilbert, 2020). Compassionate acts can be directed towards oneself or others, and both types have consistently shown a positive impact on health outcomes (MacBeth & Gumley, 2012). The current study focused on compassion towards others. Compassion toward others has emerged as an essential variable that promotes mental and physical well-being, with implications for mindfulness research, both as an outcome and as a contributor to the effectiveness of mindfulness practices. For instance, Medvedev et al. (2021) have reported that compassion toward others was positively associated with the observing and describing facets of mindfulness. This aligns with earlier findings by Jazaieri et al. (2014), which indicated that compassion toward others can enhance mindfulness, resulting in the reduction of negative emotional experiences. These studies demonstrated that compassion toward others is not merely an outcome of mindfulness practice but also a reinforcing factor that could facilitate mindfulness. Boellinghaus et al. (2014) also demonstrated that mindfulness training can foster compassion toward others, further highlighting a bidirectional link between the two constructs. These dynamic interplays have prompted a growing number of mindfulness programs to integrate explicit compassion training (i.e., either toward others or self) into their frameworks (Huppert, 2017). Ultimately, compassion enhances mindfulness by reducing negative emotions triggered by the competitive nature of our environments, leading to emotional stability and kindness towards others. Taken together, these findings underscore the importance of compassion for mindfulness research, both for its theoretical significance and for its practical implications in enhancing emotional well-being, interpersonal connection, and the broader impact of mindfulness-based interventions.

Additionally, the interaction between compassion towards others and social support was found to act as a buffer against physiological reactivity to stress (Cosley et al., 2010). Compassion towards others was also related to lower levels of loneliness (Lee et al., 2021). Likewise, Mongrain et al. (2011) have demonstrated that practicing compassion towards others leads to sustained happiness, enhanced self-esteem, and decreased levels of depressive symptoms. During the COVID-19 pandemic, compassion towards others was identified as an important factor in promoting resilience against the harmful effects of the pandemic on mental health and social well-being across 21 countries

(Matos et al., 2022). Individuals who exhibited higher levels of compassion towards others were more likely to adhere to COVID-19 prevention protocols, including a willingness to stay home to protect others (Karnaze et al., 2022). So far, the findings have highlighted the substantial benefits of compassion towards others for well-being research, including mindfulness literature. Hence, adapting a scale that measures compassion towards others into different languages and validating it across multiple countries spanning continents, while assessing its measurement precision using the Rasch method, as the focus of this study, is essential for advancing research on psychological health.

One of the few scales for measuring compassion towards others is the 5-item Santa Clara Brief Compassion Scale (SCBCS; Hwang et al., 2008), demonstrating sound validity and reliability (Cronbach's alpha of 0.90) among college students. The SCBCS is a short version of the Compassionate Love Scale (CLS; Sprecher & Fehr, 2005). This self-reported scale captures the extent to which individuals perceive themselves as compassionate toward others in their thoughts, feelings, and behaviors. Ever since the development of the SCBCS, validation studies using the traditional Classical Test Theory (CTT) methods, such as the confirmatory factor analysis (CFA), have provided evidence supporting its psychometric properties across different countries. For instance, the Peruvian version of the SCBCS was found to be a valid and reliable instrument for measuring compassion towards others, further exhibiting invariance among men and women (Caycho-Rodríguez et al., 2022). These results were similar to the Farsi (Hoseininezhad et al., 2023) and the Czech versions (Novak et al., 2021) of the SCBCS.

Nevertheless, CFA evidence has been criticized as assuming that observed scores encompass true scores and measurement error without accounting for this error, resulting in an inaccurate conclusion about the true scores of a latent construct (Demars, 2013; Prudon, 2015). Furthermore, CFA results are highly contingent on the sample used. This means that estimating parameters may excessively reflect the unique characteristics of a specific sample rather than precisely capturing the underlying structure of the broader population (Goudarzian, 2023). For instance, it is not uncommon to achieve high reliability coefficients for bigger samples (López et al., 2015; Prudon, 2015), as reliability under this method only reflects consistency of scores. Similarly, CFA models cannot accurately differentiate between the actual constructs being measured and extraneous factors, making it susceptible to misleading correlations, which raises concerns about the dimensionality of a scale (Trninić et al., 2013). CFA also does not account for the differential contribution of individual items to the overall scale score if all items contribute equally. For instance, in a scale measuring depression, items related to suicidal thoughts may be more important than those related to sleep patterns. Additionally,

CFA does not consider that differences between response categories may not be uniform within and between items (Hays et al., 2000). Such challenges associated with the CTT methods may limit scale generalizability across diverse populations.

Rasch methodology is essentially similar to the one-parameter Item Response Theory (IRT) model and provides a robust approach to assessing measurement statistics of a scale compared to the CTT methods (Demars, 2013; Rasch, 1993). For example, the Rasch model accounts for random error to provide a more accurate estimate of the true scores and consequently a more precise measurement of the latent trait (Demars, 2013). Rasch methodology assumes that the probability of a person endorsing an item response is modeled as a logistic function of the person's ability and the item's difficulty (Medvedev & Krägeloh, 2022; Tennant & Küçükdeveci, 2023). The Rasch model approach to psychometric assessment involves estimating both item difficulty and person ability, using specific models for both polytomous or dichotomous data (Masters, 1982; Rasch, 1993). The Rasch model aligns closely with the fundamental principles of measurement, including the invariance of item parameters across respondent groups and the consistent interval structure of scale units (Thurstone, 1931). One notable aspect of the Rasch model is its capability to convert ordinal responses into interval-level data, so long as the data adhere closely to the Rasch model's strict requirements. This transformation offers several advantages, including increased reliability, better sample targeting, reduced measurement error, and suitability for parametric analyses (Medvedev & Krägeloh, 2022).

Despite the benefits of Rasch methodology for enhancing measurement, no study has been identified that applied Rasch analysis to the SCBCS. We have also observed the absence of a German version of this scale in the literature. Hence, our study was guided by two primary aims. First, we aimed to adapt the SCBCS into the German language by ensuring linguistic and conceptual equivalence. Second, we conducted a cross-cultural psychometric evaluation of the SCBCS across samples from Ghana, Germany, India, and New Zealand, using Rasch analysis to improve its measurement precision. As part of the cross-cultural validation, we focused on assessing the scale's reliability, structural validity, convergent validity, and divergent validity. We also investigated measurement invariance of the SCBCS across key sociodemographic groups, including country of residence, age, educational level, employment status, marital status, and sex. We expected a positive correlation between the scores of the SCBCS and related constructs such as positive affect, religiosity, self-compassion, and dispositional optimism. This assessment aimed to establish the convergent validity of the SCBCS. Existing literature has consistently documented a positive association between

these constructs and compassion towards others (Chan et al., 2022; López et al., 2018). However, we expected a weak to zero non-significant correlation between the scores of the SCBCS and unrelated measures of psychological distress and negative affect (Chan et al., 2022; Medvedev et al., 2021). This underscores the SCBCS's capacity to differentiate from theoretically uncorrelated constructs, evaluating the divergent validity of the SCBCS.

## Method

### Participants

The current study used data from a total sample of 1822 recruited from Germany (475), Ghana (523), India (411), and New Zealand (413) for the Rasch analysis. Initially, we utilized Microsoft Excel to randomly select 1000 participants (250 from each country) from the total sample to ensure equal representativeness from each country. The Rasch Unidimensional Measurement Model (RUMM; Andrich et al., 2009) was used to randomly select the final 500 from our sample for the Rasch analysis. The selection of the sample size was based on the suggested sample size requirement for Rasch analysis especially when using the RUMM (i.e., 250 to 500; Hagell & Westergren, 2016). The final study sample was balanced between the benefits of larger sample sizes and the need to address the over-sensitivity of the chi-square test. The chi-square test has the tendency to magnify statistical significance with larger sample sizes (e.g., > 500), even in situations where the practical impact of data misfit is insignificant (Pelton, 2002). The current data were gathered in mid-2022 and form part of a larger dataset on the interplay between psychological variables and vaccination attitudes used in a conceptually and methodologically different study (e.g., Adu et al., 2023). The age of participants ranged from 18 to 80 years in India ( $M_{age} = 26.14$ ;  $SD = 8.57$ ), 18 to 89 years in New Zealand ( $M_{age} = 46.35$ ;  $SD = 18.07$ ), 18 to 63 years in Ghana ( $M_{age} = 29.48$ ;  $SD = 5.69$ ), and 18 to 87 years in Germany ( $M_{age} = 44.09$ ;  $SD = 5.57$ ). Participants differed significantly in terms of level of education ( $\chi^2(66) = 405.39$ ,  $p < 0.001$ ), sex ( $\chi^2(27) = 91.94$ ,  $p < 0.001$ ), age ( $\chi^2(6) = 364.15$ ,  $p < 0.001$ ), marital status ( $\chi^2(63) = 117.74$ ,  $p < 0.001$ ), and employment status ( $\chi^2(63) = 164.96$ ,  $p < 0.001$ ) across the countries.

### Procedure

Data collection in Ghana and India utilized various online platforms, including Facebook, WhatsApp, Twitter, Instagram, and Email. This cross-sectional online survey employed convenience sampling via a snowballing

technique, offering no incentives to participants from Ghana and India. For data collection in New Zealand and Germany, a data collection company “Qualtrics” facilitated the process, and participants received a minor monetary compensation (e.g., NZ\$5 or a possibility to participate in a draw). The use of online data collection was prompted by COVID-19 preventive protocols like social distance. Initially, participants were requested to provide demographic details before proceeding to the main survey, which typically lasted around 15min. Questionnaires were presented in English for participants in Ghana, India, and New Zealand, while a German version was provided for participants in Germany.

## Measures

### Compassion Towards Others

The Santa Clara Brief Compassion Scale (SCBCS; Hwang et al., 2008) is a 5-item measure scored on a 7-point Likert scale, ranging from 1 = *not at all true of me* to 7 = *very true of me*. Example of an item found on the scale is “I tend to feel compassion for people, even though I do not know them.” The reliabilities of the scale were found to be very good for the total sample (Cronbach’s alpha [ $\alpha$ ] = 0.86, McDonald’s omega [ $\omega$ ] = 0.89, Person Separation Index (PSI) = 0.81;  $M = 24.10$ ,  $SD = 7.00$ ).

### Life Orientation Test

The Life Orientation Test Revised (LOT-R; Glaesmer et al., 2008; Scheier et al., 1994) was used to measure dispositional optimism using a 10-item scale, rated on a 5-point Likert scale ranging from 0 = *strongly disagree* to 4 = *strongly agree*. For instance, a positively worded item on this scale is “In uncertain times, I usually expect the best.” While the scale exhibited relatively low reliability coefficients according to  $\alpha = 0.57$  and  $\omega = 0.58$  ( $M = 16.10$ ,  $SD = 3.62$ ), it demonstrated acceptable reliability when assessed using the Rasch-based  $PSI = 0.73$  (Table 2).

### Self-Compassion

The 12-item Self-Compassion Scale-Short Form (SCS-SF; Raes et al., 2011) was employed to measure self-compassion. This scale uses a 5-point Likert-scale response format: 1 = *almost never* to 5 = *almost always*. An example of a positively worded item on this scale is “I try to be understanding and patient towards those aspects of my personality I don’t like.” The reliabilities of the scale were acceptable ( $\alpha = 0.71$ ,  $\omega = 0.72$ ,  $PSI = 0.71$ ;  $M = 31.40$ ,  $SD = 7.40$ ; Adu et al., 2024).

### Psychological Distress

The 21-item Depression Anxiety Stress Scales (DASS-21; Lovibond & Lovibond, 1995), rated on a 4-point Likert-scale response option, from 0 = *did not apply to me at all* to 3 = *applied to me very much*, was utilized to measure psychological distress. An example of this scale is as follows: depression (“I couldn’t seem to experience any positive feeling at all”), anxiety (“I was aware of dryness in my mouth”), and stress (“I found it hard to wind down”). The DASS-21 demonstrated very good to excellent reliability coefficients ( $\alpha = 0.98$ ,  $\omega = 0.98$ ,  $PSI = 87$ ;  $M = 30.80$ ,  $SD = 18.00$ ; Adu et al., 2025).

### Religiosity

The Centrality of Religiosity Scale (CRS-10; Huber & Huber, 2012) was used to assess religiosity in this study. This scale has been employed in international research contexts (e.g., Adu et al., 2021). The scale consists of 10 self-reported items, such as “To what extent do you believe that God or something divine exists?” Participants responded on a 5-point Likert scale, ranging from 1 = *not at all* to 5 = *very much so*. The internal consistency of the CRS-10 was excellent ( $\alpha = 0.96$ ,  $\omega = 0.96$ ,  $PSI = 93$ ;  $M = 30.24$ ,  $SD = 50.00$ ).

### Positive Affect and Negative Affect

We used the well-known 20-item Positive Affect (PA) and Negative Affect (NA) Scale (PANAS; Watson et al., 1988) to assess positive affect and negative affect. Participants rated each adjective on a 5-point Likert scale, ranging from 1 = *very slightly* to 5 = *extremely*. Examples of adjectives measuring PA include “interested,” “strong,” and “proud,” while NA includes “anger,” “fear,” and “sadness.” The reliability coefficients for the PA subscale were excellent ( $\alpha = 0.91$ ;  $\omega = 0.93$ ,  $PSI = 0.91$ ;  $M = 31.40$ ,  $SD = 7.40$ ). The NA subscale also showed very good to excellent reliability ( $\alpha = 0.89$ ;  $\omega = 0.92$ ,  $PSI = 81$ ;  $M = 23$ ,  $SD = 8.70$ ).

### Instrument Translation

Scales that do not have existing German versions, such as the SCBCS, were adapted into the German language following established guidelines for the validation and adaptation of cross-cultural instruments (Adu et al., 2023; Brislin, 1970). Initially, the SCBCS questionnaire was translated from English to German by a bilingual native speaker. Then, it was back-translated by another bilingual individual with expertise in instrument validation (Brislin, 1970). The resulting German version was carefully evaluated by three independent bilingual health researchers to confirm

its comprehensibility and ensure cross-cultural equivalence with the original version (Borsa et al., 2012). This evaluation process included assessing semantic accuracy, idiomatic expressions, and cultural nuances. The final questionnaire was piloted with a group of students and academic colleagues to verify the appropriateness of its formatting and instructions (see Appendix for the German version of the SCBCS).

## Data Analyses

### Data Preparation

Imported data in the IBM Statistical Package for the Social Sciences (SPSS) version 28 were examined for missing values utilizing the Little (1988) missing completely at random (MCAR) test. Missing data were handled using the expectation maximization algorithm (Dellaert, 2002), as the data were not MCAR ( $p < 0.05$ ). Descriptive statistics and correlational analyses were then conducted using SPSS, while the Rasch analysis was performed using the RUMM (Andrich et al., 2009). The skewness and kurtosis were all within the acceptable range of  $-2$  to  $+2$  (George, 2011).

### Rasch Analysis

An initial likelihood ratio test indicated that the unrestricted partial credit model is the most appropriate model as there were significant differences between response options thresholds across individual items ( $p = 0.001$ ), which violated the assumptions of the rating scale model. Initially, the overall Rasch model fit was evaluated by observing the significance of the item-trait interaction chi-square reflecting Rasch model misfit, aiming for a non-significant result ( $p > 0.05$ ; Tennant & Küçükdeveci, 2023; Wilkinson et al., 2023). Estimated residual values falling within the average range of  $-2.50$  to  $+2.50$  indicate an acceptable item fit, while residual correlations between individual items are expected to be below 0.20 (Bartholomew et al., 2023; Christensen et al., 2017). The presence of unidimensionality is confirmed when  $\leq 5\%$  of cases show potential for measuring multiple dimensions or the lower bound of confidence intervals around such cases overlaps 5% (Hagell, 2015; Smith, 2002).

We examined correct threshold ordering by evaluating *item characteristic curves* (ICCs) for any signs of disordered thresholds (Andrich, 1978). Sample targeting was evaluated by aiming for values between  $+0.50$  and  $-0.50$  for an item mean of 0.00 to avoid floor and ceiling effects (Tennant & Conaghan, 2007). Local item dependencies (i.e., residual correlation matrix above 0.20) were addressed using the advanced testlet methodology (Tennant & Küçükdeveci, 2023; Wilkinson et al., 2023). Reliability was estimated using the PSI, which provides values interpreted akin to those obtained from  $\alpha$  and  $\omega$ . To ensure the robustness of our findings, we triangulated our results by providing reliability estimates using these three different estimators. Invariance reflected by lack of differential item functioning (DIF) was investigated through ANOVA and Bonferroni adjusted pairwise comparison  $t$ -tests to identify any item bias across sociodemographic factors. We used the 33rd and 66th percentiles standard method to create three balanced age categories: 18–35 years (33%), 36–50 years (33%), and 61–89 years (34%) to assess age group invariance (Hagquist & Andrich, 2017; Pratscher et al., 2019). An achievement of optimal Rasch model fit favors the development of the algorithm to transform ordinal to interval-level data (Medvedev & Krägeloh, 2022). Finally, convergent and divergent validity analyses were conducted using Pearson's correlation coefficient.

## Results

### Overall Model Fit, Reliability, and Unidimensionality

Findings from Table 1 (A1) showed an overall Rasch model misfit, indicated by a significant interaction between items and the latent trait ( $\chi^2(30) = 58.48$ ,  $p < 0.001$ ). However, the analysis revealed acceptable unidimensionality, with the lower bound of significant  $t$ -tests ( $-1.5\%$ ) within the 5% cutoff point. The reliability coefficient was also satisfactory (PSI = 0.78), and all individual item fit residuals fell within the range of  $-2.50$  to  $+2.50$ . As the overall Rasch model fit could have been affected by local dependency between items, the residual correlation matrix was examined and indicated residual correlations between four individual items exceeding 0.20. To address this issue, a modification

**Table 1** Overall Rasch model fit statistics for the initial and final analysis of the Santa Clara Brief Compassion Scale (SCBCS;  $n = 500$ )

Analyses (A)	Item fit residual		Person fit residual		Goodness of fit		PSI	Unidimensionality $t$ -test	
	Mean	SD	Mean	SD	$\chi^2$ (df)	$p$		%	Lower bound
A1 Initial	-0.24	0.92	-0.53	1.41	58.48(30)	<0.001	0.78	0.4	-1.5 (yes)
A2 Final	-0.40	2.07	-0.61	1.17	24.80(24)	0.09	0.81	3.0	1.1 (yes)

PSI Person Separation Index without extremes

analysis of testlet creation was applied, combining locally dependent items. Items 1 + 5 and 2 + 4 were combined based on residual correlation exceeding the 0.20 threshold. This modification resulted in the overall satisfactory Rasch model fit ( $\chi^2(24) = 24.80, p = 0.09$ ), and no misfitting items or local item dependency. It also resulted in improved reliability (PSI = 0.81) and strict unidimensionality, demonstrated by a lower bound of significant *t*-tests of 1.1% falling within the 5% cutoff point (Table 1; final). Table 2 provides an overview of individual item fit statistics within the Rasch model, including item locations, fit residuals, and chi-square parameters.

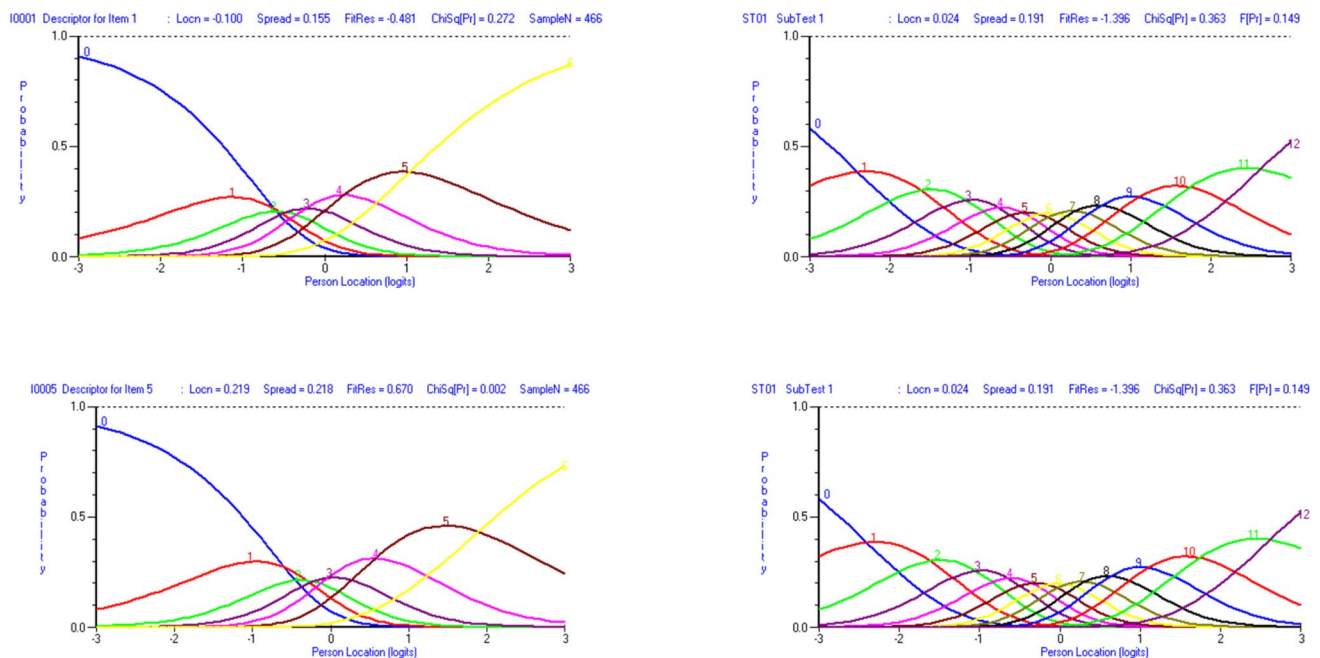
Additionally, the modification resolved the issue of disordered thresholds for Items 1 and 2 (Fig. 1; left), with no response options functioning inappropriately in the testlets (Fig. 1; right), as shown by the monotonic patterns in the ICCs for all items.

### Differential Item Functioning, Person-Item Trait, and Ordinal-to-Interval Conversion

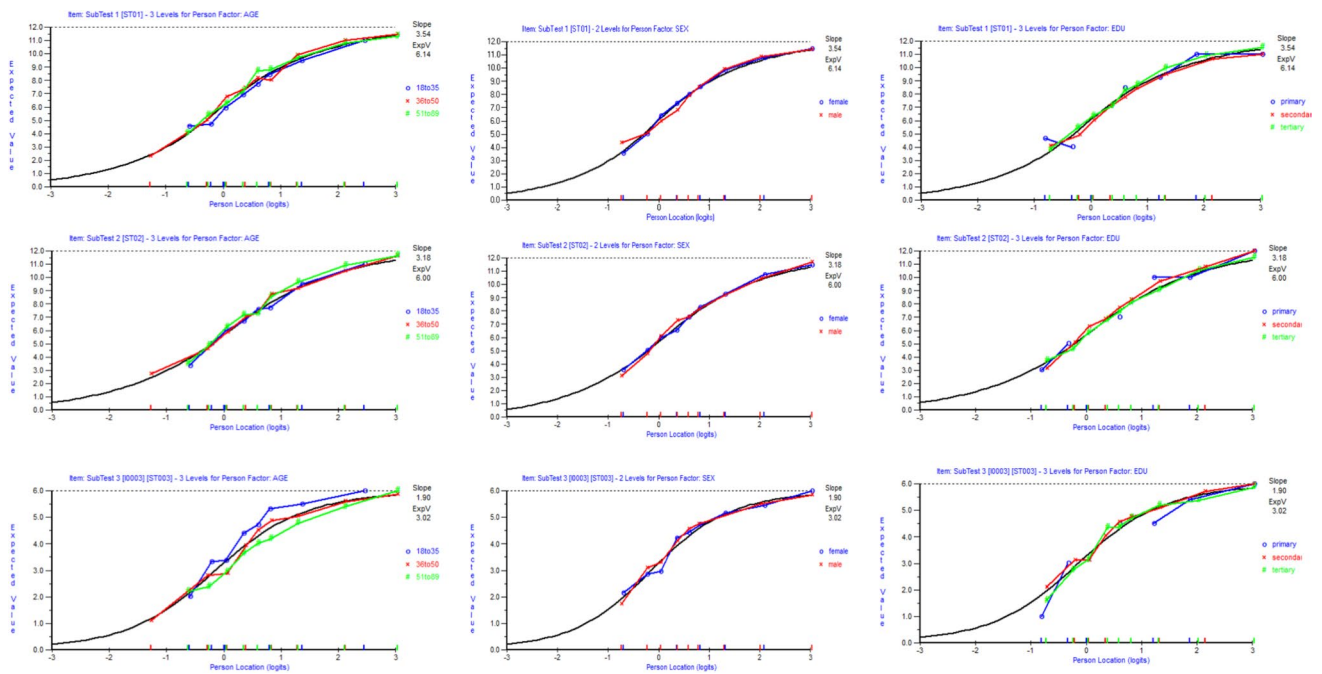
Figures 2 and 3 indicate no DIF across country of origin, age, sex, level of education, employment status, and marital status for all items. The person-item threshold distributions showed no ceiling or floor effects, with satisfactory targeting

**Table 2** Items fit statistics including the initial and final analyses of the Santa Clara Brief Compassion Scale (SCBCS;  $n = 500$ )

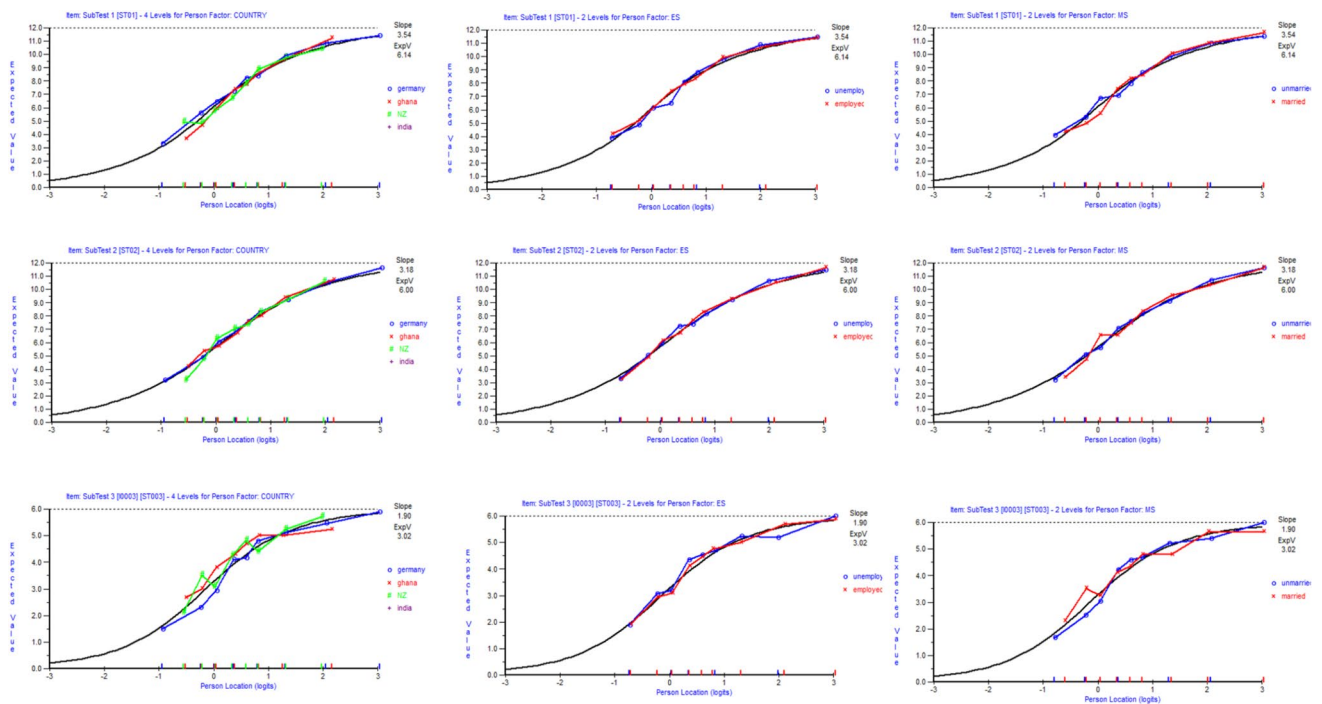
Item no	Initial analysis (5 items)	Location	Fit residual	Chi-square
1	When I hear about someone (a stranger) going through a difficult time, I feel a great deal of compassion for him or her	-0.10	-0.48	7.57
2	I tend to feel compassion for people, even though I do not know them	-0.12	-1.30	15.43
3	One of the activities that provide me with the most meaning to my life is helping others in the world when they need help	-0.24	-0.87	11.60
4	I would rather engage in actions that help others, even though they are strangers, than engage in actions that would help me	0.24	0.76	3.62
5	I often have tender feelings toward people (strangers) when they seem to be in need	0.22	0.67	20.26*
<b>Final analysis: testlets (T)</b>				
T1	Items: 1 + 5	0.02	-1.40	8.76
T2	Items: 4 + 2	0.11	-1.78	8.63
3		-0.14	1.98	8.40



**Fig. 1** Item characteristics curves (ICC) for the final items of the Santa Clara Brief Compassion Scale (SCBCS); disordered ICC for items 1 and 5 (left); ordered testlets (right)



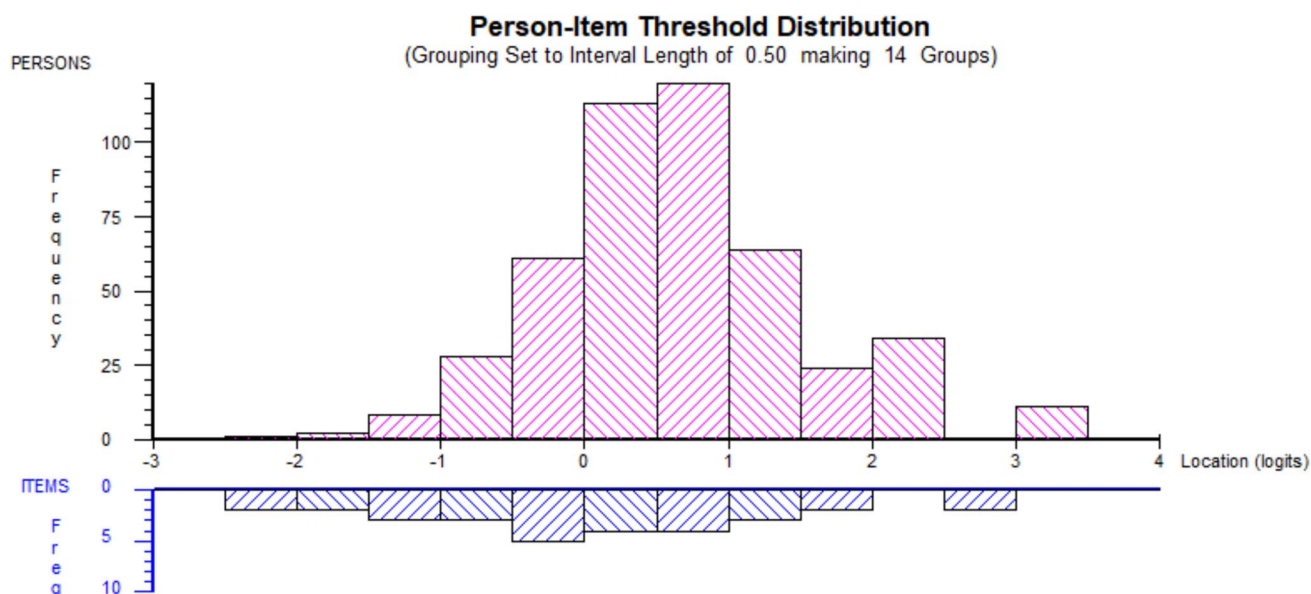
**Fig. 2** Differential Item Functioning (DIF) curves of the Santa Clara Brief Compassion Scale (SCBCS) for age (left), sex (middle), and education (right)



**Fig. 3** Differential Item Functioning (DIF) curves of the Santa Clara Brief Compassion Scale (SCBCS) for country (left), employment status (middle), and marital status (right)

(mean person location = 0.63,  $SD = 0.91$ ; Fig. 4). Following the Rasch model fit indices, an ordinal-to-interval conversion algorithm was developed based on person estimates of the

Rasch model. This algorithm allows for the transformation of ordinal scores into interval-level data (Table 3). We also found a statistically significant difference ( $t(1821) = -24.44$ ,



**Fig. 4** Person-item thresholds distributions for the Santa Clara Brief Compassion Scale (SCBCS)

$p < 0.001$ ) with a large effect size of  $d = 2.56$  between the mean scores of the ordinal scores ( $M = 24.09$ ,  $SD = 7.00$ ) and Rasch-transformed interval scores ( $M = 22.63$ ,  $SD = 5.24$ ) within the same SCBCS range. This provides indirect evidence for enhancement of psychometric statistics when using interval-transformed scale score reflected by lower  $SE$  indicative of reduced measurement error.

### Convergent and Divergent Validity

Pearson's correlation coefficient ( $r$ ) results showed positive associations between the SCBCS scores and measures of dispositional optimism ( $r = 0.10$ ,  $p < 0.001$ ), self-compassion ( $r = 0.07$ ,  $p < 0.001$ ), positive affect ( $r = 0.36$ ,  $p < 0.001$ ), negative affect ( $r = -0.15$ ,  $p < 0.001$ ), and religiosity ( $r = 0.45$ ,  $p < 0.001$ ). Conversely, there were weak negative relations between the SCBCS scores and measures of anxiety ( $r = -0.02$ ,  $p = 0.44$ ), stress ( $r = -0.03$ ,  $p = 0.30$ ), and depression ( $r = -0.03$ ,  $p = 0.16$ ). All correlations were computed using Rasch-transformed interval scores for both the SCBCS and the other measures.

### Discussion

We used Rasch methodology to adapt the SCBCS into German while simultaneously conducting a cross-cultural validation and enhancing its measurement precision using

Rasch methodology across samples from Germany, Ghana, India, and New Zealand. Our initial analysis indicated an overall misfit of the SCBCS to the Rasch model, primarily due to item local dependency. We employed the testlet creation modification technique, which effectively resolved the item local dependency issues. This modification resulted in achieving the optimal overall Rasch model fit and confirmed strict unidimensionality, good reliability, and absence of ceiling and floor effects, including good targeting for the SCBCS. The formation of testlets was deliberately employed to address the observed disordered thresholds, and this improvement was theoretically predictable rather than coincidental. Disordered thresholds typically occur when items exhibit local dependency, causing an overlap in the variability attributed to random error rather than the intended latent trait. By aggregating items into testlets, we effectively reduced local item dependency, since the shared variance among locally dependent items is encapsulated within the testlet itself. Theoretically, this aggregation reduces random error variance by stabilizing response patterns, thereby clarifying the distinction between response categories and resulting in better-ordered thresholds. Thus, the observed improvement in model fit and threshold ordering following testlet creation is not an incidental finding, but rather an expected outcome informed by Rasch measurement theory, emphasizing the methodological rigor of our analytical approach (Wilkinson et al., 2023). Additionally, we developed an

**Table 3** Ordinal-to-interval transformation for the 5-item Santa Clara Brief Compassion Scale (SCBCS) ordinal total scores in logits and the scale metric ( $n = 500$ )

Ordinal scores	Interval	
	logits	Scale
5.00	-3.68	5.00
6.00	-2.88	8.19
7.00	-2.328	10.39
8.00	-1.947	11.91
9.00	-1.652	13.08
10.00	-1.409	14.05
11.00	-1.201	14.88
12.00	-1.019	15.60
13.00	-0.858	16.24
14.00	-0.712	16.83
15.00	-0.578	17.36
16.00	-0.453	17.86
17.00	-0.334	18.33
18.00	-0.22	18.79
19.00	-0.109	19.23
20.00	0.001	19.67
21.00	0.111	20.11
22.00	0.223	20.55
23.00	0.34	21.02
24.00	0.461	21.50
25.00	0.591	22.02
26.00	0.732	22.58
27.00	0.886	23.19
28.00	1.058	23.88
29.00	1.253	24.66
30.00	1.477	25.55
31.00	1.739	26.59
32.00	2.055	27.85
33.00	2.458	29.46
34.00	3.031	31.74
35.00	3.849	35.00

*Note:* This conversion table can only be used for complete responses to each of 5 scale items. To use this table, ordinal raw scores (left column) should be obtained by adding the observed scores for all 5 items. Next, match the ordinal total score (5–35) to the corresponding interval score in the right column (scale 5–35). A final converted score between 5 and 35 will be obtained, with higher scores corresponding to higher levels of compassion towards others. Researchers can be contacted for guidance on implementing these scores, including the provision of syntax for use in analysis

ordinal-to-interval transformation algorithm to enhance the precision of the SCBCS, valid for the four countries as well as representations from all these countries. The scale also demonstrated convergent validity with measures of optimism, self-compassion, religiosity, and positive affect across the countries. Furthermore, we established divergent

validity of the SCBCS using measures of psychological distress.

It is important to emphasize that observing misfit indices at the initial stage of our analyses, despite achieving unidimensionality, indicates a deviation from the expected fundamental assumption of monotonicity of the Rasch model. This means that while the scale effectively captures a single underlying construct, the responses do not consistently align with the principle that as the level of compassion towards others increases, individuals are more likely to endorse items that reflect this trait (Junker & Sijtsma, 2000). The use of testlets in modern Rasch analysis to improve a scale's psychometric parameters, as demonstrated in the current study, has emerged as a powerful methodological approach with several benefits for scale development and validation (Adu et al., 2024; Bartholomew et al., 2023). Testlets are useful for addressing local item dependence, which occurs when residual errors of individual items are correlated with each other, independently of the underlying latent trait being measured (Tennant & Küçükdeveci, 2023; Wang & Wilson, 2005). Ideally, each item is expected to contribute uniquely to the SCBCS. This local dependency violated the Rasch model's assumption of local independence, leading to spurious correlations and model misfit of the SCBCS. Such item dependencies can arise from shared wording, overlapping content, or response patterns not directly tied to the latent construct, which increases measurement error (Wang & Wilson, 2005). By grouping these locally dependent items into testlets, we reduced their shared error variance and preserved their theoretical relevance without discarding them from the scale (Boateng et al., 2018; Tennant & Küçükdeveci, 2023).

Testlet creation stabilizes response patterns by averaging out random fluctuations across items, which reduces the measurement error, enhances the model fit, and improves the precision of person estimates. The testlet aggregation tends to smooth irregularities that might have caused threshold disordering at the item level (Tennant & Küçükdeveci, 2023). Importantly, once such spurious measurement errors due to local dependency are reduced using testlets and the data fit the unidimensional Rasch model, such best-fitting model is used to develop an algorithm to transform ordinal scores potentially affected by spurious correlations and unequal distances between item response thresholds into interval level scores already accounting for such errors.

This highlights why testlets are particularly beneficial for interval-level scores. Testlets also help to maintain construct validity by retaining items that are conceptually important, but may not meet strict Rasch fit criteria when analyzed individually. Rather than eliminating misfitting items, which can compromise the content coverage and complexity of the constructs, testlets offer a balanced solution (Boateng et al., 2018). Testlets allow the researchers to uphold the fundamental measurement principles of the Rasch model

while preserving the theoretical soundness and complexities embedded in individual items (Medvedev et al., 2021).

It is important to emphasize that testlet creation was based solely on statistical indicators such as residual correlations between items exceeding a magnitude of 0.20, as relying on content-based judgments can be misleading due to a variety of reasons for such method effect (Boateng et al., 2018). While item content may suggest theoretical groupings, these deductions may not necessarily reflect actual error variance or item dependencies in the data. Statistical analyses provide empirical evidence of how items function, revealing local dependencies that might not be apparent from content alone. Moreover, testlets can only be effectively created when actual error variance is known. Therefore, using statistical criteria ensured a more accurate and objective basis for grouping items into testlets. The attainment of model fit indices in this study suggests that the Rasch model addresses a limitation of CCT methods regarding the SCBCS. Unlike CCT, the Rasch model recognizes the unequal intervals between response categories of the SCBCS and considers the probabilistic nature of the responses (Magno, 2009). This approach demonstrated higher precision in measurement, aligning with the fundamental principles of measurement (Thurstone, 1931).

Moreover, the use of Rasch-transformed interval-level data is important for accurately assessing meaningful changes in the scores of the SCBCS. Given that the Rasch model was successfully applied to widely used mindfulness measures such as the Five Facet Mindfulness Questionnaire (FFMQ; Medvedev et al., 2017) and the Comprehensive Inventory of Mindfulness Experiences (CHIME; Wilkinson et al., 2023) with interval-level scores available, it contributes to valid comparisons of compassion and mindfulness effects with a higher precision, particularly with the available bidirectional evidence showing that individuals who exhibit greater compassion toward others are also more likely to display mindfulness-related traits, and vice versa (Boellinghaus et al., 2014; Jazaieri et al., 2014). This reciprocal link underscores the ongoing integration of compassion (both towards self and others) training into mindfulness-based interventions, which has been found to yield positive outcomes such as enhanced emotional regulation, improved interpersonal connection, and greater psychological well-being (Huppert, 2017; Medvedev et al., 2021). Therefore, acknowledging compassion and mindfulness as mutually reinforcing constructs supports their conceptual and practical value in therapeutic and contemplative contexts.

Consequently, advances in psychometric methods, such as the application of the Rasch model to the SCBCS in the current study, are essential for improving the precision with which changes in compassion are measured during intervention studies. For example, consider two scenarios in

which compassion scores increase from 5 to 10 and from 10 to 15 on an ordinal scale. Although these changes appear equal, the Rasch-transformed interval scores reveal that the first change represents a difference of 9.05 units, while the second reflects only 3.31 units. This demonstrates how interval-level measurement offers a more comprehensive understanding of meaningful change over time. Such precision is important for accurately capturing the effects of mindfulness-based interventions on compassion toward others and vice versa, especially in both research and clinical settings. In sum, compassion toward others is an increasingly important construct in mindfulness research. When assessed with precise measurement tools, it will provide valuable insights into the personal and relational benefits of mindfulness and the mechanisms through which these outcomes are achieved. Of note, researchers do not need to develop new interval-level scores for the SCBCS. They simply need to obtain the total ordinal score from the SCBCS and refer to the provided conversion table to retrieve the corresponding interval-level score. This approach saves time and facilitates easier application of the results in both research and practice (refer to Table 3).

Additionally, the absence of ceiling and floor effects indicates that the SCBCS perfectly targeted the sample, demonstrating the precision of the scale in measuring the construct of interest across a wide range of levels of compassion towards others within the sample. Thus, the SCBCS is sensitive enough to distinguish between individuals at varying levels of compassion towards others (Medvedev et al., 2019; Tennant & Küçükdeveci, 2023). This quality demonstrates the reliability of the SCBCS. Scales with a PSI of  $> 0.80$  are suitable for assessing individuals, while a PSI of  $> 0.70$  is preferable for assessing groups of individuals (Fisher, 1992; Mallinson et al., 2022). This indicates that the current PSI positions the SCBCS as a suitable scale for consistently and accurately measuring compassion towards others at both individual and group levels. It is important to note that reliability in CCT methods focuses on score consistency and assumes that measurement errors are random and uncorrelated. On the other hand, Rasch methodology relies on item ability to differentiate between individuals with varying levels of a trait, while considering the non-random nature of errors. Therefore, the Rasch model tends to provide more accurate reliability estimates, such as the current PSI of the SCBCS, compared to traditional CTT (Mallinson et al., 2022).

Furthermore, the absence of DIF suggests that regardless of individuals' sex, age, country of origin, marital status, and employment status, the items on the scale performed similarly on average. This shows the SCBCS is invariant across individuals with different demographic characteristics. Achieving measurement invariance across such numerous personal factors enhances the scale's utility, credibility, and

consistency, enabling reliable comparisons of study results (Welzel et al., 2023). This is especially important considering the significant differences observed in these personal factors across countries. The significant positive correlation between the SCBCS scores and related constructs, including positive affect, self-compassion, religiosity, and optimism, aligned with existing literature and supports the scale's convergent validity (Chan et al., 2022; Medvedev et al., 2021). While some of the correlation coefficients were relatively low, they were statistically significant, thereby meeting the criteria for convergent validity (Jones et al., 2024). On the other hand, the low and non-significant negative correlations between the SCBCS scores and measures of psychological distress support the scale's divergent validity. These findings suggest that the SCBCS measures compassion towards others in a way that is consistent with well-established measures of related constructs, while also distinguishing itself from measures of unrelated constructs (Cooper, 2023).

### Strengths, Limitations and Future Research

Our study has translated and validated the German version of the SCBCS for the first time, which is contributing to enhancing well-being research among the German-speaking population. Additionally, we provided new psychometric evidence for reliability, internal validity, divergent validity, and convergent validity of the SCBCS using Rasch methodology across four countries. By developing an algorithm for transforming ordinal to interval-level scores, studies can now assess compassion towards others with greater precision across a wide range of contexts using interval-level scores. Furthermore, our study included samples from four distinct countries across four continents, increasing the external validity of our study. However, our study has limitations, including the use of a relatively homogeneous convenience sample recruited through social media, which makes it challenging to estimate response rates. While Rasch methodology is not heavily influenced by sample characteristics, there may still be cultural and language influences on how individuals respond to scale items. Therefore, our findings may not be easily generalized to other populations, for example clinical samples.

In summary, we utilized Rasch methodology to adapt, assess, and enhance the measurement precision of the SCBCS. Our data fit the Rasch model well, supporting the scale's unidimensionality, as well as its convergent and divergent validity. The scale demonstrated reliability and measurement invariance across various demographic factors. Additionally, we developed an algorithm to convert ordinal scores to interval-level scores, improving the scale's precision. Further research, encompassing multi-lingual versions

of the SCBCS, is necessary to establish its universal acceptance as a measure for assessing compassion towards others and related health outcomes.

## Appendix

### German Version of the Santa Clara Brief Compassion Scale (SCBCS)

Bitte beantworten Sie die nachfolgenden Fragen ehrlich und spontan auf Basis der vorgegebenen Skalierung—trifft gar nicht auf mich zu (1) bis trifft voll und ganz auf mich zu (7).

1. Wenn ich von jemandem (einer/einem Fremden) höre, dass sie/er durch eine schwere Zeit geht, empfinde ich starkes Mitgefühl.
2. Ich neige dazu, Mitgefühl zu empfinden, auch wenn ich jemanden gar nicht kenne.
3. Anderen zu helfen, wenn sie mich brauchen, ist für mich eine der sinnvollsten Tätigkeiten im Leben.
4. Ich würde mich eher für andere einsetzen, auch wenn es Fremde sind, als mich für Dinge einzusetzen, die mir helfen.
5. Ich entwickle oft Gefühle für Menschen (Fremde), die scheinbar Hilfe brauchen.

**Author Contribution** P.A., T.P., O.N.M., and C.R.S. involved in the conception of the study; P.A. led the writing of the manuscript; O.N.M. oversaw the analysis of data; N.I. helped with data collection in India; A.R. helped with the adaptation of some instruments into German D.C. and S.C. were advisors who provided comments to improve the manuscript; C.R.S. edited the final manuscript and was P.A.'s lead PhD supervisor. All authors contributed to the study design. All authors contributed to drafting the paper and revised the manuscript for important intellectual content. All authors gave final approval for this version to be published.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions. The authors received no direct financial support for the research, authorship, or publication of this article. However, the lead author acknowledges the Wellington Doctoral Scholarship, which supported their research during the study.

**Data Availability** The study participants did not consent to having their data shared publicly. The deidentified participant dataset from the current study can be made available to researchers upon a reasonable request to the corresponding author.

### Declarations

**Ethics Approval and Consent to Participate** The study was approved by the institutional Human Research Ethics Committee of Victoria University of Wellington, Wellington, New Zealand (#0000029770). The study is in line with the Declaration of Hel-

sinki, which outlines fundamental ethical principles for health research involving the use of human participants (World Medical Association, 2001).

**Informed Consent** Participants provided informed consent by clicking a button after reading the consent information. They agreed for their results to be published or used for academic purposes such as reports, presentations, and public documentation, with data presented in aggregate form (i.e., combined and analyzed with others).

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adu, P., Jurcik, T., & Grigoryev, D. (2021). Mental health literacy in Ghana: Implications for religiosity, education and stigmatization. *Transcultural Psychiatry*, *58*(4), 516–531. <https://doi.org/10.1177/13634615211022177>
- Adu, P., Popoola, T., Bartholomew, E., Iqbal, N., Roemer, A., Jurcik, T., Collings, S., Aspin, C., Medvedev, O. N., & Simpson, C. R. (2024). Enhancing the precision of the self-compassion scale short form (SCS-SF) with Rasch methodology. *Mindfulness*, *15*(11), 2779–2793. <https://doi.org/10.1007/s12671-024-02462-y>
- Adu, P., Popoola, T., Iqbal, N., Medvedev, O. N., & Simpson, C. R. (2025). Validating the depression anxiety stress scales (DASS-21) across Germany, Ghana, India, and New Zealand using Rasch methodology. *Journal of Affective Disorders*. <https://doi.org/10.1016/j.jad.2025.04.099>
- Adu, P., Popoola, T., Roemer, A., Collings, S., Aspin, C., Medvedev, O. N., & Simpson, C. R. (2023). Validation and cultural adaptation of the Motors of COVID-19 Vaccination Acceptance Scale (movac-COVID19S) in German. *Psychological Test Adaptation and Development*. <https://doi.org/10.1027/2698-1866/a000064>
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561–573. <https://doi.org/10.1007/BF02293814>
- Andrich, D., Sheridan, B., & Luo, G. (2009). *RUMM 2030*. RUMM Laboratory.
- Bartholomew, E., Roemer, A., & Medvedev, O. N. (2023). Validation and enhancement of the 6-Item Nature-Relatedness Scale using Rasch analysis. *Ecopsychology*, *15*(1), 81–90. <https://doi.org/10.1089/eco.2022.0031>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Boellinghaus, I., Jones, F. W., & Hutton, J. (2014). The role of mindfulness and loving-kindness meditation in cultivating self-compassion and other-focused concern in health care professionals. *Mindfulness*, *5*(2), 129–138. <https://doi.org/10.1007/s12671-012-0158-6>
- Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Cross-cultural adaptation and validation of psychological instruments: Some considerations. *Paidéia*, *22*(53), 423–432. <https://doi.org/10.1590/S0103-863X2012000300014>
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216. <https://doi.org/10.1177/135910457000100301>
- Caycho-Rodríguez, T., Vilca, L. W., Plante, T. G., Carbajal-León, C., Cabrera-Orosco, I., García Cadena, C. H., & Reyes-Bossio, M. (2022). Spanish version of the Santa Clara Brief Compassion Scale: Evidence of validity and factorial invariance in Peru. *Current Psychology*, *41*, 4431–4446. <https://doi.org/10.1007/s12144-020-00949-0>
- Chan, K. K. S., Lee, J.C.-K., Yu, E. K. W., Chan, A. W., Leung, A. N. M., Cheung, R. Y., Li, C. W., Kong, R.H.-M., Chen, J., & Wan, S. L. Y. (2022). The impact of compassion from others and self-compassion on psychological distress, flourishing, and meaning in life among university students. *Mindfulness*, *13*(6), 1490–1498. <https://doi.org/10.1007/s12671-022-01891-x>
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q 3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement*, *41*(3), 178–194. <https://doi.org/10.1177/014662161667752>
- Cooper, C. (2023). *An introduction to psychometrics and psychological assessment: Using, interpreting and developing tests*. Routledge. <https://doi.org/10.4324/9781003240181>
- Cosley, B. J., McCoy, S. K., Saslow, L. R., & Epel, E. S. (2010). Is compassion for others stress buffering? Consequences of compassion and social support for physiological reactivity to stress. *Journal of Experimental Social Psychology*, *46*(5), 816–823. <https://doi.org/10.1016/j.jesp.2010.04.008>
- Dellaert, F. (2002). *The expectation maximization algorithm*. College of Computing, Georgia Institute of Technology.
- Demars, C. E. (2013). A comparison of confirmatory factor analysis and multidimensional Rasch models to investigate the dimensionality of test-taking motivation. *Journal of Applied Measurement*, *14*(2), 179–196.
- Fisher Jr, W. (1992). Reliability statistics. *Rasch Measurement Transactions*. *MESA Press*, *6*, 238.
- George, D. (2011). *SPSS for windows step by step: A simple study guide and reference, 17.0 update, 10/e*. Pearson Education India.
- Gilbert, P. (2020). Compassion: From its evolution to a psychotherapy. *Frontiers in Psychology*, *11*, 3123. <https://doi.org/10.3389/fpsyg.2020.586161>
- Glaesmer, H., Hoyer, J., Klotsche, J., & Herzberg, P. Y. (2008). Die deutsche Version des Life-Orientierung-Tests (LOT-R) zum dimensionellen Optimismus und Pessimismus. *Zeitschrift für Gesundheitspsychologie*, *16*(1), 26–31. <https://doi.org/10.1026/0943-8149.16.1.26>
- Goudarzian, A. H. (2023). Challenges and recommendations of exploratory and confirmatory factor analysis: A narrative review from a nursing perspective. *Journal of Nursing Reports in Clinical Practice*, *1*(3), 133–137. <https://doi.org/10.32598/JNRCP.23.44>
- Hagell, P. (2015). Testing unidimensionality using the PCA/t-test protocol with the Rasch model: A cautionary note. *Rasch Measurement Transactions*, *28*(4), 1487–1489.
- 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.

- Hagquist, C., & Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and Quality of Life Outcomes, 15*, 181. <https://doi.org/10.1186/s12955-017-0755-0>
- Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care, 38*(9 Suppl), II28–42. <https://doi.org/10.1097/00005650-200009002-00007>
- Hoseininezhad, N., Hosseini, S., & Abdollahi, A. (2023). Psychometric properties of the Santa Clara Brief Compassion Scale-Farsi version. *Pastoral Psychology, 71*(1), 1–11. <https://doi.org/10.1007/s11089-023-01054-2>
- Huber, S., & Huber, O. W. (2012). The centrality of religiosity scale (CRS). *Religions, 3*(3), 710–724. <https://doi.org/10.3390/rel3030710>
- Huppert, F. A. (2017). Mindfulness and compassion as foundations for well-being. In *Future directions in well-being: Education, organizations and policy* (pp. 225–233). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-56889-8\\_39](https://doi.org/10.1007/978-3-319-56889-8_39)
- Hwang, J. Y., Plante, T., & Lackey, K. (2008). The development of the Santa Clara brief compassion scale: An abbreviation of Sprecher and Fehr's compassionate love scale. *Pastoral Psychology, 56*(4), 421–428. <https://doi.org/10.1007/s11089-008-0117-2>
- Jazaieri, H., McGonigal, K., Jinpa, T., Doty, J. R., Gross, J. J., & Goldin, P. R. (2014). A randomized controlled trial of compassion cultivation training: Effects on mindfulness, affect, and emotion regulation. *Motivation and Emotion, 38*(1), 23–35. <https://doi.org/10.1007/s11031-013-9368-z>
- Jones, S., Yu, M. L., & Brown, T. (2024). Convergent validity between the school-age versions of the Sensory Processing Measure 2 (SPM2) and the Sensory Profile 2 (SP2): A pilot study. *Australian Occupational Therapy Journal, 71*(5), 718–732. <https://doi.org/10.1111/1440-1630.12958>
- Junker, B. W., & Sijtsma, K. (2000). Latent and manifest monotonicity in item response models. *Applied Psychological Measurement, 24*(1), 65–81. <https://doi.org/10.1177/014662160002410>
- Karnaze, M. M., Bellettiere, J., & Bloss, C. S. (2022). Association of compassion and empathy with prosocial health behaviors and attitudes in a pandemic. *PLoS ONE, 17*(7), e0271829. <https://doi.org/10.1371/journal.pone.0271829>
- Kim, J. J., Parker, S. L., Doty, J. R., Cunningham, R., Gilbert, P., & Kirby, J. N. (2020). Neurophysiological and behavioral markers of compassion. *Scientific Reports, 10*(1), 6789. <https://doi.org/10.1038/s41598-020-63846-3>
- Lee, E. E., Govind, T., Ramsey, M., Wu, T. C., Daly, R., Liu, J., Tu, X. M., Paulus, M. P., Thomas, M. L., & Jeste, D. V. (2021). Compassion toward others and self-compassion predict mental and physical well-being: A 5-year longitudinal study of 1090 community-dwelling adults across the lifespan. *Translational Psychiatry, 11*(1), 397. <https://doi.org/10.1038/s41398-021-01491-8>
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>
- López, A., Sanderman, R., Ranchor, A. V., & Schroevers, M. J. (2018). Compassion for others and self-compassion: Levels, correlates, and relationship with psychological well-being. *Mindfulness, 9*(1), 325–331. <https://doi.org/10.1007/s12671-017-0777-z>
- López, A., Sanderman, R., Smink, A., Zhang, Y., Van Sonderen, E., Ranchor, A., & Schroevers, M. J. (2015). A reconsideration of the Self-Compassion Scale's total score: Self-compassion versus self-criticism. *PLoS ONE, 10*(7), e0132940. <https://doi.org/10.1371/journal.pone.0132940>
- 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](https://doi.org/10.1016/0005-7967(94)00075-U)
- MacBeth, A., & Gumley, A. (2012). Exploring compassion: A meta-analysis of the association between self-compassion and psychopathology. *Clinical Psychology Review, 32*(6), 545–552. <https://doi.org/10.1016/j.cpr.2012.06.003>
- Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment, 1*(1), 1–11.
- Mallinson, T., Kozłowski, A. J., Johnston, M. V., Weaver, J., Terhorst, L., Grampurohit, N., Juengst, S., Ehrlich-Jones, L., Heinemann, A. W., & Melvin, J. (2022). Rasch reporting guideline for rehabilitation research (RULER): The RULER statement. *Archives of Physical Medicine and Rehabilitation, 103*(7), 1477–1486. <https://doi.org/10.1016/j.apmr.2022.03.013>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. <https://doi.org/10.1007/BF02296272>
- Matos, M., McEwan, K., Kanovský, M., Halamová, J., Steindl, S. R., Ferreira, N., Linhares, M., Rijo, D., Asano, K., & Márquez, M. G. (2022). Compassion protects mental health and social safety during the COVID-19 pandemic across 21 countries. *Mindfulness, 13*(4), 863–880. <https://doi.org/10.1007/s12671-021-01822-2>
- Medvedev, O. N., & Krägeloh, C. U. (2022). Rasch measurement model. In O. N. Medvedev, C. U. Krägeloh, R. J. Siebert, & N. N. Singh (Eds.), *Handbook of assessment in mindfulness research*. Springer. [https://doi.org/10.1007/978-3-030-77644-2\\_4-1](https://doi.org/10.1007/978-3-030-77644-2_4-1)
- Medvedev, O. N., Cervin, M., Barcaccia, B., Siebert, R. J., Roemer, A., & Krägeloh, C. U. (2021). Network analysis of mindfulness facets, affect, compassion, and distress. *Mindfulness, 12*(4), 911–922. <https://doi.org/10.1007/s12671-020-01555-8>
- Medvedev, O. N., Krägeloh, C. U., Hill, E. M., Billington, R., Siebert, R. J., Webster, C. S., Booth, R. J., & Henning, M. A. (2019). Rasch analysis of the Perceived Stress Scale: Transformation from an ordinal to a linear measure. *Journal of Health Psychology, 24*(8), 1070–1081. <https://doi.org/10.1177/135910531668960>
- Medvedev, O. N., Siebert, 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>
- Mongrain, M., Chin, J. M., & Shapira, L. B. (2011). Practicing compassion increases happiness and self-esteem. *Journal of Happiness Studies, 12*, 963–981. <https://doi.org/10.1007/s10902-010-9239-1>
- Novak, L., Malinakova, K., Mikoska, P., Furstova, J., Cann, R., & Tavel, P. (2021). Psychometric evaluation of the Santa Clara Brief Compassion Scale in the Czech environment (SCBCS). *Mental Health, Religion & Culture, 24*(3), 294–305. <https://doi.org/10.1080/13674676.2020.1850666>
- Pelton, T. (2002). Where are the limits to the Rasch advantage? In *Paper presented at the International objective measurement workshop*.
- Perez-Bret, E., Altisent, R., & Rocafort, J. (2016). Definition of compassion in healthcare: A systematic literature review. *International Journal of Palliative Nursing, 22*(12), 599–606. <https://doi.org/10.12968/ijpn.2016.22.12.599>
- Pratscher, S. D., Wood, P. K., King, L. A., & Bettencourt, B. A. (2019). Interpersonal mindfulness: Scale development and initial construct validation. *Mindfulness, 10*(6), 1044–1061. <https://doi.org/10.1007/s12671-018-1057-2>
- Prudon, P. (2015). Confirmatory factor analysis as a tool in research using questionnaires: A critique. *Comprehensive Psychology, 4*, 03. <https://doi.org/10.2466/03.CP.4.10>
- Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests (Expanded ed.)*. MESA Press.

- 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>
- Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67(6), 1063. <https://doi.org/10.1037/0022-3514.67.6.1063>
- Smith, E. V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3(2), 205–231.
- Sprecher, S., & Fehr, B. (2005). Compassionate love for close others and humanity. *Journal of Social and Personal Relationships*, 22(5), 629–651. <https://doi.org/10.1177/0265407505056439>
- 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 Care & Research*, 57(8), 1358–1362. <https://doi.org/10.1002/art.23108>
- Tennant, A., & Küçükdeveci, A. A. (2023). Application of the Rasch measurement model in rehabilitation research and practice: Early developments, current practice, and future challenges. *Frontiers in Rehabilitation Sciences*, 4, 1208670. <https://doi.org/10.3389/fresc.2023.1208670>
- Thurstone, L. L. (1931). The measurement of social attitudes. *The Journal of Abnormal and Social Psychology*, 26(3), 249–269. <https://doi.org/10.1037/h0070363>
- Trninić, V., Jelaska, I., & Štalec, J. (2013). Appropriateness and limitations of factor analysis methods utilized in psychology and kinesiology: Part II. *Fizička Kultura*, 67(1), 1–17.
- Wang, W. C., & Wilson, M. (2005). The Rasch testlet model. *Applied Psychological Measurement*, 29(2), 126–149. <https://doi.org/10.1177/0146621604271105>
- 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>
- Welzel, C., Brunkert, L., Kruse, S., & Inglehart, R. F. (2023). Non-invariance? An overstated problem with misconceived causes. *Sociological Methods & Research*, 52(3), 1368–1400. <https://doi.org/10.1177/0049124121995521>
- Wilkinson, S., Ribeiro, L., Krägeloh, C. U., Bergomi, C., Parsons, M., Siegling, A., Tschacher, W., Kupper, Z., & Medvedev, O. N. (2023). Validation of the comprehensive inventory of mindfulness experiences (CHIME) in English using Rasch methodology. *Mindfulness*, 14(5), 1204–1218. <https://doi.org/10.1007/s12671-023-02099-3>
- World Medical Association. (2001). World medical association declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, 79(4), 373.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.