ORIGINAL PAPER

Enhancing the Precision of the Self-Compassion Scale Short Form (SCS-SF) with Rasch Methodology

Peter Adu[1] · Tosin Popoola[2] · Emerson Bartholomew[3] · Naved Iqbal[4] · Anja Roemer[5] · Tomas Jurcik[6] · Sunny Collings[1] · Clive Aspin[1] · Oleg N. Medvedev[7] · Colin R. Simpson[1,8]

Abstract
Objectives Precise measurement of self-compassion is essential for informing well-being–related policies. Traditional assessment methods have led to inconsistencies in the factor structure of self-compassion scales. We used Rasch methodology to enhance measurement precision and assess the psychometric properties of the Self-Compassion Scale Short Form (SCS-SF), including its invariance across Ghana, Germany, India, and New Zealand.
Method We employed the Partial Credit Rasch model to analyse responses obtained from 1000 individuals randomly selected (i.e. 250 from each country) from a total convenience sample of 1822 recruited from the general populations of Germany, Ghana, India, and New Zealand.
Results The initial identification of local dependency among certain items led to a significant misfitting of the SCS-SF to the Rasch model (χ^2 (108) = 260.26, $p < 0.001$). We addressed this issue by merging locally dependent items, using testlets. The solution with three testlets resulted in optimal fit of the SCS-SF to the Rasch model (χ^2 (27) = 23.84, $p = 0.64$), showing evidence of unidimensionality, strong sample targeting ($M = 0.20$; $SD = 0.72$), and good reliability (Person Separation Index = 0.71), including invariance across sociodemographic factors. We then developed ordinal-to-interval conversion tables based on the Rasch model's person estimates. The SCS-SF showed positive correlations with measures of compassion towards others, optimism, and positive affect, alongside negative associations with psychological distress and negative affect.
Conclusions The current study supports the reliability, as well as the structural, convergent, and external validity of the SCS-SF. By employing the ordinal-to-interval conversion tables published here, the precision of the measure is significantly enhanced, offering a robust tool for investigating self-compassion across different cultures.

Keywords Measurement · Self-Compassion Scale · Rasch Analysis · Mindfulness · Psychometrics

Research on self-compassion has gained substantial interest in the international literature due to its impact across diverse dimensions of well-being. For instance, a meta-analysis of 14 studies found a large negative effect size for the relation between self-compassion and psychopathology (MacBeth & Gumley, 2012). Hwang et al. (2019) also identified self-compassion as the most influential predictor of reduced educator stress in Australian students. In addition, Lefebvre et al. (2020) established a connection between individuals' workplace resilience and self-compassion. Similarly, a review of 28 studies revealed that self-compassion protects against the development of poor body image and the emergence of risk factors for maladaptive behaviours (Braun et al., 2016).

The concept of self-compassion refers to the internal nurturing of emotional well-being and mental health. It involves fully accepting and openly understanding an individual's life adversities without self-judgment or excessive self-criticism (Neff, 2003a). In other words, self-compassion entails the treatment of oneself with the same kindness, care, love, and understanding that one will offer to a significant close relation such as a friend in times of life setbacks. The three proposed key subconstructs of self-compassion encompass being kind to oneself (self-kindness), acknowledging that life challenges are part of common human experience (common humanity), and practicing awareness of thoughts and feelings without overly identifying with them (mindfulness; Neff, 2003a). Notably, the positive impacts of the mindfulness subconstruct on overall well-being have received significant attention in the literature. For example, mindfulness

---

Springer

has been associated with reduced levels of depression, anxiety, and stress in challenging situations, as exemplified during the COVID-19 pandemic (Hartstone & Medvedev, 2021).

Accurate measurement of this essential positive psychological resource is vital for advancing our understanding of its impact on mental health, guiding interventions and treatments, and informing policies related to well-being. To date, the assessment of self-compassion in the existing literature primarily relies on two widely recognised versions of the same psychometric scale: the 26-item Self-Compassion Scale (Neff, 2003b, 2016) and a 12-item Self-Compassion Scale Short Form (SCS-SF; Raes et al., 2011). Using either of these versions, the conceptually separate yet overarching aspects of self-compassion are measured through positively worded items related to self-kindness, common humanity, and mindfulness, as well as negatively worded items related to self-judgment, feelings of isolation, and over-identification. The negative items reflect behaviours and thought patterns that are less compassionate in nature (Neff, 2003a). As such, the scales can be employed as a unidimensional or multi-dimensional measure. Notably, the scale has been instrumental in the study of self-compassion (Neff & Tóth-Király, 2022).

Furthermore, evaluation of the psychometric properties of the scale has predominantly been conducted with confirmatory factor analyses (CFA; e.g. Rahman et al., 2023), a method etymologised in Classical Test Theory (CTT). This method has so far provided evidence supporting the reliability and validity of the Self-Compassion Scale, demonstrating that the scale is a suitable measure for assessing self-compassion across various samples (Babenko & Guo, 2019; Neff et al., 2021). However, an ongoing controversy exists regarding the factor structure of the scale. While some studies have supported the six-factor structure, others have reported a two-factor structure for this scale. For instance, López et al. (2015) found that the Dutch version of the Self-Compassion Scale supported a two-factor model, self-compassion and self-criticism, rather than the traditional six-factor structure. This pattern was supported by Costa et al. (2016), including the Bangla and Turkish versions of the scale (Koğar & Koğar, 2023; Rahman et al., 2023). A two-factor solution, involving positive and negative self-compassion, was also confirmed among Spanish nurses (Lluch-Sanz et al., 2022).

Studies have highlighted possible reasons for the controversies surrounding the structure of the scale. For example, studies have reported that including the uncompassionate components items (self-judgment, feelings of isolation, over-identification), which initially served as reversed items for the main self-compassion measure (self-kindness, common humanity, mindfulness) in the Self-Compassion Scale, unintentionally introduced new dimensions alongside the compassionate components (self-kindness, common humanity,

mindfulness; Wong et al., 2003). Given that uncompassionate components are reverse coded, they intend to complement compassionate aspects by providing more comprehensive coverage of the construct, rather than measuring two conceptually opposing constructs. However, the uncompassionate items were found to be linked to negative health outcomes such as depression symptoms, while the compassionate items are associated with positive outcomes like happiness. The Self-Compassion Scale now includes multiple dimensions with different relations to external constructs, providing an advantage of measuring specific aspects of self-compassion independently and as a total score by reverse coding uncompassionate items. This may create confusion for researchers less familiar with psychometrics and measurement (Muris & Otgaar, 2020), which can be addressed by applying the unidimensional Rasch measurement model using testlets. Ultimately, if an acceptable fit to the Rasch model is achieved using testlets, this finding would be taken as evidence of unidimensionality and support for using the total scale score (Sutton & Medvedev, 2023).

Notwithstanding, methods of analysis based on CTT are prone to spurious correlations due to method effects, possibly resulting in incongruent findings regarding the dimensionality of the scales. These methods are also largely sample dependent, which leads to the estimation of parameters that can over-represent the idiosyncrasies of a specific sample rather than accurately representing the underlying structure of the population, resulting in poor model generalisability (Magno, 2009). Again, since instruments are not perfect, observed scores could be divergent from the true ability, state, or trait of an individual; thus, estimation of true score under CTT does not exclude measurement error (Eluwa et al., 2011; Magno, 2009). These inconsistencies in the factor structures of the self-compassion scales and the limitations of CTT raise questions about the unidimensionality and measurement accuracy of the Self-Compassion Scale, including the extent to which these scales maintain measurement invariance across diverse countries (Hambleton, 1994).

Rasch analysis, a psychometric analytical technique, incorporates probabilistic modeling to assess and enhance the measurement properties of items and scales (Medvedev & Krägeloh, 2022; Tennant & Küçükdeveci, 2023). It is part of a family of models under the Modern Test Theory (MTT) umbrella (Ellis & Mead, 2004). MTT techniques such as Rasch analysis transcend many of the limitations of the CTT-derived methods such as CFA. Thus, Rasch models provide a set of criteria, including the consideration of respondents' abilities and item difficulties, assessing response options for polytomous items (i.e. items with more than two response categories), and converting ordinal-level data into a more reliable interval-level scale. This ensures a robust assessment of an instrument. Furthermore, Rasch methods align

strictly with the fundamental measurement principles laid out by Thurstone (1931), emphasizing non-discrimination between instrument users (invariance), unidimensionality, and equally proportioned scale units (i.e. concatenability).

Despite the advantages of Rasch analysis in psychometric assessment, only one study (Finaulahi et al., 2021) in the existing literature has applied this methodology to the assessment of the self-compassion scales. However, this study lacked cross-country validation of these scales, as the study was conducted with an English-speaking population only, predominantly composed of individuals of White British ethnicity, with an overrepresentation of females. Additionally, the researchers only assessed invariance based on two demographic factors: age and sex. These pitfalls lessen the confidence of applying these scales in other contexts. The study also did not evaluate convergent or divergent validity. Therefore, there is lack of robust evidence regarding the psychometric characteristics of the self-compassion scales. Hence, we sought to provide further psychometric assessment of the 12-item SCS-SF employing the Rasch analysis, with a primary focus on evaluating reliability and various forms of validity, including structural, convergent, and divergent validities. In addition to these, we examined the scale's invariance across sociodemographic factors such as country, age, education, and sex using data from diverse samples in Ghana, Germany, India, and New Zealand.

Based on evidence in the literature, we anticipated a positive association between the SCS-SF scores and measures of compassion towards others, optimism, and positive affect, examining the convergent validity of the scale (Neff et al., 2007). However, we assessed divergent validity of the SCS-SF by hypothesising a weak to zero correlation between the SCS-SF scores and measures of psychological distress, negative affect, and pessimism (Medvedev et al., 2021; Shapira & Mongrain, 2010).

Method

Participants

We randomly selected 1000 (e.g. 250 from each country) participants from a total sample of 1822 recruited from Germany (475), Ghana (523), India (411), and New Zealand (413) during the months of June and July 2022 for our Rasch analysis (Fig. 1). 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 ($M_{age} = 44.09$; $SD = 5.57$). The randomly selected participants differed significantly in terms of education (χ^2 (6) = 708.48, $p < 0.001$), sex (χ^2 (3) = 21.72, $p < 0.001$), and age (χ^2 (6) = 387.74, $p < 0.001$) across the sample.

Fig. 1 Flowchart for participant sampling process from the four countries ($n = 250$ per country)
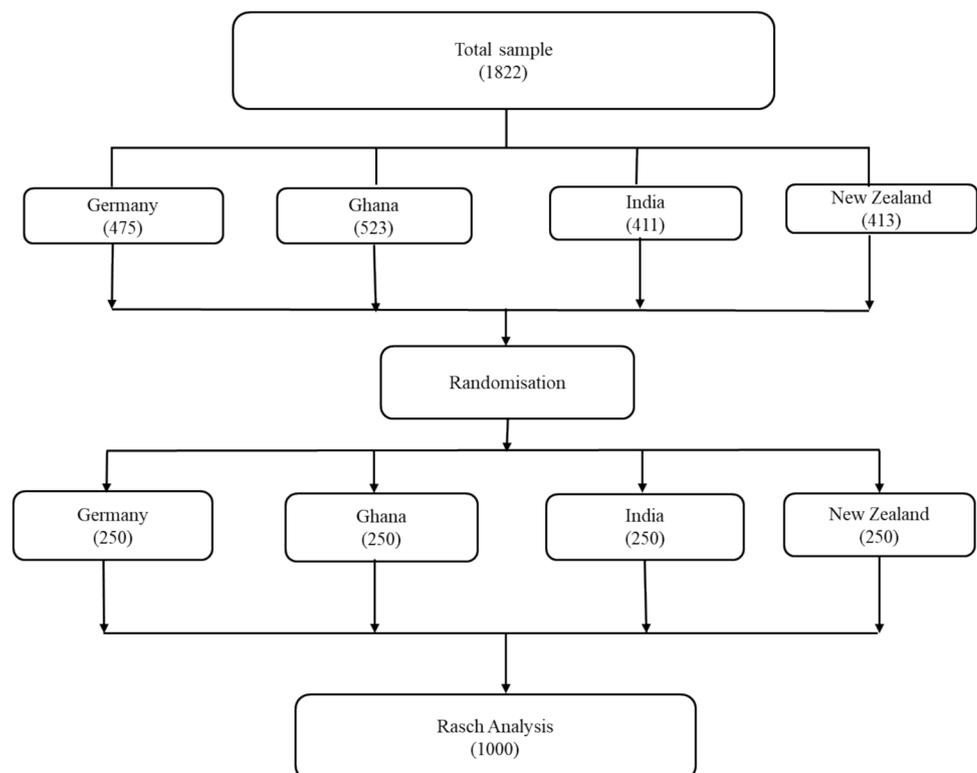
Power Analysis

Rasch models are less reliant on sample size since they estimate parameters from individual responses rather than data volume (Hagell & Westergren, 2016; Tennant & Küçükdeveci, 2023), allowing for precise estimations, even with smaller sample sizes. This approach reduces sensitivity to chi-square values, which can inflate statistical significance without practical impact (Pelton, 2002). Therefore, our sample selection was to effectively balance the benefits of larger samples with the challenge of chi-square sensitivity. To ensure parameter accuracy, a sample size of around 250 to 500 is recommended for Rasch analyses using the Rasch Unidimensional Measurement Model (RUMM; Hagell & Westergren, 2016).

Procedure

The data from Ghana and India were collected utilising *SelectSurvey.net* software via various online platforms, including Facebook, WhatsApp, Twitter, Instagram, and email, using convenience sampling. We relied on our social network for data from these two countries; as such, participants were not rewarded. In New Zealand and Germany, data collection was facilitated by the Qualtrics data collection company, and participants were remunerated. Online data collection offers a cost-effective means of reaching diverse populations in various locations (Lefever et al., 2007). The questionnaires were presented in English for participants in Ghana, India, and New Zealand, while a German version was provided for participants in Germany. Participants initially provided demographic information and then completed the main survey, which typically took around 15 min. The data used for the current paper were part of a larger international dataset on psychological factors and COVID-19 vaccination attitudes. Sections of the current data have been analysed using different methods and concepts, triangulating the results. For instance, previous studies have utilised this data to establish links between psychological factors and COVID-19 vaccination attitudes (Adu et al., 2024b) and have adapted and validated the COVID-19 vaccination attitudes scales using CFA (Adu et al., 2023, 2024a).

Measures

Self-Compassion

The 12-item SCS-SF (Raes et al., 2011) is a self-report questionnaire designed to assess self-compassion. It comprises six subscales (self-kindness, self-judgment, common humanity, isolation, mindfulness, and over-identification), each consisting of two items. Table 2 provides detailed information regarding the sub- and full scales.

This scale is the shortened version of the main and initial 26-item Self-Compassion Scale (Neff, 2003b). The scale uses a 5-point Likert-scale response format: 1 = *Almost Never* to 5 = *Almost Always*. To calculate the total scores for the SCS-SF, negative items (self-judgment, isolation, and over-identification) are reverse-scored. Refer to the "Results" section for the reliability (Person Separation Index) of this scale.

Psychological Distress

We measured psychological distress using the Depression Anxiety Stress Scale (DASS-21; Lovibond & Lovibond, 1995). This 21-item instrument is 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*. Sample items from the scale encompass: 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"). This scale demonstrated excellent reliability for the overall sample (Cronbach's $\alpha = 0.98$, McDonald's $\omega = 0.98$; $M = 30.80$, $SD = 18.00$).

Positive Affect and Negative Affect

We assessed Positive Affect and Negative Affect with the popular 20-item Positive Affect (PA) and Negative Affect (NA) Scale (PANAS; Watson et al., 1988). Each adjective on this scale is rated on 5-point Likert scale ranging from 1 = *very slightly* to 5 = *extremely*. Examples of adjectives measuring PA include "interested", "strong", and "proud", while NA comprises "anger", "fear", and "sadness". In this study, the reliability coefficient for the PA subscale for the whole sample was excellent ($\alpha = 0.91$; $\omega = 0.93$, $M = 31.40$, $SD = 7.40$). The NA subscale also ranged from very good to excellent ($\alpha = 0.89$; $\omega = 0.92$, $M = 23.00$, $SD = 8.70$).

Optimism Versus Pessimism

The revised version of the Life Orientation Test (LOT-R; Scheier et al., 1994) was employed to measure optimism and pessimism. This 10-item is rated on a 5-point Likert scale from 0 = *strongly disagree* to 5 = *strongly agree*. An example of a positively worded item on this scale is "In uncertain times, I usually expect the best", and a negatively worded item is "If something can go wrong for me, it will". The scale showed relatively low reliability for the total sample ($\alpha = 0.57$; $\omega = 0.58$, $M = 16.10$, $SD = 3.62$). It is not uncommon to find such reliability for scales with few items (Lee et al., 2016).

Compassion Towards Others

We utilised the Santa Clara Brief Compassion Scale (SCBCS; Hwang et al., 2008) to evaluate Compassion towards others. This five-item measure is scored on a 7-point Likert scale ranging from 1 = *not at all true of me* to 7 = *very true of me*. A sample item found on the scale is: "I tend to feel compassion for people, even though I do not know them". The scale exhibited very good reliability for the total sample ($\alpha = 0.86$; $\omega = 0.89$, $M = 24.10$, $SD = 7.00$).

Data Analyses

Data Preparation and Partial Credit Model

Data imputation was carried out using IBM SPSS (version 28); the Expectation Maximization (EM) algorithm was employed for this purpose (Dellaert, 2002; Little, 1988). Descriptive statistics and correlational analysis were performed using SPSS. Total scores for all the multi-item scales were calculated, and an examination of Q-Q plots, skewness, and kurtosis (i.e. all within -2 to $+2$) demonstrated normally distributed variables (George & Mallery, 2011). The advanced Rasch analysis utilised RUMM2030 (Andrich et al., 2009), while applying the unrestricted Partial Credit model for parameter estimations (Masters, 1982). This specialised statistical model used in item response theory was suitable for our dataset, as it incorporates varying levels of individual items and responses without assuming uniformity of items. It further allows for modification strategies to improve the overall scale and individual item functioning (Bartholomew et al., 2023; Tennant & Küçükdeveci, 2023).

Overall Model Fit Estimate

Rasch analysis involves an initial assessment of the overall model fit using a chi-square test to check how well items interact with the latent trait. Then, each item's fit to the model is evaluated using item fit residuals, and a chi-square value is calculated for each item. To confirm the Rasch model's overall fit, a non-significant interaction between items and the latent trait ($p > 0.05$) is required (Tennant & Küçükdeveci, 2023; Wilkinson et al., 2023). Individual item fit residuals should fall within the range of -2.50 to $+2.50$, and the residual correlations between individual items below 0.20 (Bartholomew et al., 2023; Christensen et al., 2017). Local dependencies (i.e. item redundancy) can introduce misleading (spurious) correlations affecting the overall measurement and dimensionality. Fortunately, this concern can be effectively handled using testlet creation methodology (i.e. combining multiple individual items into a single, more comprehensive assessment; Lundgren Nilsson & Tennant, 2011; Tennant & Küçükdeveci, 2023).

Invariant Measurement

Differential item functioning (DIF) in Rasch analysis assesses the consistency of a measure across various sample groups such as country, age, sex, and education (i.e. primary, secondary, and tertiary), with the aim of avoiding any DIF in individual items (Sutton & Medvedev, 2023; Tennant & Küçükdeveci, 2023). To examine age group invariance, we applied a standard approach, creating three balanced age groups based on the 33rd and 66th percentiles, ensuring roughly three distinct age groups: 18–29 years, 30–45 years, and 46–89 years. DIF was assessed using between groups ANOVA and visual inspection of individual item plots (Hagquist & Andrich, 2017; Pratscher et al., 2022).

Reliability

The Person Separation Index (PSI) is used to evaluate the scale's reliability and indicates its effectiveness in distinguishing between different levels of an individual's traits. PSI values, on a scale from 0 to 1, are interpreted akin to Cronbach's alpha. Values exceeding 0.70 signify acceptable reliability for group measurements, and values at or above 0.80 indicate suitability for individual assessments (Fisher, 1992).

Unidimensionality

The assessment of unidimensionality in Rasch analysis involves the use of principal components analysis and *t*-tests (Hagell, 2015). Unidimensionality is supported when $\leq 5\%$ of *t*-tests yield statistically significant results when comparing person estimates between sets of items with high and low loadings on the first principal component of residuals (Smith, 2002). Additionally, if the lower boundary of confidence intervals calculated for the number of significant *t*-tests falls within the range of 5%, it indicates unidimensionality. When data adhere to Rasch model assumptions, an ordinal-to-interval transformation table is constructed using person estimates to enhance the precision of the scale (Medvedev et al., 2020). The current study applied the conventional threshold for statistical significance (*p*-value < 0.05).

Convergent and Divergent Validity

We established convergent and divergent validity by computing Pearson's correlations between the SCS-SF interval scores and various measures, including psychological distress (depression, stress, and anxiety), positive and negative

affect, compassion towards others, and life orientation scale (i.e. optimism versus pessimism).

Results

Initial Analysis

Our initial analysis showed the SCS-SF's misfit to the overall Rasch model, as there was evidence of a significant interaction observed between the items and the latent trait of self-compassion ($\chi^2(108) = 260.26$, $p < 0.001$). The SCS-SF demonstrated a reasonable level of reliability with a PSI = 0.65, including no evidence of unidimensionality (Table 1; A1 Initial). Inspection of individual items revealed that Items 1, 7, and 11 displayed a significant misfit to the model, items exceeding -2.50 to $+2.50$ thresholds. The items with their misfitting coefficients are marked with an asterisk in Table 2. Table 2 provides detailed information on individual item fit statistics from the initial analysis, inclusive of item location, fit residual, and Chi-square values for item-trait interaction.

Table 1 Rasch model fit statistics for the initial and final analyses of the SCS-SF ($n = 1000$)

Analyses	Item fit residual		Person fit residual		Goodness of fit		PSI	Unidimensionality t-test	
	Mean	SD	Mean	SD	χ^2 (df)	p		%	Lower bound
A1 Initial	0.16	1.93	-0.78	2.18	260.26 (108)	<0.001	0.65	9.2	7.8% (no)
A2 6 Items	0.03	1.94	-0.66	1.50	141.64 (54)	<0.001	0.51	7.3	5.9% (no)
A3 Final	-0.11	1.00	-0.57	1.05	23.84 (27)	0.64	0.71	4.7	3.3% (yes)

PSI = Person Separation Index without extremes

Table 2 Individual items fit statistics including the initial and final analyses of the SCS-SF ($n = 1000$)

No	Initial analysis: 12 items	Location	Fit residual	Chi Square
1	When I fail at something important to me, I become consumed by feelings of inadequacy*	0.338	3.65*	32.03*
2	I try to be understanding and patient towards those aspects of my personality I don't like	-0.125	0.097	13.62
3	When something painful happens, I try to take a balanced view of the situation	-0.366	0.974	18.95
4	When I'm feeling down, I tend to feel like most other people are probably happier than I am*	0.092	-0.708	6.48
5	I try to see my failings as part of the human condition	-0.081	2.052	22.77
6	When I'm going through a very hard time, I give myself the caring and tenderness I need	-0.120	-1.646	21.59
7	When something upsets me, I try to keep my emotions in balance	-0.369	-1.025	30.16*
8	When I fail at something that's important to me, I tend to feel alone in my failure*	0.300	-0.853	13.33
9	When I'm feeling down, I tend to obsess and fixate on everything that's wrong*	0.132	-1.189	11.31
10	When I feel inadequate in some way, I try to remind myself that feelings of inadequacy are shared by most people	0.257	3.47*	52.48
11	I'm disapproving and judgmental about my own flaws and inadequacies*	0.090	-1.837	24.91*
12	I'm intolerant and impatient towards those aspects of my personality I don't like*	-0.148	-0.999	12.64
Analysis 2: 6 super-items (Si)				
Si1	Items: 2 + 6 (Self-Kindness subscale)	-0.16	-1.55	27.27*
Si2	Items: 11 + 12 (Self-Judgment subscale)	-0.02	-1.31	7.50
Si3	Items: 5 + 10 (common humanity subscale)	0.09	3.76*	51.12*
Si4	Items: 4 + 8 (Isolation subscale)	0.13	-0.54	10.27
Si5	Items: 3 + 7 (Mindfulness subscale)	-0.23	0.20	25.99
Si6	Items: 1 + 9 (Over-identified subscale)	0.19	-0.39	19.49
Final analysis: 3 super-items				
Si1	Items: Si1 + Si4	0.05	-1.26	5.11
Si2	Items: Si2 + Si3	-0.02	0.45	9.47
Si3	Items: Si5 + Si6	-0.03	0.48	9.25

Items with asterisks should be reverse coded before computing the total ordinal scores

Initial Testlet Creation

To enhance the SCS-SF's fit to the Rasch model, we examined the residual correlation matrix, revealing local dependencies between items with correlations surpassing the 0.20 threshold. Such local dependencies can affect the overall fit and dimensionality of a scale. To maintain the scale's validity, we addressed this issue in our subsequent analysis by creating six testlets, aligning with the six subscales of the SCS-SF (self-kindness, self-judgement, common humanity, isolation, mindfulness, over-identification; Table 1: A2 6 Items and Table 2). This combination of items (i.e. items that share higher error variability) aimed to reduce measurement error. However, goodness of fit to the Rasch model was not achieved (χ^2 (54) = 141.64, $p < 0.001$). We achieved a reasonable level of reliability with a PSI of 0.51. The assumption of unidimensionality remained unmet, necessitating further analysis.

Final Analysis

The testlets, self-kindness, and common humanity (Table 2) showed a significant misfit to the model. Further assessment of the residual correlation matrix involving the six testlets revealed persistent local dependency among some testlets. We improved the model further following the same abovementioned procedure to resolve this issue. This involved the creation of three final testlets (self-kindness + isolation, self-judgement + common humanity, mindfulness + over-identification) from the initial six testlets. This modification resulted in achieving overall best fit of the SCS-SF to the Rasch model (χ^2 (27) = 23.84, $p = 0.64$), indicated by the lower bound of significant t-tests (3.3%) overlapping the 5% cut-off point (Table 1: A3 Final); strong evidence of unidimensionality was obtained, including the absence of misfitting items and local dependency. A notable improvement in reliability (PSI = 0.71) was observed at this stage. Figure 2, the item characteristic curve (ICC), illustrates that all testlets were working appropriately across different levels of the latent trait.

DIF, Person-Item Trait, and Ordinal-to-Interval Conversion

Our DIF analysis for age, sex, education (Fig. S1 in Supplementary Information), and country (Fig. 3) indicated no notable differences across any of the derived final testlets. The person-item trait distribution of the final testlets showed no ceiling or floor effects (Fig. 4), demonstrating that 100% of the sample were effectively targeted by items thresholds of the SCS-SF with a person location mean of 0.20 ($SD = 0.72$). The best fit indices of the SCS-SF led to the development of the ordinal-to-interval conversion algorithm, which was based on the Rasch model's person estimates, allowing for the transformation of the ordinal scores into interval-level data. Table 3 provides detailed information about this transformation, including how to use the table and the scores. A paired samples t-test comparing the means of the ordinal ($M = 37.81$; $SD = 5.94$) and Rasch-transformed interval ($M = 36.67$; $SD = 5.60$) scores using the same scale range revealed a true statistical difference between the interval and ordinal scores ($t(1821) = 42.23$, $p < 0.000$), with a large effect size of $d = 1.00$. A significant difference of 0.03 in the standard error was observed, favouring the interval scores.

This conversion table can only be used for complete responses to each of 12-item SCS-SF. To use this table, ordinal raw scores (left column) should be obtained by adding the observed scores for all 12 items. Next, match the ordinal total score (12–60) to the corresponding interval score in the right column (scale 12–60). A final converted score between 12 and 60 will be obtained, with higher scores corresponding to higher levels of self-compassion.

## Convergent and Divergent Validity

Pearson's correlation coefficient analysis revealed positive associations between SCS-SF scores and measures of positive affect ($r = 0.37$, $p < 0.001$), optimism ($r = 0.51$, $p < 0.001$), and compassion towards others ($r = 0.05$, $p = 0.02$). Conversely, negative correlations were observed between SCS-SF scores and measures of negative affect ($r = -0.39$, $p < 0.001$), and psychological distress ($r = -0.43$, $p < 0.001$).

Discussion

We used Rasch methodology to assess the psychometric properties, measurement invariance, and enhanced the measurement precision of the SCS-SF using a sample from four diverse countries. Optimal Rasch model fit was attained for the SCS-SF after combining items with high shared variability into three testlets without removing items from the scale. This was done to preserve the validity of the SCS-SF, mitigate spurious correlations resulting from method effects, and reduce measurement error (Medvedev & Krägeloh, 2022; Wilkinson et al., 2023). These findings were consistent with previous Rasch investigations of the SCS-SF (Finaulahi et al., 2021).

In this study, we combined items with high shared variances unrelated to the overarching latent trait into testlets to effectively reduce spurious correlations and related measurement error. This approach has significant implications for the ongoing debate about the dimensionality of the SCS-SF. Specifically, the high variance shared among

Fig. 2 SCS-SF item characteristic curve (ICC) for the final three testlets
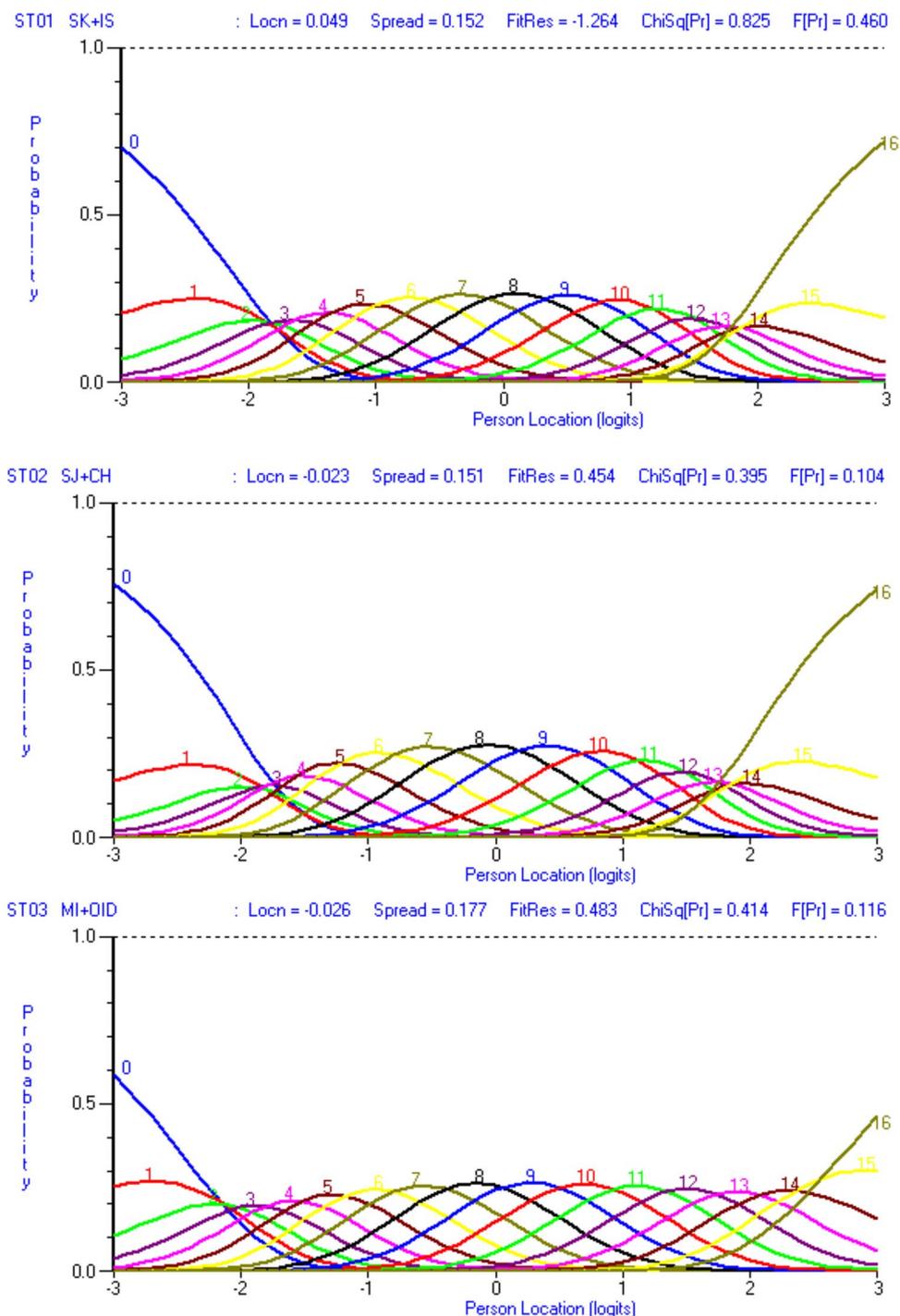
uncompassionate components (e.g. self-judgment, feelings of isolation, and over-identification) supported the findings by Wong et al. (2003) that these items represent a unique underlying latent variable, suggesting they should be treated as a distinct factor rather than simply reversed items. However, the latter observation that both uncompassionate and compassionate components share high variances for testlet creation could also imply the existence

of a common overarching factor that encompasses the six compassionate and uncompassionate components of the scale. This finding aligns with Neff's (2016) perspective, which supports the idea of a unidimensional measure of self-compassion. Arguably, from a psychometric perspective, a construct does not exist if it cannot be measured using the total score. Therefore, this evidence lends further credence to using the unidimensional SCS-SF in research,
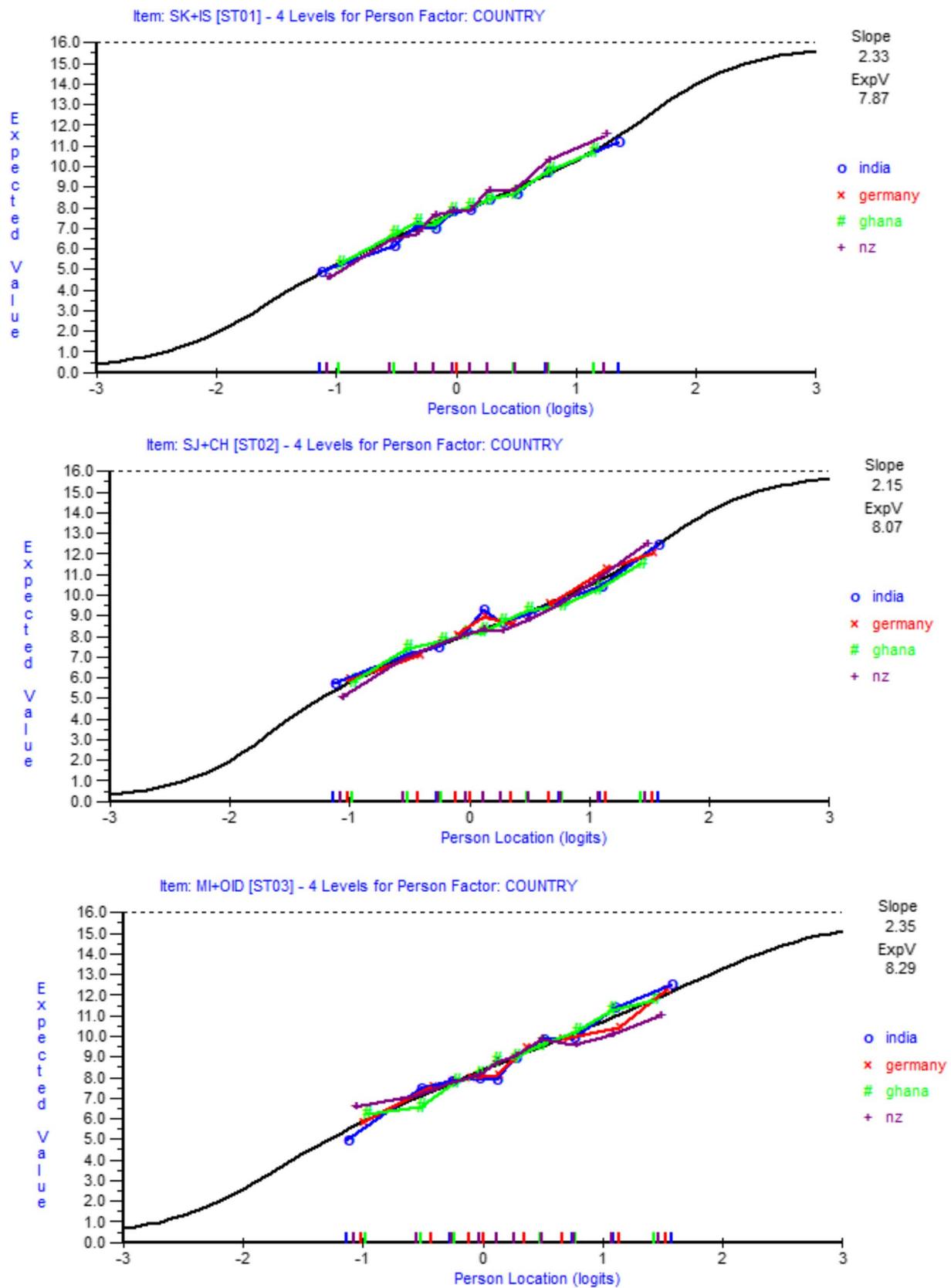
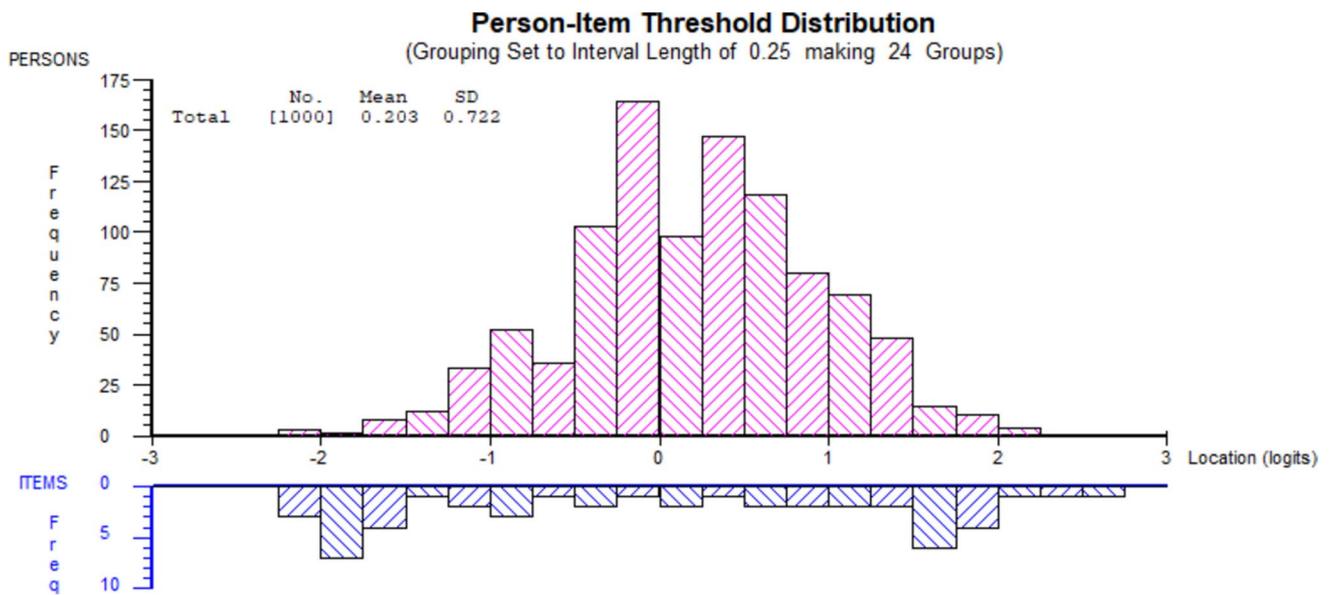Fig. 3 Differential item functioning (DIF) curves for country

Fig. 4 Person-item thresholds distributions for the SCS-SF

Table 3 Ordinal-to-interval conversion for the 12-item SCS-SF

Ordinal scores	Interval		Ordinal scores	Interval	
	logits	Scale		logits	Scale
12	−3.26	12.00	37	0.12	36.04
13	−2.82	15.15	38	0.27	37.11
14	−2.54	17.17	39	0.41	38.17
15	−2.35	18.47	40	0.56	39.21
16	−2.21	19.46	41	0.70	40.22
17	−2.10	20.26	42	0.84	41.21
18	−2.00	20.98	43	0.97	42.14
19	−1.91	21.61	44	1.09	43.01
20	−1.83	22.21	45	1.21	43.81
21	−1.75	22.78	46	1.31	44.55
22	−1.67	23.35	47	1.41	45.23
23	−1.59	23.92	48	1.50	45.88
24	−1.51	24.50	49	1.58	46.49
25	−1.42	25.11	50	1.67	47.09
26	−1.33	25.75	51	1.75	47.69
27	−1.23	26.44	52	1.84	48.29
28	−1.13	27.20	53	1.93	48.93
29	−1.01	28.01	54	2.02	49.61
30	−0.89	28.89	55	2.13	50.39
31	−0.76	29.82	56	2.26	51.29
32	−0.62	30.79	57	2.42	52.41
33	−0.48	31.81	58	2.63	53.93
34	−0.33	32.85	59	2.96	56.29
35	−0.19	33.90	60	3.48	60.00
36	−0.04	34.97			

despite the ongoing debate regarding its dimensionality (Muris & Otgaar, 2020).

The SCS-SF also demonstrated strong sample targeting, meaning that item difficulty levels were appropriately distributed across the range of participants' abilities in the present sample. In essence, the difficulty levels of the items in the scale accurately match the various levels of proficiency and knowledge within our sample (Sutton & Medvedev, 2023; Tennant & Küçükdeveci, 2023). The ICCs confirmed that an item's probability of endorsement varies across different levels of the latent trait being measured, suggesting that the item effectively distinguishes between individuals with varying trait levels (Tennant & Küçükdeveci, 2023; Wilkinson et al., 2023). In other words, items on the SCS-SF effectively discriminate between individuals with differing levels of self-compassion, accurately capturing the range and nuances between levels of self-compassion within the sample, an essential aspect of assessment lacking in CTT methodology. As CTT methods primarily focus on validity and consistency of scores, MTT methods such as the Rasch analysis assess a wider and more detailed array of psychometric properties (Eluwa et al., 2011; Magno, 2009).

The established unidimensionality in the SCS-SF implies that items measure a single overarching latent trait of self-compassion. Hence, a single score obtained from this version of the scale more accurately represents an individual's self-compassion level (Finaulahi et al., 2021; Medvedev & Krägeloh, 2022). The use of the unidimensional SCS-SF is particularly recommended for assessing self-compassion as the factor structure of the self-compassion scales is unclear and often varies between studies in the literature (Muris &

Otgaar, 2020). We observed sound reliability for the SCS-SF that fulfils the conservative criteria for group assessments (PSI ≥ 0.70) as outlined by Tennant and Conaghan (2007). In other words, this version of the Self-Compassion Scale is well-suited for evaluating self-compassion at a group level in research or clinical settings. However, this reliability was not sufficiently high for within-group assessment (i.e. repeated measures or pre- versus post-intervention). Finaulahi et al. (2021) found the SCS-SF to be a reliable measure of both groups and individuals. The sightly varying results observed between these two studies could potentially be attributed to differences in the samples and languages across the studies. These variations in findings further complicate the controversies about the reliability and validity of the SCS-SF across samples (Muris & Otgaar, 2020).

Furthermore, it is essential to emphasise the attainment of measurement invariance for the SCS-SF in our study across four countries, and other sociodemographic factors such as age, sex, and education level. This underscores the scale's strength in its ability to be used across a wide spectrum of individuals, spanning various countries, age groups, sexes, and educational backgrounds especially following the statistically significant difference observed between these sociodemographic factors across countries. Measurement invariance increases the applicability, acceptability, and robustness of the SCS-SF, suggesting that study outcomes stemming from the use of this scale can be confidently compared (Welzel et al., 2023). Finaulahi et al. (2021) similarly confirmed the invariance of this scale, yet this pertained specifically to age and sex. Available studies using a CFA approach established similar measurement invariance for this scale, but the original factor structure of the SCS-SF was not achieved (Meng et al., 2019).

Moreover, we utilised Rasch methodology to transform the ordinal scores of the SCS-SF into interval-level data, acknowledging the presence of varying intervals between response categories (Magno, 2009; Pratscher et al., 2022). This method provides a real-life precision in measurement (Magno, 2009; Tennant & Küçükdeveci, 2023) diverging from the conventional hierarchical assumption among response categories prevalent in CTT (Courville, 2004). Notably, the interval scores were found to exhibit reduced measurement error compared to the ordinal scores, signifying that the interval scores provide a more precise and less variable estimation of scores compared to the ordinal scores (Bartholomew et al., 2023). Interval transformation enhances score precision, ensuring a more accurate representation of individual responses in a group in research or clinical settings (Barber et al., 2022; Medvedev et al., 2018). Furthermore, interval-level data is appropriate for use with parametric statistical tests, as they do not violate their underlying assumptions. Below is an illustration of how the interval scores demonstrate advantage over the ordinal scores.

Imagine person A's initial score was 20 and person B's initial score was 30 before taking part in a self-compassion intervention. Following the intervention, person A's score rose to 35, and person B's score increased to 45. Solely relying on ordinal scores might suggest that both individuals experienced a similar level of change in their self-compassion levels. However, Rasch interval scores present a different scenario. Person A's score increased by 11.69 units, while person B's score increased by 14.92 units (Table 3). Despite the seemingly comparable changes, person B's transformation was more than person A's, indicating a quite different outcome that may be clinically significant. This emphasises the accurate measurements of the Rasch interval scores in group studies to better discern authentic changes in attitudes and behaviours.

Additionally, we established the convergent validity of the SC-SF scores, demonstrating a positive correlation with related measures such as positive affect, optimism, and compassion towards others. Past research consistently indicates that self-compassion has a significant direct connection with self-reported measures of positive affect, optimism, and compassion towards others (Neff et al., 2007). Notwithstanding, there is strong evidence regarding the negative association between self-compassion and psychological distress, as well as negative affect (Medvedev et al., 2021; Shapira & Mongrain, 2010). While this finding aligns with our study results and supports the external validity of the SCS-SF, the evidence of divergent validity was not present. This evidence suggests that the SCS-SF is an accurate, relevant, and applicable scale for measuring self-compassion (Stöber, 2001).

Limitations and Future Research

While our study utilised samples from four distinct locations, signifying the robustness of our findings, it is essential to note that our instruments were predominately administered in the English language across three of these countries, representing disparate ethnocultural groups. Of note, the reliability coefficient of the current scale was not high enough for within-group assessments. Considering the potential cultural and language influences in responding to scale items, additional research involving diverse participant groups and translated versions of the SCS-SF using MTT is necessary to ascertain cross-cultural consistency, applicability, and the overall robustness of this scale. Another inconsistency involved the slightly different approaches used in recruiting the samples (e.g. rewards). The present study primarily involved a non-clinical sample, emphasising the need for future research to validate these results in clinical settings, particularly among groups affected by mood disorders or other psychological health conditions.

In summary, we used Rasch methodology to assess the psychometric properties of the SCS-SF across four distinct

countries. The SCS-SF exhibited a strong fit to the Rasch model and demonstrated unidimensionality. The SCS-SF remained consistent across various demographic factors such as country, age, sex, and educational background. We then developed an algorithm for converting ordinal to interval-level data, thereby enhancing the measurement precision of the scale. The unidimensional SCS-SF was found to be well-suited for assessing group-level self-compassion, and displayed strong convergent and divergent validity. While our large sample size increases confidence in our findings, we encourage further research to provide similar evidence using diverse ethnic groups and clinical samples to further strengthen and broaden the universal tenability of the scale.

Data Availability 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

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 analysed with others).

Ethics Approval The study received approval from the Human Research Ethics Committee at Victoria University of Wellington, New Zealand (#0000029770). The study was also in line with the Declaration of Helsinki, which outlines fundamental ethical principles for health research involving the use of human participants (World Medical Association, 2001).

Conflict of Interest The authors declare no competing interests.

References

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, 4*(1), 319–329. https://doi.org/10.1027/2698-1866/a000064

Adu, P., Popoola, T., Collings, S., Aspin, C., Medvedev, O. N., & Simpson, C. R. (2024a). Psychometric properties of the Motors of COVID-19 Vaccination Acceptance Scale (MoVac-COVID19S) in New Zealand: Insights from confirmatory factor analysis. *Current Psychology, 43*, 26628–26638. https://doi.org/10.1007/s12144-024-05877-x

Adu, P., Popoola, T., Iqbal, N., Roemer, A., Collings, S., Aspin, C., Medvedev, O. N., & Simpson, C. R. (2024b). Cross-country assessment of the unique contributions of psychological factors to vaccination: Perspectives on the COVID-19 pandemic. *Journal of Health Psychology.* https://doi.org/10.1177/13591053241266592

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

Babenko, O., & Guo, Q. (2019). Measuring self-compassion in medical students: Factorial validation of the Self-Compassion Scale-Short Form (SCS-SF). *Academic Psychiatry, 43*, 590–594. https://doi.org/10.1007/s40596-019-01095-x

Barber, C. C., Middlemiss, W., & Medvedev, O. N. (2022). Applying Rasch methodology to examine and enhance precision of the Edinburgh postnatal depression scale. *Journal of Affective Disorders, 308*, 391–397. https://doi.org/10.1016/j.jad.2022.04.009

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

Braun, T. D., Park, C. L., & Gorin, A. (2016). Self-compassion, body image, and disordered eating: A review of the literature. *Body Image, 17*, 117–131. https://doi.org/10.1016/j.bodyim.2016.03.003

Christensen, K. B., Makransky, G., & Horton, M. (2017). 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/10.1177/0146621616677775

Costa, J., Marôco, J., Pinto-Gouveia, J., Ferreira, C., & Castilho, P. (2016). Validation of the psychometric properties of the Self-Compassion Scale. Testing the factorial validity and factorial invariance of the measure among borderline personality disorder, anxiety disorder, eating disorder and general populations. *Clinical Psychology & Psychotherapy, 23*(5), 460–468. https://doi.org/10.1002/cpp.1974

Courville, T. G. (2004). *An empirical comparison of item response theory and classical test theory item/person statistics* (Publication No. 305067822) [Doctoral dissertation, Texas A&M University]. ProQuest One Academic.

Dellaert, F. (2002). The expectation maximization algorithm. *College of Computing, Georgia Institute of Technology*.

Ellis, B. B., & Mead, A. D. (2004). Item analysis: Theory and practice using classical and modern test theory. In *Handbook of research methods in industrial and organizational psychology* (pp. 324–343). https://doi.org/10.1002/9780470756669.ch16

Eluwa, O. I., Eluwa, A. N., & Abang, B. K. (2011). Evaluation of mathematics achievement test: A comparison between classical test theory (CTT) and item response theory (IRT). *Journal of Educational and Social Research, 1*(4), 99–106.

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

Fisher, W., Jr. (1992). Reliability, separation, strata statistics. *Rasch Measurement Transactions, 6*(3), 238.

George, D., & Mallery, P. (2011). *SPSS for Windows step by step: A simple study guide and reference, 17.0 update* (10th ed.). Pearson Education India.

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

Hambleton, R. K. (1994). Item response theory: A broad psychometric framework for measurement advances 1, 2. *Psicothema, 6*(3), 535–556.

Hartstone, J. M., & Medvedev, O. N. (2021). The role of mindfulness and life satisfaction in psychological distress during the COVID-19 lockdown in New Zealand: A quasi-experimental study. *Mindfulness, 12*(11), 2693–2706. https://doi.org/10.1007/s12671-021-01731-4

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

Hwang, Y.-S., Medvedev, O. N., Krägeloh, C., Hand, K., Noh, J.-E., & Singh, N. N. (2019). The role of dispositional mindfulness and self-compassion in educator stress. *Mindfulness, 10*(8), 1692–1702. https://doi.org/10.1007/s12671-019-01183-x

Koğar, E. Y., & Koğar, H. (2023). A validation study of the Self-Compassion Scale-Short Form (SCS-SF) with ant colony optimization in a Turkish sample. *Psihologija, 56*(1), 89–114. https://doi.org/10.2298/PSI211127015K

Lee, C.-T., Lin, C.-Y., Tsai, M.-C., Strong, C., & Lin, Y.-C. (2016). Psychometric evaluation and wording effects on the Chinese version of the parent-proxy Kid-KINDL. *Health and Quality of Life Outcomes, 14*, 123. https://doi.org/10.1186/s12955-016-0526-3

Lefebvre, J.-I., Montani, F., & Courcy, F. (2020). Self-compassion and resilience at work: A practice-oriented review. *Advances in Developing Human Resources, 22*(4), 437–452. https://doi.org/10.1177/152342232094914

Lefever, S., Dal, M., & Matthíasdóttir, Á. (2007). Online data collection in academic research: Advantages and limitations. *British Journal of Educational Technology, 38*(4), 574–582. https://doi.org/10.1111/j.1467-8535.2006.00638.x

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

Lluch-Sanz, C., Galiana, L., Vidal-Blanco, G., & Sansó, N. (2022). Psychometric properties of the Self-Compassion Scale—Short Form: Study of its role as a protector of Spanish nurses professional quality of life and well-being during the COVID-19 pandemic. *Nursing Reports, 12*(1), 65–76. https://doi.org/10.3390/nursrep12010008

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

Lundgren Nilsson, A., & Tennant, A. (2011). Past and present issues in Rasch analysis: The functional independence measure (FIM™) revisited. *Journal of Rehabilitation Medicine, 43*(10), 884–891. https://doi.org/10.2340/16501977-0871

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.

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

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

Medvedev, O. N., Turner-Stokes, L., Ashford, S., & Siegert, R. J. (2018). Rasch analysis of the UK Functional Assessment Measure in patients with complex disability after stroke. *Journal of Rehabilitation Medicine, 50*(5), 420–428. https://doi.org/10.2340/16501977-2324

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

Medvedev, O. N., Cervin, M., Barcaccia, B., Siegert, 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

Meng, R., Yu, Y., Chai, S., Luo, X., Gong, B., Liu, B., Hu, Y., Luo, Y., & Yu, C. (2019). Examining psychometric properties and measurement invariance of a Chinese version of the Self-Compassion Scale–Short Form (SCS-SF) in nursing students and medical workers. *Psychology Research and Behavior Management, 12*, 793–809. https://doi.org/10.2147/PRBM.S216411

Muris, P., & Otgaar, H. (2020). The process of science: A critical evaluation of more than 15 years of research on self-compassion with the Self-Compassion Scale. *Mindfulness, 11*(6), 1469–1482. https://doi.org/10.1007/s12671-020-01363-0

Neff, K. (2003a). Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity, 2*(2), 85–101. https://doi.org/10.1080/15298860309032

Neff, K. D. (2003b). The development and validation of a scale to measure self-compassion. *Self and Identity, 2*(3), 223–250. https://doi.org/10.1080/15298860309027

Neff, K. D. (2016). The Self-Compassion Scale is a valid and theoretically coherent measure of self-compassion. *Mindfulness, 7*(1), 264–274. https://doi.org/10.1007/s12671-015-0479-3

Neff, K. D., & Tóth-Király, I. (2022). Self-compassion scale (SCS). In O. N. Medvedev, C. U. Krägeloh, R. J. Siegert, & N. N. Singh (Eds.), *Handbook of assessment in mindfulness research*. Springer International Publishing. https://doi.org/10.1007/978-3-030-77644-2_36-1

Neff, K. D., Rude, S. S., & Kirkpatrick, K. L. (2007). An examination of self-compassion in relation to positive psychological functioning and personality traits. *Journal of Research in Personality, 41*(4), 908–916. https://doi.org/10.1016/j.jrp.2006.08.002

Neff, K. D., Tóth-Király, I., Knox, M. C., Kuchar, A., & Davidson, O. (2021). The development and validation of the Self-Compassion Scale (long-and short form). *Mindfulness, 12*(1), 121–140. https://doi.org/10.1007/s12671-020-01505-4

Pelton, T. (2002). *Where are the limits to the Rasch advantage?* Paper presented at the International Objective Measurement Workshop.

Pratscher, S. D., Oyler, D. L., Bettencourt, B. A., & Medvedev, O. N. (2022). Interpersonal Mindfulness Scale-Short Form development using Rasch analyses. *Mindfulness, 13*(8), 2007–2021. https://doi.org/10.1007/s12671-022-01937-0

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

Rahman, M. M., Huq, M. M., Shakib, S., & Shahrier, M. A. (2023). Psychometric evaluation of the Bangla version of the Self-Compassion Scale-Short Form (SCS-SF). *Mindfulness*, *14*(11), 2795–2806. https://doi.org/10.1007/s12671-023-02237-x

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A re-evaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67*(6), 1063–1078. https://doi.org/10.1037/0022-3514.67.6.1063

Shapira, L. B., & Mongrain, M. (2010). The benefits of self-compassion and optimism exercises for individuals vulnerable to depression. *The Journal of Positive Psychology, 5*(5), 377–389. https://doi.org/10.1080/17439760.2010.516763

Smith Jr, E. V. (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.

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment, 17*(3), 222–232. https://doi.org/10.1027/1015-5759.17.3.222

Sutton, A., & Medvedev, O. N. (2023). Development and validation of the Awareness Outcomes Measure (AOM) using Rasch approach. *Mindfulness, 14*(2), 473–481. https://doi.org/10.1007/s12671-022-02047-7

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

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/10.1177/0049124121995

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

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the Material Values Scale. *Journal of Consumer Research, 30*(1), 72–91. https://doi.org/10.1086/374697

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.

Authors and Affiliations

Peter Adu[1] · Tosin Popoola[2] · Emerson Bartholomew[3] · Naved Iqbal[4] · Anja Roemer[5] · Tomas Jurcik[6] ·
Sunny Collings[1] · Clive Aspin[1] · Oleg N. Medvedev[7] · Colin R. Simpson[1,8]

✉ Peter Adu
adupe@staff.vuw.ac.nz

Tosin Popoola
Tosin.Popoola@newcastle.edu.au

Emerson Bartholomew
emerson.bartholomew@auckland.ac.nz

Naved Iqbal
niqbal@jmi.ac.in

Anja Roemer
A.Roemer@massey.ac.nz

Tomas Jurcik
dr.tomas.jurcik@gmail.com

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, Wellington, New Zealand

[2]   School of Nursing and Midwifery, College of Health,
       Medicine and Wellbeing, The University of Newcastle,
       Callaghan, Australia

[3]   Department of Psychological Medicine, Faculty of Medical
       and Health Sciences, University of Auckland, Auckland,
       New Zealand

[4]   Jamia Millia Islamia, New Delhi, India

[5]   School of Psychology, Massey University, Palmerston North,
       New Zealand

[6]   University of British Columbia, Vancouver, Canada

[7]   School of Psychological and Social Sciences, University
       of Waikato, Hamilton, New Zealand

[8]   Usher Institute, The University of Edinburgh, Edinburgh, UK