Mindfulness (2021) 12:730–740
https://doi.org/10.1007/s12671-020-01539-8

ORIGINAL PAPER

Investigating Psychometric Properties of the Self-Compassion Scale Using Rasch Methodology

Kalolaine P. Finaulahi[1] · Alexander Sumich[2] · Nadja Heym[2] · Oleg N. Medvedev[1]

Accepted: 27 October 2020 / Published online: 2 January 2021

Abstract

Objectives The 26-item Self-Compassion Scale (SCS) and its short 12-item version (SCS-SF) were reported to have acceptable psychometric properties, and both scales are widely used to assess self-compassion in individuals. However, recent investigations were inconsistent regarding factor structure of the SCS, and psychometric properties of the scale were not tested for consistency with principles of fundamental measurement using appropriate methodology such as Rasch analysis.

Methods A partial credit Rasch model was used to evaluate psychometric properties of the SCS and SCS-SF with the sample of 743 respondents randomly divided into two equal subsamples (A and B) to replicate the results for the purpose of robustness.

Results Initially, there were no misfitting items but the local dependency between various items affected Rasch model fit. This issue was resolved by combining locally dependent items into four super-items resulting in the best fit to the Rasch model of both SCS and SCS-SF, with evidence of unidimensionality and an excellent sample targeting. Although both scale versions had strong reliability satisfactory for individual and group assessment, the original SCS demonstrated superior psychometric properties reflected by higher reliability indicated by Person Separation Index (PSI) of 0.90 compared to the SCS-SF (PSI = 0.85). These analyses were replicated with the sample B for both scale versions, resulting in equally good fit. This permitted generating ordinal-to-interval conversion tables based on Rasch model estimates.

Conclusions The current study supported the reliability and internal validity of both the SCS and SCS-SF. Accuracy of these assessment instruments can be further improved by using the ordinal-to-interval conversion tables published here.

Keywords Mindfulness · Measurement · Self-Compassion Scale · Rasch analysis · Reliability

Compassion, from Buddhist philosophy, refers to "the wish for all sentient beings to be free from suffering and its causes" (Shonin et al. 2014, p. 1162) and "…the urge to help in some way" (Neff and Dahm 2015, p. 121). According to Neff and Dahm (2015), compassion is extended to our own, as well as other's suffering. Thus, self-compassion refers to the inward direction of those feelings and motivations (Neff 2003b; Neff and Dahm 2015). Here, individual suffering is experienced through non-avoidance and open understanding of one's own pain, inadequacies and failings, alongside its full acceptance, in a manner that is non-judgemental and distinct from self-pity, selfishness or self-centeredness (Neff 2003b). Neff proposed three distinct interacting subconstructs within self-compassion: self-kindness, common humanity and mindfulness (Neff 2003a, b; Neff and Dahm 2015). Based on this conceptualisation, a 26-item unidimensional Self-Compassion Scale (SCS; Neff 2003a, 2016) and its short-form 12-item version (SCS-SF; Raes et al. 2011) have been developed. A rich literature has ensued highlighting the centrality of self-compassion to well-being (Durkin et al. 2016; Neff and Dahm 2015; Shrestha 2016; Sinclair et al. 2017). However, there is ongoing debate regarding the psychometric properties of the SCS and the most efficient way to evaluate this construct.

According to Neff and Dahm (2015), self-kindness means that rather than rebuking ourselves for failures or being harshly critical of personal shortfalls (self-judgement), we kindly acknowledge that we are performing as well as we can, and soothe/nurture ourselves during times of difficulty and adversity. Compassion for our own distress motivates us to

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

[1] School of Psychology, University of Waikato, Gate 1, Knighton Road, Hamilton, New Zealand

[2] School of Social Sciences, Nottingham Trent University, Nottingham, UK

ameliorate our own suffering. However, rather than the autocentric "woe is me" view present during self-pity that may lead to a sense of isolation, self-compassion recognises that we all suffer. By fostering an inclusive and connected mindset through common humanity and reflecting on the shared human experience, we feel less isolated when we are in pain. Mindful awareness of our negative thoughts and emotions allows us to be experientially open to reality and approach ourselves compassionately in the moment, without *over-identifying* with a negative self-concept.

Neff initially proposed that these conceptually distinct, yet interacting, components of self-compassion can be assessed using positively worded (self-kindness, common humanity and mindfulness) and juxtaposed negatively worded (self-judgement, isolation and over-identification, representing uncompassionate behaviour) items (Neff 2003a). These subscales can be interpreted individually or within a unitary self-compassion score. However, recent evidence supporting the multi-dimensionality of the SCS suggested separate constructs for negatively and positively worded subscales. For example, whilst Kumlander et al. (2018) found acceptable fit for the 6-factor model, they noted that the negative items correlated more strongly with each other, suggesting a method effect, which is a systematic influence due to the properties of the scale/items. Specifically, they proposed that at least two separate latent constructs, namely self-compassion and self-criticism, are measured by the scale. López et al. (2015) examined the factor structure, reliability and construct validity of SCS in a large community sample, applying classical test theory (CTT) approach. Using confirmatory factor analysis (CFA), the 6-factor model could not be replicated, whereas a subsequent exploratory factory analysis (EFA) suggested a 2-factor model comprising positively (self-compassion) and negatively (self-criticism) worded items (López et al. 2015). Similarly, Costa et al. (2016) applied CFA in four different samples (borderline personality disorder, anxiety disorder, eating disorder and general population) and could not replicate the 6-factor structure. Consequently, they also reported a 2-factor model—self-compassionate attitude and self-critical attitude—as best fit across all groups. Moreover, self-compassion and self-criticism/self-coldness have been differentiated using physiological measures (Gilbert et al. 2011). Together these findings have formed the basis for the over-arching argument against unidimensionality of SCS that comprises both positive and negative affective traits.

The majority of studies examining psychometric properties of the SCS to date have used CTT, which is limited in distinguishing between true multidimensionality and spurious correlations attributable to method effects (e.g. due to the scale/item properties). Indeed, Kumlander et al. (2018) found strong correlations between negatively worded items, suggesting a potential method effect due to reverse coding. Recently, Neff et al. (2019) investigated the factor structure of the SCS using statistically adequate methodology and samples and found satisfactory fit for the correlated 6-factor model and the single bi-factor model, which explained 95% of the true item variance. These results provided support for the over-arching latent trait of self-compassion operationalised by the SCS items, which can be further verified using more advanced methodology such as Rasch analysis. Rasch methodology is more specifically suited to differentiate true multidimensionality from spurious correlations (Lundgren-Nilsson and Tennant 2011; Medvedev et al. 2017a, b, 2018). Rasch analysis is a robust probabilistic psychometric method that applies an iterative process allowing for strategies to improve the overall scale and individual item functioning (Balalla et al. 2019). These advantages of Rasch analysis over CTT have been demonstrated across various samples, scales and disciplines (e.g. medicine, rehabilitation, psychology and education; Lundgren-Nilsson and Tennant 2011; Medvedev et al. 2017a, b; Norquist et al. 2004).

The Rasch model contains a set of fitness criteria, which can be applied to the SCS and SCS-SF to investigate characteristics of each individual item and subscale, and their unique contributions to measuring an over-arching trait of self-compassion as a unidimensional construct. Rasch holds advantages over CTT, because it can estimate the location or difficulty of every item, test the order of response options of the items that are polytomous and transform ordinal-level data to an interval-level scale to enhance reliability (Hobart and Cano 2009; Nijsten et al. 2006; Rasch 1960; Wilson 2004; Wright and Stone 1979). The process involves taking ordinal-level data and analysing responses whilst considering respondents' ability or traits as well as item difficulty or location (Tennant and Conaghan 2007). It also involves dealing with locally dependent items, which occur when responses to one item influence responses to a different item (e.g. method effect). An effective method to deal with local dependency is to create super-items by combining locally dependent items to improve the model fit (Medvedev et al. 2018). Removal of items is an alternative, but the last resort because it may affect the validity of the construct (Pratscher et al. 2019). When Rasch model criteria are met, transformation from the ordinal-level data into interval-level scale can be achieved, which improves precision of instruments to better assess latent traits (Medvedev et al. 2017a, b). Rasch methods are aligned with the fundamental measurement principles laid out by Thurstone (1931), which imply that (i) scales should not discriminate between users (e.g. sex differences), (ii) the model should measure one parameter (unidimensionality) and (iii) the units of the scale should be equally proportioned, such as an interval-level scale.

Recent studies used CTT methodology and were inconsistent regarding factor structure of the SCS because CTT is unable to distinguish between true factors and spurious correlations due to method effect. The aim of the current study was

to apply modern Rasch methodology to investigate psychometric properties and dimensionality of the SCS and SCS-SF, and compliance of these instruments with the principles of fundamental measurement. Essentially, we tested the hypothesis that spurious correlations between SCS items were due to local dependency, which refers to a method effect rather than multidimensionality. Therefore, if modified SCS fit the unidimensional Rasch model, it will provide evidence for unidimensionality as a multidimensional measure cannot fit the Rasch model. If satisfactory fit to the Rasch model is achieved, ordinal-to-interval conversion tables can be produced from the Rasch model estimates to enhance precision of SCS.

Method

Participants

Participants were recruited as part of a larger data collection through several means (e.g. university research participation scheme, radio advertising, local posters and word-of-mouth), including an online recruitment platform (Mechanical Turk (MTurk), https://www.mturk.com/). In all cases, participants were directed to an online survey for a study on nutrition and mood (Heym et al. 2019). Participants provided online informed consent prior to completing the survey. The sample included $n = 743$ participants, $n = 246$ males (31.8%) and $n = 6$ participants who did not provide sex information. The majority (72.2%) identified as 'White British', with 9.3% identified as 'Other White' background, 1.2% as 'Irish', 1.6% as 'White and Black Caribbean' and 15.7% as 'Other ethnicity'. Age ranged from 18 to 77 years (mean = 30.44; SD = 11.33) and was normally distributed, with skewness and kurtosis ranging between 0.28 and 0.98. For differential item functioning (DIF) testing, 3 age categories were created, each comprising 33% of the total samples (18–21, 22–33 and 34–80).

Procedures

Online data collection was initially conducted via Bristol online surveys and then moved to Qualtrics (https://www.qualtrics.com/) due to departmental policy. Participants completed online consent and provided a unique code to protect anonymity. Participation was voluntary. Participants who completed the survey through MTurk were paid US$0.50 for their time. Those that completed the survey through other platforms did not receive any direct payment; however, they had an opportunity to win Amazon vouchers through a raffle or receive research credits (e.g. psychology students from the institution). Initial items in the questionnaire asked for demographic-related (sex, age, ethnicity, country of residence, socio-economic status, education) and health-related (diet, history of illness) information. Participants then completed a series of scales on depression and related constructs, including the SCS (Neff 2003a). Of these, the SCS is currently reported.

Measures

The SCS (Neff 2003a) is a 26-item self-report questionnaire of self-compassion comprising six subscales including self-kindness (5 items; $\alpha = 0.84$; "I try to be loving towards myself when I'm feeling emotional pain"), self-judgement (5 items; $\alpha = 0.83$; "I'm disapproving and judgmental about my own flaws and inadequacies"), common humanity (4 items; $\alpha = 0.80$; "When things are going badly for me, I see the difficulties as part of life that everyone goes through"), isolation (4 items; $\alpha = 0.81$; "When I think about my inadequacies, it tends to make me feel more separate and cut off from the rest of the world"), mindfulness (4 items; $\alpha = 0.79$; "When something upsets me I try to keep my emotions in balance") and over-identification (4 items; $\alpha = 0.83$; "When I'm feeling down I tend to obsess and fixate on everything that's wrong"). The SCS-SF (Raes et al. 2011) is the 12-item scale version which consists of the same subscales as the SCS. Both questionnaires use a 5-point Likert-scale response format (1 = 'Almost Never' to 5 = 'Almost Always'). Subscale scores are obtained by calculating the mean of the subscale item responses. The total SCS and SCS-SF score can be found after reverse scoring negative items (self-judgement, isolation and over-identification), then calculating the total mean (Neff 2003a).

Data Analyses

Descriptive statistics, loadings on the first principal component, item-to-total correlations and internal consistency of the SCS were computed using IBM SPSS v.25, and Rasch analysis employed RUMM2030 (Andrich et al. 2009). The total sample was split randomly into two groups, sample A ($n = 372$) and sample B ($n = 371$) for the purpose of replication. Each sample size is appropriate for Rasch analysis of 26-item scales using RUMM2030 (Andrich et al. 2009), which should be between 250 and 500 cases to achieve a balance between minimising type I errors whilst having a sufficient number of cases for item calibrations (Hagell and Westergren 2016). A likelihood-ratio test was conducted and showed significant differences in response option thresholds of individual items across the scale items ($p < 0.01$), which means that the unrestricted partial credit model (Masters 1982) will be more appropriate for the current data (Lundgren-Nilsson and Tennant 2011).

Rasch analysis follows a sequential and logical process of iteration that is set out by Leung et al. (2013). This starts with evaluating the threshold ordering and identifying any disordered thresholds. Thresholds are estimated individual scores on a construct when the chances of choosing either of two

consequent response options are the same (Andrich 1978). When thresholds are disordered, individual scores on a construct do not increase progressively with response options (Andrich 1978). Item location mean is always set to zero, and an expected person location mean is ± 0.50, which means that respondents' abilities are covered by the scale range. Next, individual item fit to the Rasch model is tested using residual statistics. A perfect fit for items and respondents is seen when the residual is equal to 0 (SD = 1), and individual fit residuals should range from − 2.50 to + 2.50. An overall model fit is reported with chi-square. Good fit to the Rasch model occurs when item-trait interaction is not significant ($p > 0.05$). A DIF is also assessed. DIF arises when there are distinct individual items/traits levels displayed by sample groups such as age and sex with the same levels of the underlying latent trait. To test for DIF and distinguish method effects, a post hoc test of significance is used (Balalla et al. 2019). A Person Separation Index (PSI) tests the scales' ability to distinguish the levels of individual traits. A reliability coefficient is produced (Fisher 1992), which is interpreted in a similar way to Cronbach's alpha; however, instead of internal consistency, it identifies how well individuals are spread on the scale (Fisher 1992). An independent samples *t* test comparison of the person estimates for a group of items with the highest negative and highest positive loadings on the first principal component was used to determine dimensionality following the methodology of Smith (2002). Evidence of unidimensionality requires the number of significant *t* test comparisons not exceeding 5% or lower bound of binominal confidence interval computed around percentage of significant *t* tests (Tennant and Pallant 2006).

Residual correlations between items can affect unidimensionality. One way to address this is to examine local response dependency using the residual correlation matrix. Essentially, there should be no evidence for local dependency between items (Christensen et al. 2017; Marais and Andrich 2008). That is, the amount of residual correlations should not exceed the mean of all residual correlations by 0.20 (Christensen et al. 2017). If local dependency is present, items can be added together to create super-items (Lundgren-Nilsson and Tennant 2011). When the criteria of Rasch model are met, the ordinal scale scores can be transformed into interval-level data based on Rasch model person estimates. Throughout this study, the conventional criteria for statistical significance of $p < 0.05$ were applied.

Results

We have evaluated the potential suitability of the SCS data for the unidimensional Rasch model from CTT perspective by examining loadings of all SCS items on the first principal component, which were ranging from 0.52 to 0.71 and supported the overarching latent construct of self-compassion. In addition, we have computed item-to-total correlations for items of the full SCS that were all above 0.49 with the overall Cronbach's alpha of 0.94, further supporting the overarching self-compassion trait. Moreover, removing any of the items did not improve but weaken initially strong internal consistency of the full scale. This evidence was consistent across both subsamples and sufficiently robust, making the SCS a perfect candidate for Rasch analysis.

Rasch Analysis of the SCS (Sample A)

The initial analysis (sample A, Table 1) shows good reliability with PSI = 0.94. However, there was misfit to the Rasch model due to a significant interaction between items and the latent trait of self-compassion (χ^2 (130) = 196.46, $p < 0.001$).

Table 1 Summary of fit statistics for the initial and the final Rasch analyses of the 26-item Self-Compassion Scale and 12-item short version (sample A [$n = 372$] and sample B [$n = 371$])

Analyses	Person mean		Goodness of fit		PSI	Significant *t* tests (unidimensionality)	
	Value	SD	χ^2 (*df*)	*p*		%	Lower bound
26-item SCS (sample A)							
Initial (26)	− 0.26	0.92	196.46 (130)	0.00	0.94	30.7	28.4 (no)
Final (26)	− 0.19	0.58	21.02 (20)	0.40	0.90	5.1	2.9 (yes)
26-items SCS (sample B)							
Initial (26)	− 0.20	0.89	179.03 (130)	0.00	0.94	26.4	24.1 (no)
Final (26)	− 0.09	0.55	22.52 (20)	0.31	0.90	6.7	4.5 (yes)
12-item SCS version (sample A)							
Initial (12)	− 0.29	0.95	100.38 (60)	0.00	0.87	16.4	14.2 (no)
Final (12)	− 0.32	0.84	23.29 (20)	0.27	0.85	4.3	2.1 (yes)
12-item SCS version (sample B)							
Initial (12)	− 0.23	0.97	102.85 (60)	0.00	0.87	14.0	11.8 (no)
Final (12)	− 0.17	0.79	18.13 (20)	0.58	0.84	3.8	1.7 (yes)

Examination of thresholds showed no significant signs of disordering on any of the SCS items. Individual item fit statistics from the initial analysis are presented in Table 2 including item location, fit residual and chi-square for the item-trait interaction showing no items with significant misfit. The residual correlation matrix was examined and displayed local dependencies between items with a correlation above the 0.20 requirement (Christensen et al. 2017), and unidimensionality was not confirmed.

Both the overall fit and dimensionality can be affected by local dependency (residual correlations between items). Instead of removing these locally dependent items, super-items can be created using locally dependent items that exceed correlations by 0.20 (Lundgren-Nilsson et al. 2013). Therefore, a second analysis was conducted where items that were identified as locally dependent on each other were combined to create 6 super-items, reflecting 6 subscales of the SCS (self-kindness, self-judgement, common humanity,

Table 2 Rasch model fit statistics including item locations, fit residuals and chi-square for the 26-item version of the SCS (sample A) before creating super-items

Facets/item no.	Items	Item location	Item fit residual	chi-square
Self-kindness				
5	I try to be loving towards myself when I'm feeling emotional pain	− 0.31	2.62	9.72
12	When I'm going through a very hard time, I give myself the caring and tenderness I need	0.14	− 0.86	7.02
19	I'm kind to myself when I'm experiencing suffering	0.01	− 2.37	11.68
23	I'm tolerant of my own flaws and inadequacies	0.05	0.83	6.78
26	I try to be understanding and patient towards those aspects of my personality I don't like	− 0.09	− 0.18	3.09
Self-judgement				
1	I'm disapproving and judgemental about my own flaws and inadequacies[R]	0.58	− 0.71	16.72
8	When times are really difficult, I tend to be tough on myself[R]	0.59	− 0.43	18.17
11	I'm tolerant and impatient towards those aspects of my personality I don't like[R]	0.15	2.63	20.13
16	When I see aspects of myself that I don't like, I get down on myself[R]	0.29	− 0.09	4.44
21	I can be a bit cold-hearted towards myself when I'm expecting suffering[R]	0.46	− 0.50	2.21
Common humanity				
3	When things are going badly for me, I see the difficulties as part of life that everyone goes through	− 1.05	1.16	7.75
7	When I'm down, I remind myself that there are lots of other people in the world feeling like I am	− 0.37	2.49	11.71
10	When I feel inadequate in some way, I try to remind myself that feelings of inadequacy are shared by most people	− 0.35	0.54	7.67
15	I try to see my feelings as part of the human condition	− 0.38	1.09	4.76
Isolation				
4	When I think about my inadequacies, it tends to make me feel more separate and cut off from the rest of the world[R]	0.18	2.01	4.38
13	When I'm feeling down, I tend to feel like most other people are probably happier than I am[R]	0.22	− 0.05	2.58
18	When I'm really struggling, I tend to feel like other people must be having an easier time of it[R]	0.08	2.93	5.89
25	When I fail at something that's important to me, I tend to feel alone in my failure[R]	0.42	− 0.65	3.56
Mindfulness				
9	When something upsets me I try to keep my emotions in balance	− 0.86	1.44	9.58
14	When something painful happens I try to take a balanced view of the situation	− 0.65	− 0.52	6.23
17	When I fail at something important to me I try to keep things in perspective	− 0.49	− 0.57	6.72
22	When I'm feeling down I try to approach my feelings with curiosity and openness	− 0.05	0.78	1.78
Over-identified				
2	When I'm feeling down I tend to obsess and fixate on everything that's wrong[R]	0.64	− 1.69	12.21
6	When I fail at something important to me I become consumed by feelings of inadequacy[R]	0.43	0.23	4.34
20	When something upsets me I get carried away with my feelings[R]	0.45	− 0.21	4.25
24	When something painful happens I tend to blow the incident out of proportion[R]	− 0.08	2.87	3.08

Item numbers are based on the original 26-item SCS version (Neff 2003a)

[R] Reverse-scored item

isolation, mindfulness, over-identification). This resulted in satisfactory goodness of fit with non-significant item-trait interaction (χ^2 (30) = 23.27, p = 0.80), meaning that scale and individual items were functioning equally well at all levels of latent trait. There was a noticeable improvement of reliability (PSI = 0.88). However, the assumption of unidimensionality was still violated at this stage, requiring further investigation.

The residual correlation matrix including the six super-items was evaluated and indicated that there was still room for improvement as some super-items continued to show local dependency exceeding the acceptable threshold. To achieve the best fit, items with higher residual correlations were further combined to create four super-items including self-kindness combined with isolation, common humanity with over-identification, self-judgement and mindfulness, following the methodology of Lundgren-Nilsson et al. (2013) and Balalla et al. (2019). This final analysis showed strong evidence of unidimensionality with lower bound of significant t tests (2.9%) overlapping 5% cut-off point, which was associated with a further increase of reliability (PSI = 0.90). Similarly, goodness of fit indicated a further reduction of error due to the interaction between items and the latent trait (χ^2 (20) = 21.02, p = 0.40). DIF was examined for age and sex, and no significant differences were identified on any of the created super-items from the final analysis. Figure 1 presents the person-item threshold distribution of the best solution without re-scoring where items cover individual ability for the long form. It shows that 100% of the samples are perfectly targeted by item thresholds of the SCS with person location mean of − 0.10 (SD = 0.58) and no signs of either ceiling or floor effects.

Rasch Analysis of the SCS (Sample B)

For the purpose of robustness, this Rasch analysis of the SCS (Neff 2003a) was replicated on the second sample (sample B). The results of this replication are included in Table 1. Initial analysis of sample B (n = 371) showed good level of reliability with PSI = 0.94, but there was misfit to the Rasch model (χ^2 (130) = 179.03, p < 0.001). Examining residual correlation matrix showed local dependencies between the same items identified in the Rasch analysis with sample A, which resulted in formation of four super-items (self-kindness with isolation, common humanity with over-identification, self-judgement, mindfulness). Similarly, these items were combined because of local dependency. Final analysis confirmed unidimensionality of the SCS. Upon examining DIF for age and sex, there were no significant differences across items between tested groups. Both goodness of fit (χ^2 (20) = 22.52, p = 0.31) and reliability with PSI = 0.90 were satisfactory and consistent with the Rasch analysis of sample A (Table 1).

Rasch Analysis of the SCS-SF (Sample A)

The 12-item SCS-SF (Raes et al. 2011) was also examined using sample A. After deleting items from the SCS that are not included in the SCS-SF (items 3, 4, 5, 7, 8 and 16 to 24), the initial analysis was conducted. Significant error due to the interaction between items and self-compassion trait affected the Rasch model fit (χ^2 (60) = 100.38, p < 0.001), but a good level of reliability with PSI = 0.87 was evident. Similar to the SCS, in assessing the residual correlation matrix, the local dependency between the same groups of items was evident. Super-items were formed based on items sharing dependency (self-kindness, self-judgement, common humanity, isolation, mindfulness, over-identification). This modification produced an acceptable fit (χ^2 (30) = 36.61, p = 0.18) with good reliability of PSI = 0.84, but there were signs of local dependency between super-items. To resolve this issue, the final analysis combined two pairs of these six super-items to make four super-items (self-kindness with isolation, common humanity with over-identification, self-judgement, mindfulness), which further enhanced fit (χ^2 (20) = 23.29, p = 0.27) and slightly improved reliability (PSI = 0.85) while strict unidimensionality was confirmed by only 4.3% of significant t tests. Figure 2 presents the person-item threshold distribution of the best Rasch model solution for the SCS-SF. It shows a good sample

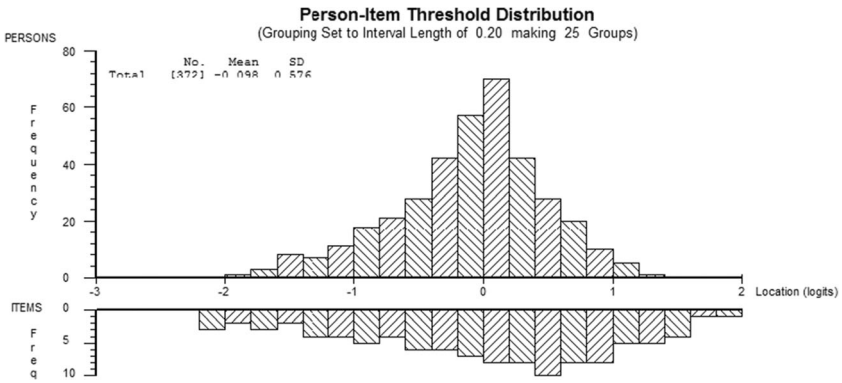

Fig. 1 Person-item threshold distribution for the final analysis of the SCS (sample A)

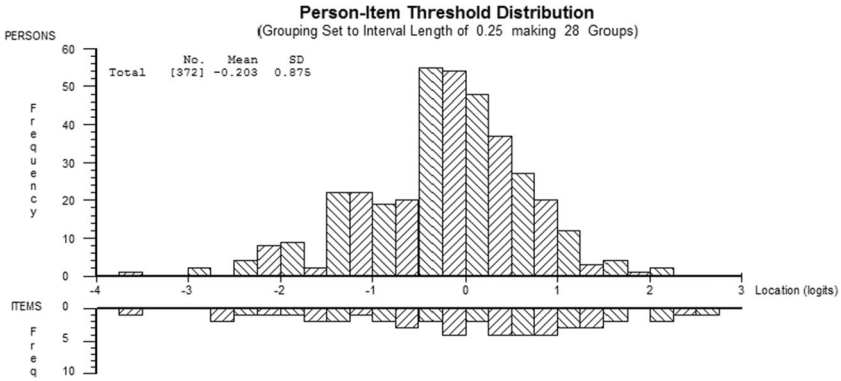

Fig. 2 Person-item threshold distribution for the final analysis of the SCS-SF (sample A)

targeting with the sample mean of − 0.20 (SD = 0.88) and coverage of 100% of the sample by the SCS-SF item thresholds.

Rasch Analysis of the SCS-SF (Sample B)

Replicating Rasch analysis of the SCS-SF with sample B (n = 371) involved deleting items (3, 4, 5, 7, 8 and 16 to 24) from the full scale. Initial analysis showed a misfit to the Rasch model (χ^2 (60) = 102.85, $p < 0.001$), but a good level of reliability with PSI = 0.87. Similar to the analysis with sample A, local dependency was identified and resolved using a super-item approach. The final analysis uses a four super-item solution and shows goodness of fit (χ^2 (20) = 18.13, p = 0.58) with a reliability PSI of 0.84 and strict unidimensionality with merely 3.8% of significant t tests.

Ordinal-to-Interval Conversion Tables

The SCS satisfied expectations of the unidimensional Rasch model, meaning that the ordinal-to-interval conversion algorithm can be generated using Rasch model person estimates. Table 3 includes interval-level scores in logit units and original scale metric that can be used to transform ordinal raw scores of the full SCS ranging from 26 to 130 located on the left-hand side into linear measure. Table 4 includes ordinal-to-interval conversion scores for the SCS-SF version. These tables allow valid interval-level scores accounting for a differential contribution of different facets to the overall self-compassion construct in both full and short scale forms. These conversion tables are easy to apply following the instructions below. All negatively worded items (1, 2, 4, 6, 8, 11, 13, 16, 18, 20, 21, 24 and 25) for the full SCS and those (1, 4, 8, 9, 11 and 12) for the SCS-SF should be reverse coded before computing the total scores. Total score is calculated by adding individual item scores together, and corresponding interval-level scores for the SCS and SCS-SF can be found on the right-hand side in Tables 3 and 4, respectively. For instance, an ordinal score of 40 corresponds to the interval score of 57.31 using the same scale range and an ordinal score of 90 will correspond to a linear score of 81.28. To compute mean interval score similar to that recommended by authors for both the SCS and SCS-SF ordinal scores (Neff 2003a), the interval scale score should be divided by the number of items (26 and 12, respectively), resulting in the interval-level scores ranging from 1 to 5.

Discussion

This study evaluated the psychometric properties and dimensionality of the SCS and SCS-SF, as well as the compliance of these measures with the fundamental principles of measurement using the Rasch analysis with two adequate independent samples for the purpose of robustness. We used advanced methodology that involves creating super-items, summarising scores of individual items, which permitted reduction of measurement error and control for spurious correlations and method effects (e.g. negatively worded items; Medvedev et al. 2017a, b, 2018). The results show that the best Rasch model fit was achieved for both SCS and SCS-SF after combining locally dependent items into four super-items, with evidence of excellent sample targeting and unidimensionality. Both scale versions demonstrated sound reliability in differentiating between individuals based on their self-compassion levels and met conservative requirements for individual (PSI ≥ 0.70) and group (PSI ≥ 0.80) assessments (Tennant and Conaghan 2007). These results were replicated with another independent sample for both scale versions and demonstrated similarly good Rasch model fit and sound reliability. Therefore, ordinal-to-interval conversion tables were produced based on Rasch model person estimates. Together these findings support the reliability and internal validity of both the SCS and SCS-SF and permit to enhance their accuracy by using the ordinal-to-interval conversion tables published here (Tables 3 and 4). Although both the original SCS and its short version (SCS-SF) met expectations of the unidimensional Rasch model, the full SCS more adequately represents self-compassion construct reflected by its higher precision in discriminating between individuals (PSI = 0.90) compared to the

Table 3 Ordinal-to-interval conversion for the 26-item SCS

Ordinal scores	Interval	
	Logits	Scale
26	−3.48	26.00
27	−2.87	34.71
28	−2.50	40.11
29	−2.27	43.46
30	−2.10	45.87
31	−1.97	47.76
32	−1.86	49.33
33	−1.76	50.68
34	−1.68	51.88
35	−1.61	52.96
36	−1.54	53.94
37	−1.47	54.87
38	−1.42	55.72
39	−1.36	56.53
40	−1.31	57.31
41	−1.25	58.04
42	−1.21	58.75
43	−1.16	59.43
44	−1.11	60.09
45	−1.07	60.74
46	−1.02	61.37
47	−0.98	61.98
48	−0.94	62.57
49	−0.90	63.15
50	−0.86	63.73
51	−0.82	64.27
52	−0.78	64.82
53	−0.75	65.36
54	−0.71	65.87
55	−0.68	66.38
56	−0.64	66.88
57	−0.61	67.37
58	−0.57	67.85
59	−0.54	68.33
60	−0.51	68.79
61	−0.48	69.25
62	−0.45	69.70
63	−0.42	70.14
64	−0.39	70.58
65	−0.35	71.02
66	−0.33	71.44
67	−0.30	71.88
68	−0.27	72.29
69	−0.24	72.71
70	−0.21	73.13
71	−0.18	73.55
72	−0.15	73.95
73	−0.12	74.36
74	−0.09	74.78
75	−0.07	75.18
76	−0.04	75.58
77	−0.01	75.99
78	0.02	76.39
79	0.05	76.80
80	0.07	77.20
81	0.10	77.60
82	0.13	78.01
83	0.16	78.41
84	0.19	78.82
85	0.21	79.22
86	0.24	79.62
87	0.27	80.04
88	0.30	80.45
89	0.33	80.86
90	0.36	81.28
91	0.39	81.70
92	0.42	82.12
93	0.45	82.55
94	0.48	82.98
95	0.51	83.42
96	0.54	83.86
97	0.57	84.30
98	0.60	84.76
99	0.63	85.21
100	0.66	85.67
101	0.69	86.14
102	0.73	86.62
103	0.76	87.11
104	0.80	87.60
105	0.83	88.11
106	0.87	88.63
107	0.90	89.14
108	0.94	89.68
109	0.98	90.23
110	1.02	90.79
111	1.06	91.38
112	1.10	91.99
113	1.14	92.59
114	1.19	93.24
115	1.23	93.92
116	1.28	94.64
117	1.34	95.39
118	1.39	96.20
119	1.45	97.08
120	1.52	98.02
121	1.59	99.07
122	1.67	100.24
123	1.76	101.58
124	1.87	103.14
125	2.00	104.98
126	2.16	107.28
127	2.36	110.19
128	2.64	114.19
129	3.07	120.38
130	3.73	130.00

SCS-SF (PSI = 0.85). Therefore, using the original SCS for measuring self-compassion is preferable to achieve higher precision and validity of assessment.

The current study initially tested the subscale items which reflected the 6-factor structure set out by Neff (2003a): self-kindness, self-judgement, common humanity, isolation, mindfulness and over-identification. Local dependency was present in some subscales (e.g. self-kindness ["I try to be loving towards myself when I'm feeling emotional pain"] with isolation ["When I think about my inadequacies, it tends to make me feel more separate and cut off from the rest of the world"], and common humanity ["When things are going badly for me, I see the difficulties as part of life that everyone goes through"] with over-identification ["When I'm feeling down I tend to

Table 4 Ordinal-to-interval conversion for the 12-item SCS-SF

Ordinal scores	Interval	
	Logits	Scale
12	−3.87	12.00
13	−3.19	15.70
14	−2.75	18.15
15	−2.45	19.77
16	−2.22	21.02
17	−2.04	22.05
18	−1.87	22.94
19	−1.73	23.74
20	−1.59	24.49
21	−1.46	25.18
22	−1.34	25.84
23	−1.23	26.47
24	−1.12	27.07
25	−1.01	27.65
26	−0.91	28.22
27	−0.81	28.77
28	−0.71	29.30
29	−0.62	29.81
30	−0.53	30.31
31	−0.44	30.80
32	−0.35	31.28
33	−0.26	31.75
34	−0.18	32.21
35	−0.10	32.67
36	−0.01	33.13
37	0.07	33.58
38	0.15	34.04
39	0.24	34.49
40	0.32	34.95
41	0.41	35.42
42	0.49	35.90
43	0.58	36.39
44	0.68	36.90
45	0.77	37.43
46	0.87	37.97
47	0.98	38.54
48	1.08	39.13
49	1.20	39.75
50	1.32	40.41
51	1.45	41.11
52	1.58	41.86
53	1.73	42.67
54	1.89	43.57
55	2.08	44.60
56	2.31	45.83
57	2.59	47.40
58	3.00	49.63
59	3.70	53.45
60	4.89	60.00

obsess and fixate on everything that's wrong"]). These were dealt with following Lundgren-Nilsson and Tennant (2011) by combining the highly correlated subscale items to create super-items. In the same way, Balalla et al. (2019) first combined the items of the World Health Organization Quality of Life measure into four super-items reflecting four subscales in their Rasch analysis. Local dependency was not present between other subscales such as the self-judgement and mindfulness, so no modifications were necessary. We also found some items between scales and within scales correlated although not meeting the margin of greater than 0.20 compared to the mean of all residual correlations, for example, self-judgement (item 1 ["I'm disapproving and judgmental about my own flaws and inadequacies"]) and over-identification (item 2 ["When I'm feeling down, I tend to obsess and fixate on everything that's wrong"]). Items measuring the same subscale also correlated highly together, which supported super-item approach used for individual subscales, for example common humanity (item 7 ["When I'm down, I remind myself that there are lots of other people in the world feeling like I am"] and item 10 ["When I feel inadequate in some way, I try to remind myself that feelings of inadequacy are share by most people"]).

It seems that Neff's (2003a) Isolation subscale encompasses alliance with others by considering that the items are negatively worded, and needs to be reverse coded to compute the total score. Therefore, combining alliance with others (isolation) and self-kindness as facets of self-compassion–based local dependency indicated that these facets share common variance after accounting for self-compassion variance present in both. Common humanity and over-identification (decentering; Lau et al. 2006) subscales seem to measure the same aspect of self-compassion; again, the reverse coding supports the argument that reduced over-identification appears as a characteristic of common humanity.

The reversal of each negatively scored subscale should be defined as an opposite of that negative construct (e.g. lack of isolation = alliance with others), meaning that isolation and over-identification in Neff's (2003a) scale compliment the measurement of self-kindness and common humanity and measure the same aspects of the construct. We have come to a similar conclusion for mindfulness and self-judgement as having a common base because all mainstream mindfulness definitions incorporate non-judgemental attitude (Segal et al. 2013).

Our study provided robust evidence of unidimensionality and invariance of the SCS and SCS-SF, indicating the overarching latent construct of self-compassion (bi-factor model) including four individual facets. The creation of super-items in this study successfully addressed local dependency issue that may explain spurious correlations affecting CFA fit in earlier studies (Apodaca and Grad 2005; Cox et al. 1996). An example where super-items did not work due to multidimensionality can be seen in Mitchell-Parker et al. (2018), where the super-item representing a subscale was removed due to poor fit to the Rasch model. Achieving good Rasch model fit, sound reliability and unidimensionality in the current study supports the argument that both the SCS and SCS-SF represent adequate measures of an overall self-compassion. This allows for an ordinal-to-interval conversion table to be generated based on person estimates of the Rasch model. The transformation of scores is important because precision of scores is improved to accurately

reflect individual responses, and the interval-level data can be used with parametric statistics without violating their assumptions. Important to note, initial individual item fit statistics presented in Table 2 showed no items with a significant misfit before any super-items were created. This supports the construct validity of the SCS items and appropriate psychometric properties of all positively and negatively worded items. We did not find evidence to suggest that positively worded items differ in psychometric properties to negatively worded items as proposed by Gilbert et al. (2011).

Limitations and Future Research

The current study was conducted with a predominantly non-clinical sample, and future research should focus on replicating these findings in a clinical sample such as groups suffering from affective disorders or other psychological health conditions. The sample used here consisted predominantly of individuals identifying as White British and were largely female. Future study could focus on generalising this study to other ethnicities. However, the Rasch analysis is considered less sample dependent compared to other psychometric methods (Tennant and Conaghan 2007) and the current sample was large enough and permitted replication of the results for the purpose of robustness contributing to the generalisability of these findings.

Acknowledgements Data collection on this project was part of a larger data collection, internally funded by Nottingham Trent University, Health and Wellbeing Grant. We would like to thank Bryony Heasman for her help in collecting data and all participants for their time in taking part in the study. We would like to acknowledge and thank the University of Waikato Māori & Psychology Research Unit for granting a scholarship to the first author to complete this study.

Authors' Contributions Initial draft of the manuscript was written by KPF and ONM, with edits from AS and NH. Data collection was conducted by AS and NH. All authors were involved in the study conceptualisation. Data cleaning, scoring and analyses were performed by KPF and ONM. All authors edited further revisions of the manuscript and approved the final version of the manuscript for submission.

Funding Part funding for the project was provided internally through Nottingham Trent University.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

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

Ethics Statement All study procedures were approved by the ethics committee for the School of Social Sciences, Nottingham Trent University, UK.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573.

Andrich, D., Sheridan, B., & Luo, G. (2009). *RUMM 2030 (beta version for Windows)*. Perth, Western Australia: RUMM Laboratory Pty Ltd.

Apodaca, P., & Grad, H. (2005). The dimensionality of student ratings of teaching: Integration of uni- and multidimensional models. *Studies in Higher Education, 30*(6), 723–748.

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

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.

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.

Cox, B. J., Parker, J. D. A., & Swinson, R. P. (1996). Anxiety sensitivity: Confirmatory evidence for a multidimensional construct. *Behaviour Research and Therapy, 34*(7), 591–598.

Durkin, M., Beaumont, E., Martin, C. J. H., & Carson, J. (2016). A pilot study exploring the relationship between self-compassion, self-judgement, self-kindness, compassion, professional quality of life and wellbeing among UK community nurses. *Nurse Education Today, 46*, 109–114.

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

Gilbert, P., McEwan, K., Matos, M., & Rivis, A. (2011). Fears of compassion: Development of three self-report measures. *Psychology and Psychotherapy: Theory, Research and Practice, 84*(3), 239–255.

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.

Heym, N., Heasman, B. C., Hunter, K., Blanco, S. R., Wang, G. Y., Siegert, R., Cleare, A., Gibson, G. R., Kumari, V., & Sumich, A. L. (2019). The role of microbiota and inflammation in self-judgement and empathy: Implications for understanding the brain-gut-microbiome axis in depression. *Psychopharmacology, 236*(5), 1459–1470.

Hobart, J., & Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods. *Health Technology Assessment, 13*(12), iii, ix–x, 1–177. https://doi.org/10.3310/hta13120.

Kumlander, S., Lahtinen, O., Turunen, T., & Salmivalli, C. (2018). Two is more valid than one, but is six even better? The factor structure of the Self-Compassion Scale (SCS). *PLoS One, 13*(12), e0207706.

Lau, M. A., Bishop, S. R., Segal, Z. V., Buis, T., Anderson, N. D., Carlson, L., Shapiro, S., Carmody, J., Abbey, S., & Devins, G. (2006). The Toronto Mindfulness Scale: Development and validation. *Journal of Clinical Psychology, 62*(12), 1445–1467.

Leung, Y.-Y., Png, M.-E., Conaghan, P., & Tennant, A. (2013). A systematic literature review on the application of Rasch analysis in musculoskeletal disease—A special interest group report of OMERACT 11. *The Journal of Rheumatology, 41*(1), 159–164.

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.

Lundgren-Nilsson, Å., & Tennant, A. (2011). Past and present issues in Rasch analysis: The Functional Independence Measure (FIM) revisited. *Journal of Rehabilitation Medicine, 43*(10), 884–892.

Lundgren-Nilsson, Å., Jonsdottir, I. H., Ahlborg, G., & Tennant, A. (2013). Construct validity of the psychological general well being index (PGWBI) in a sample of patients undergoing treatment for stress-related exhaustion: a Rasch analysis. *Health and Quality of Life Outcomes, 11*(1), 2.

Marais, I., & Andrich, D. (2008). Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model. *Journal of Applied Measurement, 9*(2), 105–124.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.

Medvedev, O. N., Siegert, R. J., Kersten, P., & Krägeloh, C. U. (2017a). Improving the precision of the Five Facet Mindfulness Questionnaire using a Rasch approach. *Mindfulness, 8*(4), 995–1008.

Medvedev, O. N., Siegert, R. J., Mohamed, A. D., Shepherd, D., Landhuis, E., & Krägeloh, C. U. (2017b). The Oxford Happiness Questionnaire: Transformation from an ordinal to an interval measure using Rasch analysis. *Journal of Happiness Studies, 18*(5), 1425–1443.

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.

Mitchell-Parker, K., Medvedev, O. N., Krägeloh, C. U., & Siegert, R. J. (2018). Rasch analysis of the Frost Multidimensional Perfectionism Scale. *Australian Journal of Psychology, 70*(3), 258–268.

Neff, K. D. (2003a). Development and validation of a scale to measure self-compassion. *Self and Identity, 2*(3), 223–250.

Neff, K. (2003b). Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity, 2*(2), 85–101.

Neff, K. D. (2016). The Self-Compassion Scale is a valid and theoretically coherent measure of self-compassion. *Mindfulness, 7*(1), 264–274.

Neff, K. D., & Dahm, K. A. (2015). Self-compassion: What it is, what it does, and how it relates to mindfulness. In B. Ostafin, M. Robinson, & B. Meier (Eds.), *Handbook of mindfulness and self-regulation* (pp. 121–137). Berlin: Springer. https://doi.org/10.1007/978-1-4939-2263-5_10.

Neff, K. D., Tóth-Király, I., Yarnell, L. M., Arimitsu, K., Castilho, P., Ghorbani, N., Guo, H. X., Hirsch, J. K., Hupfeld, J., Hutz, C. S., & Kotsou, I. (2019). Examining the factor structure of the Self-Compassion Scale in 20 diverse samples: Support for use of a total score and six subscale scores. *Psychological Assessment, 31*(1), 27.

Nijsten, T., Unaeze, J., & Stern, R. S. (2006). Refinement and reduction of the Impact of Psoriasis Questionnaire: Classical test theory vs. Rasch analysis. *British Journal of Dermatology, 154*(4), 692–700.

Norquist, J. M., Fitzpatrick, R., Dawson, J., & Jenkinson, C. (2004). Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Medical Care*, *42*(1 Suppl), I25–I36.

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.

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.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Denmark, Copenhagen: Danish Institute for Educational Research.

Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2013). *Mindfulness-based cognitive therapy for depression* (2nd ed.). New York: The Guilford Press.

Shonin, E., Van Gordon, W., Compare, A., Zangeneh, M., & Griffiths, M. D. (2014). Buddhist-derived loving-kindness and compassion meditation for the treatment of psychopathology: A systematic review. *Mindfulness, 6*(5), 1161–1180.

Shrestha, A. K. (2016). Positive psychology: Evolution, philosophical foundations, and present growth. *Indian Journal of Positive Psychology, 7*(4), 460.

Sinclair, S., Kondejewski, J., Raffin-Bouchal, S., King-Shier, K. M., & Singh, P. (2017). Can self-compassion promote healthcare provider well-being and compassionate care to others? Results of a systematic review. *Applied Psychology: Health and Well-Being, 9*(2), 168–206.

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.

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.

Tennant, A., & Pallant, J. F. (2006). Unidimensionality matters! (A tale of two Smiths?). *Rasch Measurement Transactions, 20*(1), 1048–1051.

Thurstone, L. L. (1931). The measurement of social attitudes. *The Journal of Abnormal and Social Psychology, 26*(3), 249.

Wilson, M. (2004). *Constructing measures: An item response modeling approach* (1st ed.). London: Routledge.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. San Diego: Mesa Press.