Examining the factor structure of the Self-Compassion Scale in four distinct populations: Is the use of a total scale score justified?

Kristin D. Neff and Tiffany A. Whitaker

University of Texas at Austin

Anke Karl

University of Exeter

In press, Journal of Personality Assessment

(UNCORRECTED PROOF)

The clinical sample examined in this study was drawn from the PREVENT trial, a project funded by the National Institute for Health Research Health Technology Assessment (NIHR HTA) Programme (project number 08/56/01). This trial is reported in full in the Lancet, DOI: http://dx.doi.org/10.1016/S0140-6736(14)62222-4. We are grateful to Willem Kuyken for his comments on a draft of this manuscript and the trial team for allowing us to use the data.

Correspondence concerning this article should be addressed to Kristin D. Neff, Department of Educational Psychology, University of Texas at Austin, 1912 Speedway, Stop D5800, Austin, TX 78712-1289. E-mail: kristin.neff@mail.utexas.edu

Abstract

This study examined the factor structure of the Self-Compassion Scale (SCS) using a bifactor model, a higher-order model, a six-factor correlated model, a two-factor correlated model and a one-factor model in four distinct populations: college undergraduates (N = 222), community adults (N = 1394), individuals practicing Buddhist meditation (N = 215), and a clinical sample of individuals with a history of recurrent depression (N = 390). The six-factor correlated model demonstrated the best fit across samples, while the one and two-factor models had poor fit. The higher-order model also showed relatively poor fit across samples, suggesting it is not representative of the relationship between subscale factors and a general self-compassion factor. The bifactor model, however, had acceptable fit in the student, community and meditator samples. While fit was suboptimal in the clinical sample, results suggested an overall self-compassion factor could still be interpreted with some confidence. Moreover, estimates suggested a general self-compassion factor accounted for at least 90% of the reliable variance in SCS scores across samples, and item factor loadings and intercepts were equivalent across samples. Results suggest that a total SCS score can be used as an overall measure of self-compassion.

Examining the factor structure of the Self-Compassion Scale in four distinct populations: Is the use of a total scale score justified?

The construct of self-compassion is drawn from Buddhist psychology, and was first operationally defined and introduced into the psychological literature over a decade ago. Neff (2003b, 2016) proposes that self-compassion is a type of self-to-self relating that represents a compassionate rather than uncompassionate stance toward the self when faced with personal suffering: self-kindness versus self-judgment, a sense of common humanity versus isolation, and mindfulness versus over-identification. These components combine and mutually interact to create a self-compassionate frame of mind. Self-kindness entails being gentle, supportive and understanding towards oneself. Rather than harshly judging oneself for personal shortcomings, the self is offered warmth and unconditional acceptance. Common humanity involves recognizing the shared human experience, understanding that all humans fail and make mistakes, and that all people lead imperfect lives. Rather than feeling isolated by one's imperfection - egocentrically feeling as if "I" am the only one who has failed or am suffering - one takes a broader and more connected perspective with regard to personal shortcomings and individual difficulties. Mindfulness involves being aware of one's present moment experience of suffering with clarity and balance, without running away with a dramatic storyline about negative aspects of oneself or one's life experience - a process that is termed "over-identification." Self-compassion can be directed towards the self when suffering occurs through no fault of one's own - when the external circumstances of life are simply painful or difficult to bear. Self-compassion is equally relevant, however, when suffering stems from one's own imprudent actions or

personal failures.

As Neff (2016) writes, the various elements of self-compassion are conceptually distinct and tap into different ways that individuals emotionally respond to pain and failure (with kindness or judgment), cognitively understand their predicament (as part of the human experience or as isolating), and pay attention to suffering (in a mindful or over-identified manner). While these components are separable and do not co-vary in a lockstep manner, they do mutually impact one another. Put another way, self-compassion can be seen as a dynamic system that represents a synergistic state of interaction between the various elements of self-compassion.

Over the past decade research on self-compassion has expanded exponentially. Self-compassion has been consistently related to psychological health in the research literature, including increased positive outcomes such as happiness and life satisfaction and decreased negative outcomes such as anxiety and depression (Barnard & Curry, 2011; MacBeth & Gumley, 2012; Zessin, Dickhauser & Garbadee, 2015). Most of the research on self-compassion has been conducted using the Self-Compassion Scale (SCS; Neff, 2003a), which assesses overall trait levels of self-compassion. Items are written in a face-valid manner and measure the cognitive and emotional behaviors associated with more compassionate and fewer uncompassionate responses to feelings of personal inadequacy and general life difficulties. Sample items (Neff, 2003a) are: Self-Kindness ("I try to be loving towards myself when I'm feeling emotional pain"), Self-Judgment ("I'm disapproving and judgmental about my own flaws and inadequacies"), Common Humanity ("When things are going badly for me, I see the difficulties as part of life that everyone goes through"), Isolation ("When I think about my inadequacies it tends to make me feel more separate and cut off from the rest of the world"), Mindfulness ("When I'm feeling

down I try to approach my feelings with curiosity and openness"), and Over-Identification ("When something upsets me I get carried away with my feelings"). Neff (2016) argues that the mind-state of self-compassion represents the relative balance of compassionate and uncompassionate responses to suffering, which is why the SCS measures the lack of uncompassionate responses as well as the presence of compassionate ones.

The SCS was developed in a sample of college undergraduates (Neff, 2003a). After identifying items that made sense to students, Neff (2003a) used exploratory factor analyses (EFA) to identify items that loaded best on separate subscales representing the six components of self-compassion. Confirmatory factor analyses (CFA) were used to provide support that scale items fit as intended with the proposed a priori theoretical model (Furr & Bacharach, 2008). An initial CFA found an adequate fit to a six-factor inter-correlated model (NNFI .90; CFI .91). A second CFA found a marginal fit to a higher-order factor model (NNFI .88; CFI .90) that accounted for the inter-correlations between subscales. The factor structure of the scale was cross-validated in a second student sample, and adequate fit was found for the six-factor inter-correlated model (NNFI .92; CFI .93) as well as the higher-order factor model (NNFI .90; CFI .92). Findings were interpreted as evidence that the subscales could be examined separately or else that a total score could be used to represent overall self-compassion levels, according the interests of the researcher.

In the original publication (Neff, 2003a), total SCS scores evidenced good internal reliability (Cronbach's α =.92), as did the six subscales (with Cronbach's α ranging from .75 to .81). Test-retest reliability over a three-week interval was also good for the total score (Cronbach's α =.93) and six subscale scores (with Cronbach's α ranging from .80 to .88). Moreover, the internal reliability of SCS scores has been found to be high across a wide variety

of populations (e.g., Neff & Pommier, 2013; Allen, Goldwasser & Leary, 2012; Werner et al., 2012).

There is ample evidence for the idea that interpretation of scores on the SCS have construct validity. Firstly, the large body of research indicating that higher total scores on the SCS are associated with wellbeing is a source of predictive validity. For instance, higher total scores on the SCS have been associated with greater levels of happiness, optimism, life satisfaction, body appreciation, perceived competence, and motivation (Hollis-Walker & Colosimo, 2011; Neff, Hsieh & Dejitthirat, 2005; Neff, Pisitsungkagarn & Hsieh, 2008; Neff, Rude, & Kirkpatrick, 2007); as well as lower levels of depression, anxiety, stress, rumination, body shame and fear of failure (Daye, Webb & Jafari, 2014; Finlay-Jones, Rees, & Kane, 2015; Neff, Hseih, & Dejitthirat, 2005; Raes, 2010). Higher scores on the SCS are also predictive of healthier physiological responses to stress (Breines et al., 2015; Breines, Thoma et al., 2014). Moreover, these findings have been duplicated in research using non-self-report methods, providing convergent validity for the SCS and thus support for the construct validity of self-compassion (Neff, 2016). For instance, self-compassion interventions have been found to increase optimism, happiness, life satisfaction, self-efficacy, and body appreciation, to decrease rumination, depression, anxiety, stress and body shame (Albertson et al., 2014; Neff & Germer, 2013; Shapira & Mongrain, 2010; Smeets et al., 2014), and to positively impact physiological responses to stress (Arch et al., 2014.) Similarly, experimental studies designed to induce a self-compassionate mood (i.e., responding to writing prompts that foster self-kindness, common humanity and mindfulness) have been shown to increase positive affect and motivation and also decrease negative emotions such as anxiety, shame, and depression (Breines & Chen, 2012; Diedrich et al, 2014; Johnson & O'Brien, 2013; Leary et al, 2007; Odou & Brinker, 2014).

SCS total scores also demonstrate good discriminant validity. Although a key feature of self-compassion is a lack of self-judgment, and SCS scores are moderately correlated with self-criticism (Blatt, D'Afflitti, and Quinlan, 1976), total SCS scores still negatively predict anxiety and depression when controlling for self-criticism and negative affect (Neff, 2003a; Neff, Kirkpatrick & Rude, 2007). In addition, SCS scores demonstrate known groups validity: individuals who practice Buddhist meditation have higher total SCS scores, as would be expected given the Buddhist origins of the construct (Neff, 2003a; Neff & Pommier, 2013). Scale scores demonstrate excellent convergent validity in terms of consistency with ratings by observers. For instance, therapists were able to significantly predict individuals' SCS scores after a brief interaction (Neff, Kirpatrick & Rude, 2007), and there was a strong association between self-reported and partner-reported scores on the SCS (multi-level modeling estimated the association to be .70) among couples in long-term romantic relationships (Neff & Beretvas, 2013). Similarly, high levels of agreement (ICC = .77) were found between independent coders using SCS items to rate the level of self-compassion displayed in brief verbal dialogues (Sbarra, Smith & Mehl, 2012). These findings suggest that the SCS measures behaviors that are clearly observable by others.

Because the SCS is the main measure available to assess self-compassion, it is important to determine whether its factor structure is valid and replicable across various populations. Given that the SCS is designed to assess self-compassion as an overall construct, moreover, it is especially important to determine whether use of a total scale score is psychometrically justified. As Neff (2016) points out, most studies examining the factor structure of the SCS have been conducted in the course of validating translations of the scale. The large majority of translations have replicated the six-factor structure of the SCS (e.g., Castilho, Pinto-Gouveia, & Duarte,

2015; Chen, Yan & Zhou, 2011; Garcia-Campayo et al., 2014; Hupfield & Ruffieux, 2011; Petrocchi, Ottaviani & Couyoumdjian, 2013). While not all examined the higher-order model, those which did yielded inconsistent findings. For example, a higher-order factor was found with a Chinese student and Portuguese clinical and community samples (Castilho, Pinto-Gouveia, & Duarte, 2015; Chen et al., 2011), but not with German and Italian student and community samples (Hupfield & Ruffieux, 2011; Petrocchi et al, 2013), a Dutch community sample (López et al., 2015) or a second Portuguese clinical sample (Costa, Marôco, Pinto-Gouveia, Ferreira & Castilho, 2015).

In the course of examining these translations, López et al., (2015) conducted EFA and found that all the positive items loaded on one factor and all the negative items loaded on a second factor, and argued that the SCS was bi-dimensional and that a total scale score should not be used (no CFA was conducted to support this factor structure). Of course, findings could have been due to a method effect, given that EFA is highly sensitive to positive versus negative affect (DiStefano & Motl, 2006). Costa et al. (2015) used CFA to compare a six-factor uncorrelated model, a higher order model, a two factor uncorrelated model that separated positive and negative items, and a two factor model that included correlated errors designed to improve model fit, and found that the two-factor model with correlated errors had the best fit. Note that the researchers' choice to examine an uncorrelated six-factor model was inconsistent with Neff's (2003a) original approach, in which the six factors are theoretically presumed to correlate. As Neff (2016) cautions, moreover, care must be used before assuming that findings obtained with translations generalize to the original language version of a scale given potential issues with the quality of translations or cultural factors that may impact findings (Behling & Law, 2000).

Surprisingly few studies have attempted to replicate the factor structure of the original

English SCS. An important exception is a study by Williams, Dalgleish, Karl, and Kuyken (2014) that examined the factor structure of the SCS in a community (N = 821), meditator (N = 211) and clinical sample of individuals with recurrent depression (N = 390) living in the United Kingdom. Williams et al. (2014) report that CFAs were used in each sample to examine SCS item fit to a one-factor model, a six-factor correlated model, and a higher-order model. The authors concluded that the one-factor and higher-order models did not fit the data acceptably. The six-factor correlated model fit the data more favorably than the remaining models in all populations examined, and demonstrated an adequate fit for the community sample. The authors concluded that "the SCS may be better suited to measuring six hypothesized facets of self-compassion... than for measuring an overarching construct (i.e., self-compassion)" and that "further research is needed to develop a more psychometrically robust measure of self-compassion" (p. 10).

Because of the good reliability and strong support for the predictive, convergent and discriminant validity of the SCS, and because the measure has been used in such a large number of empirical studies - the vast majority of which have used a total scale score (Neff, 2016) - more investigation is warranted before concluding that the SCS should be redesigned. Some of the limitations of relying on model fit as a means of evaluating the adequacy of a factor structure representing the items on a scale are that it depends on sample size and other model properties. For instance, simulation studies in this area have found that model fit indices suggest poor model fit of the correct model with an increase in the number of indicators per factor (Marsh, Hau, Balla, & Grayson, 1998) and that model fit decreases as the size of factor loadings increase (Saris, Satorra, & van der Veld, 2009), both of which are counterintuitive findings. This is echoed in applied research settings in which traditional cutoff values do not tend to support

complex, multi-factor (e.g., five or more factors with five or more indicators per factor) structures associated with good scales (see e.g., Marsh, Hau, & Grayson, 2005). This suggests that the cutoff values associated with model fit indices in structural equation modeling (SEM) are dependent upon the model under investigation and its properties. As such, the traditional model fit cutoff values are recommended as broad guidelines and not stringent benchmarks (Marsh, Hau, & Wen, 2004; West, Taylor, & Wu, 2012). Other markers of adequate model fit, particularly the effect sizes associated with the relevant parameter estimates, should also be examined. Given the complex factors involved in determining the quality of a psychometric measure - including the soundness of its theoretical underpinnings and its validity and utility for facilitating research - it is useful to have a more substantive estimate of the ability of a self-report measure to assess its intended target than model fit alone.

One model that may support the interpretation of an overarching self-compassion factor in addition to six separate subscale factors is a bifactor model (Reise, Bonifay & Haviland, 2013). In a bifactor model, individual scale items load on a general or "target" factor as well as a subscale or "group" factor. Considering how this model applies to the SCS, the target factor is the general self-compassion factor and the group factors consist of the six subscale factors. As Reise, Moore, and Haviland (2010) discuss, in a bifactor model the general factor is modeled as directly influencing individual item responses, and the ways in which individual items form group factors are also modeled. Moreover, none of the factors (target or group) are allowed to correlate in a bifactor model. This is because the subscale factors are posited to account for the shared variance in their respective set of item responses over and above the variance accounted for by the target (e.g., self-compassion) factor. In contrast, a higher-order model posits that while the target factor explains the correlation among group factors, there is no direct effect of

the target factor on individual items - the link between the target and items is modeled as being only indirectly related through the group factors. See Figure 1 for examples of a bifactor model and higher-order model as they apply to the SCS. It should be noted that the theoretical assumption that the target factor has no direct influence on individual item responses is a strong one. In contrast, "the bifactor model specifies that there is a single (general) trait explaining some proportion of common item variance for all items, but that there also are group traits explaining additional common variance for item subsets. The general and group factors are on equal conceptual footing and compete for explaining item variance—neither is 'higher' or 'lower' than the other (Reise et al., 2010, p. 547).

Although the bifactor model was first developed in the early 20th century (Holzinger & Swineford, 1937), it was not well known or commonly used in the psychometric literature (at least in the United States) when the SCS was first developed (Reise et al., 2013). Neff (2016) argues that the bifactor model is a more accurate way to represent her original conceptualization of self-compassion (Neff, 2003a), given that self-compassion is theorized to directly manifest in the particular ways that individuals respond to suffering (as represented by SCS scale items).

One benefit of using a bifactor model is that it allows for the straightforward calculation of the percent of total score variance that is accounted for by the general target factor, each group factor, and error. Although Williams et al. (2014) did not find support for the use of a total SCS score based on traditional model fit criteria, it might still be defensibly used. For instance, if the large majority of observed variance in SCS scores is explained by the general target factor, this would provide some sense of confidence that the scale could be used to measure the general factor even in the presence of multi-dimensionality. In contrast, if the majority of variance in scores is not attributable to the general target factor, this would argue against the use of a general

score. Reise et al. (2013) suggest .75 or higher as an ideal amount of variance explained by a general factor in order to confidently use a total scale score.

The current study examined the psychometric properties of the original English version of the SCS using a bifactor , higher-order, and six-factor correlated model in various populations, with the main goal of determining whether or not the use of an overall SCS score (in addition to the six subscale scores) is justified. We also examined the fit of a one-factor model (to be consistent with Williams et al., 2014), and a two-factor correlated model that separated positive and negative items in order to test the claims of López et al. (2015) and Costa et al. (2015) that the SCS is bi-dimensional. We examined the properties of the SCS in a sample of undergraduates (termed "Students"), community adults (termed "Community"), and individuals practicing Buddhist meditation (termed "Meditators"). Because we did not have direct access to a clinical sample, we obtained permission to re-analyze data from the sample of individuals with a history of recurrent depression previously examined by Williams et al. (2014; termed "Clinical").

Methods

Participants

Students. This sample included a total of 222 undergraduates (84 male and 138 female, *M* age = 20.94, *SD* = 2.03), who were randomly selected from an educational-psychology subject pool at a large Southwestern university. In terms of ethnicity, 57% of the sample self-identified as White, 22% Asian, 14% Hispanic, 3% Black, and 4% Other.

Community. The sample of community adults was recruited from Mechanical Turk, an online survey research recruitment method that samples from the general public. Mechanical Turk has been found to be much more nationally representative of the general population than college samples (Buhrmester, Kwang, & Gosling, 2011). Participants were directed to Survey

Monkey in order to take the study, and were paid 75 cents for completing it (see Buhrmester et al., 2011 for supporting evidence of validity at low payment levels). The sample had 1394 participants (35% male and 65% female; *M* age = 36.01, *SD* = 12.88). The ethnic breakdown was 77% White, 7% Black, 6% Asian, 6% Hispanic, and 6% Other.

Meditators. This sample included a total of 215 meditators (30% male and 70% female*; M* age = 47.40, *SD* = 11.59). The ethnic breakdown was 87% White, 2% Asian, 2% Hispanic, 2% Black, 7% Other. Participants were recruited via an e-mail that invited them to complete an online questionnaire via Survey Monkey. E-mails were sent to individuals affiliated with Seattle Insight Meditation Society, Spirit Rock, the Insight Meditation Society, and other similar groups. Participants reported a wide range in meditation experience from beginner to advanced (1 to 20 years of meditation practice). The average length of meditation practice for the sample was 6.67 years (*SD* = 3.86). Fifty-three percent of the participants identified as Buddhist, 26% identified as having no religious affiliation, 12% identified as Christian, 6% as Other, and 3% as Jewish.

Clinical. This is the same sample that was analyzed in the Williams et al. (2014) study. It initially included 405 participants (23% male and 77% female; *M* age = 50.16, *SD* = 11.08). Participants were recruited through primary care settings in the United Kingdom. Criteria for this group included having a diagnosis of recurrent major depressive disorder in full or partial remission according to the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV;* American Psychiatric Association, 1994), having three or more previous major depressive episodes, and being 18 or older. For a full description of the sample, see Williams et al. (2014). Note that as in the original study, participants with any missing data on the SCS were excluded, leaving a remaining sample of *N* = 390.

Measures

Self-Compassion. Participants in all four samples completed the SCS (Neff, 2003a) described above. This 26 item self-report measure includes 5 self-kindness items, 5 self-judgment items, 4 common humanity items, 4 isolation items, 4 mindfulness items, and 4 over-identification items. Responses are given on a 5-point Likert scale ranging from 1 (*Almost Never*) to 5 (*Almost Always*). Negative items are reverse-coded so that higher scores represent a lower frequency of these responses. For scoring purposes, means are calculated for each subscale, and a grand mean is calculated that represents an overall self-compassion score.

Psychometric Analyses

In the current study, five different models were used to examine the factor structure of the SCS in each of the four samples. Specifically, CFAs were conducted to examine a one-factor, two-factor correlated, six-factor correlated, higher-order, and a bifactor model. M*plus* software (version 7.4; Muthén & Muthén, 1998-2016) was used when conducting all of the CFAs. Because item responses on the Self-Compassion Scale (SCS) are ordinal in nature, ranging from 1 (*Almost Never*) to 5 (*Almost Always*), maximum-likelihood robust (MLR) estimation in M*plus* was used to estimate model parameters[1].

Model fit was evaluated globally using the comparative fit index (CFI), the Tucker-Lewis Index (TLI; also known as the NNFI), the root mean squared error of approximation (RMSEA) with accompanying 90% confidence interval (CI), and the standardized root mean square residual (SRMR). Because the goal of the study was to determine if use of a total SCS score is justified, and whether the SCS is a "good enough" measure of self-compassion or if it needs to be

[1]Normal theory maximum likelihood (ML) estimation has been shown to produce accurate parameter estimates for CFAs with ordered variables having five or more categories (Beauducel & Herzberg, 2006; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Nonetheless, ML robust (MLR) estimation was used in order to adequately correct for underestimated standard errors and inaccurate test statistics which tend to occur with ordered categorical variables when using ML estimation (Rhemtulla et al., 2012).

redesigned, we evaluated fit indices using the liberal criteria used in the study conducted by Williams et al. (2014). In other words, we felt the conclusion that a total score should *not* be used and that the SCS should be redesigned should only be drawn if model fit failed to meet these more liberal criteria. Thus, models associated with CFI and TLI values greater than or equal to .90 were deemed as acceptably fitting models (Bentler & Bonett, 1980). Models associated with RMSEA values equal to or less than .10 would indicate acceptable model fit using the liberal criteria (Browne & Cudeck, 1993; MacCallum, Browne, & Sugawara, 1996). The RMSEA value is associated with a 90% confidence interval that provides further evidence of acceptable model fit if the upper limit does not exceed a value of .10 (West et al., 2012). SRMR values equal to or less than .10 would also indicate adequate model fit using the liberal criteria (Hu & Bentler, 1999). Note that chi-square test statistics were not used because they tend to be highly sensitive to sample size (Marsh, Balla, & McDonald, 1988), distributional assumption violations of the data (Bentler & Bonnett, 1990), and the type of model misspecification present in the data (Gallini & Mandeville, 1984). Akaike's (1973) information criterion (AIC) for each model was also documented. This criterion is comparative, meaning that models associated with the smallest AIC values are selected as models that would most likely cross-validate in subsequent samples. Because a unique feature of the current study involved examining the bifactor model in multiple samples, multiple-group techniques were also used to examine the equivalence of the bifactor model's parameter estimates across the four populations.

One of the advantages of a bifactor model is that it allows for the calculation of different indices that represent the percentage of variance in scores attributable to all of the factors as well as the percentage of variance in scores attributable to the general factor only. The omega index, which is a ratio of true score variance to total variance and corresponds to internal consistency

reliability (Hancock & Mueller, 2001), represents the percentage of variance in the total scores accounted for by all of the factors. Omega hierarchical (omegaH; McDonald, 1999) is an index used to estimate the percentage of variance in the total scores that is attributed to the general or target factor (Hancock & Mueller, 2001; Reise et al., 2013).

For each of the four samples examined, standardized estimates from the bifactor model were used to calculate both omega and omegaH indices in order to determine the reliability associated with the SCS as well as the amount of variance on the SCS total scores that may be attributed to an overall self-compassion factor, respectively. To determine the amount of reliable variance (i.e., not due to error) in the SCS scores attributed to the general "self-compassion" factor, omegaH is divided by omega $\left(\frac{\omega_H}{\omega}\right)$.

Results

Descriptive statistics, including means, standard deviations and internal consistency reliability alphas, are presented in Table 1. Internal consistency coefficients are generally deemed acceptable when they are .80 or above (Lance, Butts, & Michels, 2006). The internal consistency estimates associated with the subscales and the overall self-compassion scale ranged from .70 to .95 with the majority falling above .80. Thus, the SCS total and subscale scores demonstrated acceptable internal reliability estimates of Cronbach alpha.

Table 2 shows the model fit indices associated with the five models analyzed for each sample. (It should be noted that the fit indices obtained for the clinical sample in the current study differed slightly from those reported in the original Williams et al. (2014) study because slightly different model estimation procedures were used. However, results did not differ in any substantive way.) As seen in Table 2, the bifactor model demonstrated adequate fit using the liberal criteria based on the CFI, RMSEA and SRMR indices in the Student, Community and

Meditator samples. The TLI supported adequate model fit in the Meditator sample, marginal fit in the Student and Community samples, and inadequate fit in the Clinical sample. For the higher-order model, the RMSEA and SRMR indices demonstrated adequate fit across all samples, whereas the TLI suggested inadequate model fit across all samples. The CFI indices were marginal for the student, community, and meditator samples and inadequate for the clinical sample. The six-factor correlated model demonstrated adequate model fit based on the CFI, TLI, RMSEA, and SRMR indices in the Student, Community and Meditator samples. In the Clinical sample, the RMSEA and SRMR indices supported adequate model fit, whereas the CFI and TLI indices suggested inadequate fit of the six-factor correlated model. The two-factor correlated and single factor models generally had poor fit across the four samples according to all fit criteria. The AIC suggested that, in all four samples, the six-factor correlated model would likely cross-validate the best, followed by the bifactor model, followed by the higher-order factor model, followed by the two-factor inter-correlated model, and finally the one-factor model.

Standardized factor loadings obtained with the six-factor correlated model of individual SCS items on their intended subscales are presented in Table 3, along with standardized factor loadings obtained with the bifactor model of individual SCS items on the general "self-compassion" factor. The loadings on the subscales in the six-factor correlated model were all statistically significantly different from zero[2] (at $p < .001$) and ranged in magnitude from .38 to .89. The loadings on the general factor in the bifactor model were all statistically significant (at $p < .001$) and ranged from .25 to .79. Tabachnick and Fidell (2007) suggest that factor loadings of .32 or above are considered meaningful, and Comrey and Lee (1992) suggest loadings of .55 and above are good because it means that approximately 30% of the variance in

[2] Statistical significance was assessed using *z* scores associated with each of the loadings. The *z* scores are calculated by dividing the loading value by its respective standard error.

the item is explained by the factor. Using these criteria, the vast majority of items had good factor loadings on their respective subscale factors in the six-factor correlated model, and all loadings were meaningful. Similarly, the large majority of items had good factor loadings on the general factor for the student, community, and meditator samples in the bifactor model, though fewer had good loadings for the clinical sample. In all samples, however, all items except one (mindfulness item 9 in the clinical sample) had loadings on the general self-compassion factor in the bifactor model that would be considered meaningful.

Table 4 presents the factor inter-correlations for the six-factor correlated model across the four samples, which ranged from .34 to .97 and were all statistically significant (at $p < .001$).

Multiple-group analysis was then conducted in order to assess strong or scalar measurement invariance. Specifically, we tested the equivalence of the factor loadings and the item intercepts for the bifactor model across the different samples. Because the Clinical sample had slightly lower model fit, invariance was tested using the Clinical sample as the reference group compared to the other three samples. A baseline model was first estimated in which the bifactor model was fitted simultaneously to the data from the two comparison samples without imposing factor loading or item intercept equality constraints across the samples. Then, the bifactor model in which all of the factor loadings were constrained to be equal across both samples was estimated and compared to the baseline model. The bifactor model with item intercepts constrained across both samples was subsequently estimated and compared to the bifactor model with only constrained factor loadings. As proposed by Cheung and Rensvold (2002), measurement invariance was supported if the ΔCFI between comparison models was equal to .01 or less. The ΔCFI between the baseline and the constrained factor loading model comparing the Clinical to the Student, Community, and Meditator samples was less than .01,

suggesting that the relationship between corresponding items and the SCS subscale factors and the overall SC factor were similar. The ΔCFI between the constrained factor loading model and the constrained intercept model was .01 or less for all three sample comparisons, suggesting that item intercepts are equal across the sample comparisons. This indicates that there is no systematic variation across samples that would suggest that some outside variable is systematically producing differences in the scores.

The estimates from the bifactor model were used to calculate the omega indices as previously described. The overall Omega index (ω), OmegaH, and the percentage of reliable variance accounted for by the "self-compassion" factor for each sample are presented in Table 5. As seen in the first column, the overall Omega index (ω) in all of the samples was .94 or greater, indicating that the large majority of total variance in the scores may be attributed to both the self-compassion and the six subscale factors. Thus, the variance in the total score due to error was never greater than 6% in any of the samples. Column two displays the OmegaH indices for each sample which ranged from .85 to .90, indicating that a large majority of the total variance in the scores can be attributed to the general, overall self-compassion factor. Further, results presented in column three suggest that the overall self-compassion factor accounts for 90% to 95% of the reliable variance (i.e., not due to error) in the total scores.

Discussion

The purpose of this study was to examine the factor structure of the SCS across a variety of populations, and, in particular, to address the question of whether the use of an overall self-compassion score is justified. Results suggested that the SCS had relatively good psychometric properties and that a total SCS score could be reliably interpreted in four different populations: student, community, meditator, and clinical, although findings for the clinical sample were more

mixed. First, it should be highlighted that mean levels of self-compassion differed across samples in a theoretically consistent manner. As would be expected given the Buddhist roots of the self-compassion construct, the Meditator sample reported the highest overall levels of self-compassion. Likewise, the Clinical sample reported the lowest levels, which would be expected given their diagnosis of recurrent major depressive disorder. Similarly, meditators reported the highest levels of self-kindness, common humanity, and mindfulness and lowest levels of self-judgment, isolation, and over-identification while the clinical sample displayed the opposite pattern.

Liberal fit criteria were used in interpreting the meaning of results to determine if the SCS can measure a general factor of self-compassion as intended or if it needs to be redesigned. CFAs were used to examine a bifactor model, a higher-order model, a six-factor correlated model, a two-factor correlated model, and a one factor model. Because of arguments that fit measures alone should not be the deciding factor in determining the validity of a scale's factor structure (e.g., Morgan, Hodge, Wells & Watkins, 2015), we also used the bifactor model to calculate omega indices to better inform judgments concerning the dimensionality of the SCS. Results generally found evidence for the idea that a total SCS score can be interpreted using a bifactor model structure, but not a higher order model structure.

The higher-order model demonstrated relatively poor fit across samples, even in a Student sample that was similar to the one in which the SCS was first developed. While the RMSEA and SRMR indices generally suggested adequate fit of the higher-order model in all four samples, the CFI and TLI indices suggested marginal or inadequate fit in all four samples. This suggests that the higher-order model is not representing the relationships among items satisfactorily, and is not the best way to understand the relationship between subscales or to

justify the use of a total SCS score. The fact that the higher-order model does not allow for modeling of the direct impact of the target variable of "self-compassion" on individual item responses, but instead posits an indirect effect only, does not theoretically align with the conceptual underpinnings of the SCS and is most likely the source of the poor model fit.

The bifactor model, in contrast, which simultaneously examines the contribution of a general factor and group factors to item variance in multidimensional measures, was found to demonstrate better fit to the data than the higher-order model. Using liberal model fit criteria, the bifactor model generally demonstrated acceptable fit according to most of the fit indices in the Student, Community, and Meditator samples. The bifactor model demonstrated suboptimal fit in the Clinical sample, however, indicating that the SCS may be operating slightly differently for clinical populations. While loadings on the general self-compassion factor in the bifactor model were generally good for the three non-clinical samples (Comrey & Lee, 1992), there were fewer items with good loadings for the clinical sample. It may be that individuals with a history of depression have different patterns of emotional regulation meaning that some items function a bit differently in relation to a general self-compassion construct for these individuals, though the precise reasons underlying these differences will need to be examined in future research. Still, when factor loadings and item intercepts were compared across samples, those for the clinical sample were highly similar to those found for the other three samples. Moreover, all but one item (mindfulness item 9) had meaningful loadings on the general factor according to Tabachnick and Fidell's (2007) criteria. This suggests that the SCS had the same basic factor structure in all four populations, although fit was not as good for the clinical population.

Fit indices also supported the six-factor correlated structure of the SCS across samples. CFA results indicated that the six-factor correlated model demonstrated adequate model fit based

on the CFI, TLI, RMSEA, and SRMR indices in the Student, Community, and Meditator samples, though fit was deemed suboptimal based on the CFI and TLI indices in the Clinical sample. In all of the samples, however, the vast majority of factor loadings for the six-factor correlated model between an item and its respective subscale factor were good (Comrey & Lee, 1992), and all loadings were meaningful (Tabachnick & Fidell, 2007). As with the general self-compassion factor, when factor loadings and item intercepts for the six subscales were compared across samples, those for the clinical sample were highly similar to those found for the other three samples.

One might argue that because the six-factor correlated model provided the best fit to the data, that the SCS subscales should be used instead of a total score. According to Reise, Bonifay, and Haviland (2013), however, "even in the presence of multidimensionality, total scale scores justifiably can be interpreted" (p. 132). Firstly, note that the six factors were shown to have a high degree of inter-correlation across samples, suggesting that the subscale factors are operating in concert with a "system" view of self-compassion. A benefit of using a bifactor model in psychometric analyses is that instead of relying on model fit criteria alone (which is a debatable issue in itself), one can estimate how much variance in the total score is explained by a general factor as well as subscale factors, allowing for a more nuanced and tangible measure of a scale's ability to be used as intended. Omega values indicated that at least 94% of the variance in the total scores was due to the general self-compassion factor as well as the six subscale factors. Thus, the largest amount of variance in the SCS total scores that could be attributed to error in any sample was 6%. Results also indicated that 90% or more of the reliable variance in total SCS scores was attributable to an overall self-compassion factor in all populations examined, including the Clinical sample. This value greatly exceeds the value of 75% suggested by Reise

et al. (2013) as warranting confidence in the use of a total scale score. Overall, these findings provide support for the use of a total SCS score as a reliable measure of self-compassion, even in clinical populations.

Of course, one could also argue that because the vast majority of variance was explained by the general self-compassion factor, that the six subscales should not be examined independently. Our position is that because the six-factor correlated model was found to have the best fit according to model comparisons using the AIC across samples, it is valid to examine the six subscales independently. However, researchers should keep in mind that these are all aspects of self-compassion and not wholly separate entities. In other words, while it is possible to examine the six subscales to answer questions such as which aspect of self-compassion is the strongest predictor of a particular outcome, or to examine group differences in the six aspects, it is probably best to examine the six subscales in addition to an overall score rather than examining one or more of the subscales completely on their own. These are interdependent parts of a whole, and should be understood and examined as such.

Note that the one and two-factor models examined for the SCS had poor fit in all samples. While no theorists we are aware of have argued that the SCS is unidimensional, there are some who have argued that the SCS is bidimensional, with a single "self-compassion" factor consisting of all the positive items and a single "self-criticism" factor consisting of all the negative items (Costa et al., 2015; Gilbert, McEwan, Matos & Rivis, 2011; López et al., 2015; Muris, 2015). Because Social Mentality Theory (Gilbert, 2005) posits that "self-compassion" taps into the mammalian caregiving system (associated with the parasympathetic nervous system), while "self-criticism" taps into the threat defense system (associated with the sympathetic nervous system), Gilbert and colleagues have argued that positive and negative self-affect should *not* be

represented by an overall scale score (Gilbert et al., 2011). However, given that the sympathetic and parasympathetic nervous systems continuously interact and co-vary (Porges, 2001), from our point of view there is no reason why a single summary score cannot be used to assess the relative balance of system components.

Muris (2015) has also proposed that the positive and negative items of self-compassion form separate constructs, and that self-compassion should only be assessed using the positive items because the negative items are conflated with psychopathology. However, a recent study by Krieger, Berger and Holtforth (2016) which used cross-lagged analyses to determine whether changes in self-compassion led to changes in depression or the reverse, found not only that self-compassion predicted depression (rather than the reverse) but that findings were the same whether one examined a total scale score, a positive factor only or a negative factor only. They interpreted their findings as evidence that self-compassion should be considered an overall construct rather than two separate constructs, and current findings are congruent with this interpretation.

There are also theoretical problems with collapsing the three positive and three negative components into two separate factors given that it would obscure important differences between components. For instance, it would make it impossible to distinguish factors such as self-kindness and mindfulness, which are likely to tap into differing neurological and physiological systems (Engen & Singer, 2016). Our findings that the six-factor inter-correlated and bifactor models had a better fit than the two-factor model suggest that the SCS can be seen as having six subscale factors and a general factor of self-compassion simultaneously, rather than being comprised of two factors, one positive and one negative.

The large majority of researchers to date have chosen to examine self-compassion as an overall construct, most likely because it is simpler to conceptualize self-compassion as a single state of mind that encompasses the compassionate versus uncompassionate ways that individuals emotionally respond, cognitively understand, and pay attention to their feelings of personal inadequacy and experiences of suffering. Also, research interest in self-compassion is often motivated by its potential implications for psychological interventions, and programs designed to teach self-compassion tend to impact all its aspects simultaneously (Neff, 2016). It is simpler to analyze and report such findings with the use of a total SCS score, and current findings support the use of the SCS in this manner. For particular research questions, however, it may also be appropriate to analyze the six subscales separately in addition to using a total score. For instance, one might want to determine if some people struggle with certain aspects of self-compassion more than others or understand which components of self-compassion are most strongly related to particular outcome variables. Körner et al., (2015) for example, used regression analyses to examine which of the six self-compassion components most strongly predicted depressive symptoms in a large, non-clinical community sample, and found that isolation predicted 18% of the variance in symptoms, followed by over-identification and self-kindness which each predicted 2%, and mindfulness and self-judgment, which each predicted 1%. These types of studies are useful as they provide clues about which aspects of self-compassion might be important to target in interventions designed to address particular outcomes like depressive symptomology. One of the benefits of the SCS is that it can be used flexibly in order to meet particular research needs.

In summary, findings of the current study using a bifactor model support the conclusion that a total SCS scale score can be appropriately used to measure the general construct of self-

compassion, while findings using a six-factor correlated model suggest it can also be used to examine the six constituent components of self-compassion. Given that the higher-order model was not well supported in any of the samples examined, and that a bifactor model appears to be a superior way to model Neff's (2003b) conceptualization of self-compassion, results suggest that future attempts to validate translations of the SCS or to examine the properties of the SCS in specific populations should not attempt to justify the use of a total SCS score using a higher-order model. Instead, researchers should examine a bifactor model (including estimating the amount of reliable variance that can be attributed to an overall self-compassion score) in addition to a six-factor correlated model to investigate validity.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory.* Budapest: Akademiai Kiado.

Albertson, E. R., Neff, K. D., & Dill-Shackleford, K. E. (2014). Self-compassion and body dissatisfaction in women: A randomized controlled trial of a brief meditation intervention. *Mindfulness*, 1-11.

Allen, A. B., Goldwasser, E. R., & Leary, M. R. (2012). Self-compassion and well-being among older adults. *Self and Identity*, *11*(4), 428-453.

Arch, J. J., Brown, K. W., Dean, D. J., Landy, L. N., Brown, K. D., & Laudenslager, M. L. (2014). Self-compassion training modulates alpha-amylase, heart rate variability, and subjective responses to social evaluative threat in women. *Psychoneuroendocrinology*, *42*, 49-58.

Barnard, L. K., & Curry, J. F. (2011). Self-compassion: Conceptualizations, correlates, & interventions. *Review Of General Psychology, 15*(4), 289–303.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 14*, 186-203.

Behling, O. & Law, K. S. (2000). *Translating Questionnaires and Other Research Instruments: Problems and Solutions*. Thousand Oaks, CA: Sage Publications.

Bentler, P. M., & Bonnett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

Blatt, S. J., D'Afflitti, J., & Quinlan, D. (1976). Experiences of depression in young adults.

Journal of Abnormal Psychology, 65, 383–389.

Breines, J. G., & Chen, S. (2012). Self-compassion increases self-improvement motivation. *Personality and Social Psychology Bulletin*, *38*(9), 1133-1143.

Breines, J. G., McInnis, C. M., Kuras, Y. I., Thoma, M. V., Gianferante, D., Hanlin, L., ... & Rohleder, N. (2015). Self-compassionate young adults show lower salivary alpha-amylase responses to repeated psychosocial stress. *Self and Identity*, *14*(4), 390-402.

Breines, J. G., Thoma, M. V., Gianferante, D., Hanlin, L., Chen, X., & Rohleder, N. (2014). Self-compassion as a predictor of interleukin-6 response to acute psychosocial stress. *Brain, Behavior, and Immunity*, *37*, 109-114.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park, CA: Sage.

Buhrmester, M., Kwang, T., & Gosling, S.D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high quality, data? *Perspectives on Psychological Science, 6*(1), 3-5.

Castilho, P., Pinto-Gouveia, J., & Duarte, J. (2015). Evaluating the multifactor structure of the long and short versions of the Self-Compassion Scale in a clinical sample. *Journal of Clinical Psychology*.

Chen, J., Yan, L., & Zhou, L. (2011). Reliability and validity of Chinese version of Self-compassion Scale. *Chinese Journal Of Clinical Psychology*, 19(6), 734-736.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness of fit indexes for testing measurement invariance. *Structural Equation Modeling, 9(2)*, 233-255.

Comrey, A., & Lee, H. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Costa, J., Marôco, J., Pinto‐Gouveia, J., Ferreira, C., & Castilho, P. (2015). 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*.

Daye, C. A., Webb, J. B., & Jafari, N. (2014). Exploring self-compassion as a refuge against recalling the body-related shaming of caregiver eating messages on dimensions of objectified body consciousness in college women. *Body Image*, *11*(4), 547-556.

Diedrich, A., Grant, M., Hofmann, S. G., Hiller, W., & Berking, M. (2014). Self-compassion as an emotion regulation strategy in major depressive disorder. *Behaviour Research and Therapy*, *58*, 43-51.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*(3), 440-464.

Engen, H. G., & Singer, T. (2016). Affect and motivation are critical in constructive meditation. *Trends in cognitive sciences*, *20*(3), 159-160.

Finlay-Jones, A. L., Rees, C. S., & Kane, R. T. (2015). Self-Compassion, emotion regulation and stress among Australian psychologists: testing an emotion regulation model of self-compassion using structural equation modeling. *PloS One*, *10*(7).

Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An Introduction*. Thousand Oaks, CA: Sage Publications.

Gallini, J. K., & Mandeville, G. K. (1984). An investigation of the effects of sample size and specification error on the fit of structural equation models. *Journal of Experimental Education, 53*, 9-19.

Garcia-Campayo, J., Navarro-Gil, M., Andrés, E., Montero-Marin, J., López-Artal, L., &

Demarzo, M. M. (2014). Validation of the Spanish versions of the long (26 items) and short (12 items) forms of the Self-Compassion Scale (SCS). *Health and Quality of Life Outcomes*, *12*(1), 4.

Gilbert, P. (2005). *Compassion: Conceptualizations, research and use in psychotherapy*. London: Brunner-Routledge.

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.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. *Structural Equation Modeling: Present and Future*, 195-216.

Hollis-Walker, L., & Colosimo, K. (2011). Mindfulness, self-compassion, and happiness in non-meditators: A theoretical and empirical examination. *Personality and Individual Differences*, 50, 222-227.

Holzinger, K. J., & Swineford, R. (1937). The bifactor method. *Psychometrika, 2*, 41-54.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Hupfield, J. & Ruffieux, N. (2011) Validierung einer deutschen Version der Self-Compassion Scale. *Zeitschrift für Klinische Psychologie und Psychotherapie,* 40 (2), 115–123.

Johnson, E. A., & O'Brien, K. A. (2013). Self-compassion soothes the savage ego-threat system: effects on negative affect, shame, rumination, and depressive symptoms. *Journal of Social and Clinical Psychology*, *32*(9), 939-963.

Körner, A., Coroiu, A., Copeland, L., Gomez-Garibello, C., Albani, C., Zenger, M., & Brähler, E. (2015). The role of self-compassion in buffering symptoms of depression in the

general population. *PLoS ONE*, *10*(10), e0136598

Krieger, T., Berger, T., & grosse Holtforth, M. (2016). The relationship of self-compassion and depression: Cross-lagged panel analyses in depressed patients after outpatient therapy. *Journal of affective disorders*, *202*, 39-45

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria. What did they really say? *Organizational Research Methods, 9*(2), 202-220.

Leary, M. R., Tate, E. B., Adams, C. E., Allen, A. B., & Hancock, J. (2007). Self-compassion and reactions to unpleasant self-relevant events: The implications of treating oneself kindly. *Journal of Personality and Social Psychology, 92,* 887-904.

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).

MacBeth, A., & Gumley, A. (2012). Exploring compassion: A meta-analysis of the association between self-compassion and psychopathology. *Clinical Psychology Review, 32*, 545-552.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130-149.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391-410.

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, E. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*, 181-220.

Marsh, H. W., Hau, K.-T., & Grayson, E. (2005). Goodness of fit in structural equation modeling. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics* (pp. 275-340). Mahwah, NJ: Erlbaum.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of the golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in over-generalizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Erlbaum.

Morgan, G., Hodge, K., Wells, K., & Watkins, M. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence, 3*, 2-20.

Muris, P. (2015). A protective factor against mental health problems in youths? A critical note on the assessment of self-compassion. *Journal of Child and Family Studies*, 1-5.

Muthén, L. K., & Muthén, B. O. (1998-2016). *Mplus user's guide* (7th ed.). Los Angeles: Authors.

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

Neff, K. D. (2003b). Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity, 2,* 85-102.

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., Beretvas, S. N. (2013). The role of self-compassion in romantic relationships. *Self and Identity, 12(1),* 78-98.

Neff, K. D., & Germer, C. K. (2013). A pilot study and randomized controlled trial of the mindful self-compassion program. *Journal Of Clinical Psychology, 69(1)*, 28-44.

Neff, K. D., Hseih, Y., & Dejitthirat, K. (2005). Self-compassion, achievement goals, and coping with academic failure. *Self and Identity, 4,* 263-287.

Neff, K. D., Kirkpatrick, K. & Rude, S. S. (2007). Self-compassion and its link to adaptive psychological functioning. *Journal of Research in Personality, 41,* 139-154.

Neff, K. D., Pisitsungkagarn, K., & Hseih, Y. (2008). Self-compassion and self-construal in the United States, Thailand, and Taiwan. *Journal of Cross-Cultural Psychology, 39,* 267-285.

Neff, K. D., Pommier, E. (2013). The relationship between self-compassion and other-focused concern among college undergraduates, community adults, and practicing meditators. *Self and Identity, 12(2),* 160-176.

Neff, K. D., & Rude, S. S., & Kirkpatrick, K. (2007). An examination of self-compassion in relation to positive psychological functioning and personality traits. *Journal of Research in Personality, 41,* 908-916.

Odou, N., & Brinker, J. (2014). Exploring the relationship between rumination, self-compassion, and mood. *Self and Identity*, *13*(4), 449-459.

Petrocchi, N., Ottaviani, C., & Couyoumdjian, A. (2013). Dimensionality of self-compassion: translation and construct validation of the self-compassion scale in an Italian sample. *Journal of Mental Health*, (0), 1-6.

Porges, S. W. (2001). The polyvagal theory: Phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology*, *42*(2), 123-146.

Raes, F. (2010). Rumination and worry as mediators of the relationship between self-compassion and depression and anxiety. *Personality and Individual Differences, 48*, 757–761.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129-140.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544-559.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354-373.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507-514.

Sbarra, D. A., Smith, H. L. & Mehl, M. R. (2012). When leaving your Ex, love yourself: Observational ratings of self-compassion predict the course of emotional recovery following marital separation. *Psychological Science. 23(3),* 261–269.

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, 377-389.

Smeets, E., Neff, K., Alberts, H., & Peters, M. (2014). Meeting suffering with kindness: effects of a brief self-compassion intervention for female college students. *Journal of Clinical Psychology, 70*(9), 794-807.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston, MA: Pearson.

Werner, K. H., Jazaieri, H., Goldin, P. R., Ziv, M., Heimberg, R. G., Gross, J. J. (2012). Self-compassion and social anxiety disorder. *Anxiety, Stress & Coping, 25 (5),* 543-558.

West, S. G., Taylor, A. G., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R.H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209-231). New York, NY: Guilford.

Williams, M. J., Dalgleish, T., Karl, A., & Kuyken, W. (2014). Examining the factor structures of the Five Facet Mindfulness Questionnaire and the Self-Compassion Scale. *Psychological Assessment*, *26*(2), 407-418.

Zessin, U., Dickhauser, O., & Garbade, S. (2015). The relationship between self-compassion and well-being: A meta-analysis. *Applied Psychology: Health and Well-Being*. *7*(3), 340-364.

Table 1
Descriptive statistics for the SCS total and subscales across samples

	Student (N = 222)			Community (N = 1394)			Meditator (N = 215)			Clinical (N = 390)		
	M	(*SD*)	α	*M*	(*SD*)	α	*M*	(*SD*)	α	*M*	(*SD*)	α
SCS	3.11	(.67)	.94	3.00	(.76)	.94	3.66	(.61)	.95	2.56	(.62)	.91
SK	3.07	(.77)	.86	2.92	(.88)	.86	3.61	(.56)	.82	2.50	(.82)	.80
SJ	3.00	(.81)	.85	3.11	(.96)	.85	2.64	(.78)	.89	3.64	(.78)	.78
CH	3.20	(.80)	.81	3.09	(.92)	.81	3.83	(.79)	.82	2.90	(.95)	.80
IS	2.87	(.84)	.77	3.16	(1.01)	.77	2.37	(.82)	.83	3.67	(.83)	.75
MI	3.29	(.78)	.80	3.23	(.86)	.80	3.95	(.68)	.83	2.94	(.81)	.74
OI	3.05	(.90)	.80	3.01	(1.01)	.83	2.49	(.75)	.82	3.69	(.79)	.70

Note. SCS = Total SCS score; SK = Self-Kindness Subscale; SJ = Self-Judgment Subscale; CH = Common Humanity Subscale; IS = Isolation Subscale; MI = Mindfulness Subscale; OI = Over-Identification Subscale. Note that the SJ, IS, and OI subscales were reverse-coded before calculating the total SCS score.

Table 2
Model Fit across Samples

Model	CFI	TLI	RMSEA (90% CI)	SRMR	AIC
Student Sample (N = 222)					
One-Factor	.79	.77	.09 (.08, .10)	.08	14438.04
Two-Factor Correlated	.88	.87	.07 (.06, .08)	.06	14191.63
Six-Factor Correlated	.93	.92	.05 (.05, .06)	.05	14047.81
Higher-Order	.89	.88	.07 (.06, .07)	.07	14153.43
Bifactor	.91	.89	.06 (.05, .07)	.06	14098.65
Community Sample (N = 1394)					
One-Factor	.74	.72	.10 (.09, .10)	.08	100063.94
Two-Factor Correlated	.88	.87	.07 (.06, .07)	.05	97446.58
Six-Factor Correlated	.94	.93	.05 (.04, .05)	.04	96229.10
Higher-Order	.89	.88	.06 (.06, .07)	.08	97214.10
Bifactor	.91	.89	.06 (.06, .06)	.07	96823.54
Meditator Sample (N = 215)					
One-Factor	.74	.72	.11 (.10, .11)	.09	12651.00
Two-Factor Correlated	.87	.86	.08 (.07, .09)	.06	12258.42
Six-Factor Correlated	.93	.92	.06 (.05, .07)	.06	12071.63
Higher-Order	.89	.88	.07 (.06, .08)	.08	12190.95
Bifactor	.91	.90	.07 (.06, .07)	.07	12115.61
Clinical Sample (N = 390)					
One-Factor	.64	.61	.11 (.11, .12)	.10	28263.28
Two-Factor Correlated	.83	.82	.08 (.07, .08)	.07	27495.59
Six-Factor Correlated	.88	.87	.07 (.06, .07)	.06	27307.50

Higher-Order	.80	.78	.09 (.08, .09)	.10	27639.30
Bifactor	.84	.81	.08 (.07, .08)	.09	27474.55

Note. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = root-mean-square error of approximation; CI = confidence interval; SRMR = standardized root-mean-square residual; AIC = Akaike information criterion.

Table 3

Factor loadings of SCS items on their intended subscale as found in the six-factor correlated model (SF) and on the general self-compassion factor as found in the bifactor model (GF) across samples

	Student *N* = 222		Comm. *N* = 1394		Med. *N* = 215		Clinical *N* = 390	
	SF	GF	SF	GF	SF	GF	SF	GF
Self-Kindness Items								
5. I try to be loving towards myself...	.71	.62	.75	.56	.76	.51	.75	.45
12. When I'm going through a very hard...	.76	.66	.80	.58	.77	.59	.81	.52
19. I'm kind to myself when I'm...	.75	.67	.80	.60	.84	.61	.75	.53
23. I'm tolerant of my own flaws...	.74	.75	.38	.33	.40	.43	.50	.56
26. I try to be understanding and patient...	.75	.70	.76	.63	.79	.60	.56	.46
Self-Judgment Items								
1. I'm disapproving and judgmental about...	.73	.64	.76	.71	.82	.75	.68	.61
8. When times are really difficult, I tend to...	.70	.62	.76	.69	.78	.69	.63	.51
11. I'm intolerant and impatient towards...	.65	.61	.69	.62	.75	.68	.62	.49
16. When I see aspects of myself that I..	.80	.75	.81	.76	.87	.79	.75	.64
21. I can be a bit cold-hearted towards...	.74	.68	.72	.66	.71	.60	.54	.48
Common Humanity Items								
3. When things are going badly for me, I...	.57	.38	.60	.38	.66	.45	.59	.42
7. When I'm down and out, I remind...	.75	.46	.75	.43	.73	.40	.66	.34
10. When I feel inadequate in some way, I...	.80	.51	.79	.42	.80	.51	.75	.47
15. I try to see my failings as part of the...	.75	.55	.69	.49	.73	.50	.80	.51
Isolation Items								
4. When I think about my inadequacies, it...	.66	.60	.76	.72	.65	.70	.70	.63
13. When I'm feeling down, I tend to feel...	.65	.56	.75	.65	.79	.59	.62	.51
18. When I'm really struggling, I tend to...	.63	.54	.72	.60	.81	.63	.56	.44
25. When I fail at something that's...	.75	.72	.77	.72	.75	.68	.70	.66
Mindfulness Items								
9. When something upsets me I try to keep...	.68	.57	.66	.49	.59	.32	.52	.25
14. When something painful happens I try...	.75	.60	.78	.58	.84	.52	.76	.47
17. When I fail at something important...	.79	.73	.78	.61	.85	.51	.68	.54
22. When I'm feeling down I try to...	.61	.52	.65	.51	.70	.55	.62	.48
Over-Identified Items								
2. When I'm feeling down I tend to obsess...	.79	.70	.80	.77	.80	.75	.66	.64
6. When I fail at something important to...	.72	.71	.78	.78	.77	.71	.69	.66
20. When something upsets me I get...	.65	.55	.71	.64	.65	.50	.50	.44
24. When something painful happens...	.66	.59	.68	.62	.69	.53	.52	.50

Note. All reported factor loadings were significant at $p < .001$.

Table 4
Correlations between the SCS Subscale Factors in the Six-Factor Correlated Model for the Student Sample (N = 222), Community Sample (N = 1394), Meditator sample (N = 215) and Clinical Sample (N = 390)

	Student	Comm.	Meditator	Clinical
SK-SJ	-.82	-.72	-.77	-.56
SK-CH	.77	.71	.72	.81
SK-IS	-.78	-.63	-.61	-.49
SK-MI	.88	.83	.85	.76
SK-OI	-.75	-.63	-.70	-.50
SJ-CH	-.48	-.44	-.53	-.34
SJ-IS	.86	.86	.83	.81
SJ-MI	-.68	-.60	-.64	-.39
SJ-OI	.92	.91	.91	.91
CH-IS	-.50	-.48	-.54	-.46
CH-MI	.80	.79	.77	.87
CH-OI	-.48	-.48	-.46	-.46
IS-MI	-.78	-.63	-.58	-.52
IS-OI	.88	.90	.82	.97
MI-OI	-.78	-.72	-.71	-.57

Note. SK = Self-Kindness Subscale; SJ = Self-Judgment Subscale; CH = Common Humanity Subscale; IS = Isolation Subscale; MI = Mindfulness Subscale; OI = Over-Identification Subscale. All correlations were significant at $p < .001$.

Table 5
Omega Estimates of Explained Variance from the Bifactor Model across Samples

	Omega ω	OmegaH ω_H	General SC Factor
Student	.95	.90	.95
Community	.95	.89	.93
Meditator	.96	.90	.94
Clinical	.94	.85	.90

Note. Omega = the proportion of variance in the total score accounted for by all factors. OmegaH = the proportion of variance in the total score accounted for by the SC factor. General SC Factor = the amount of reliable variance (not due to error) that is accounted for by the SC factor.

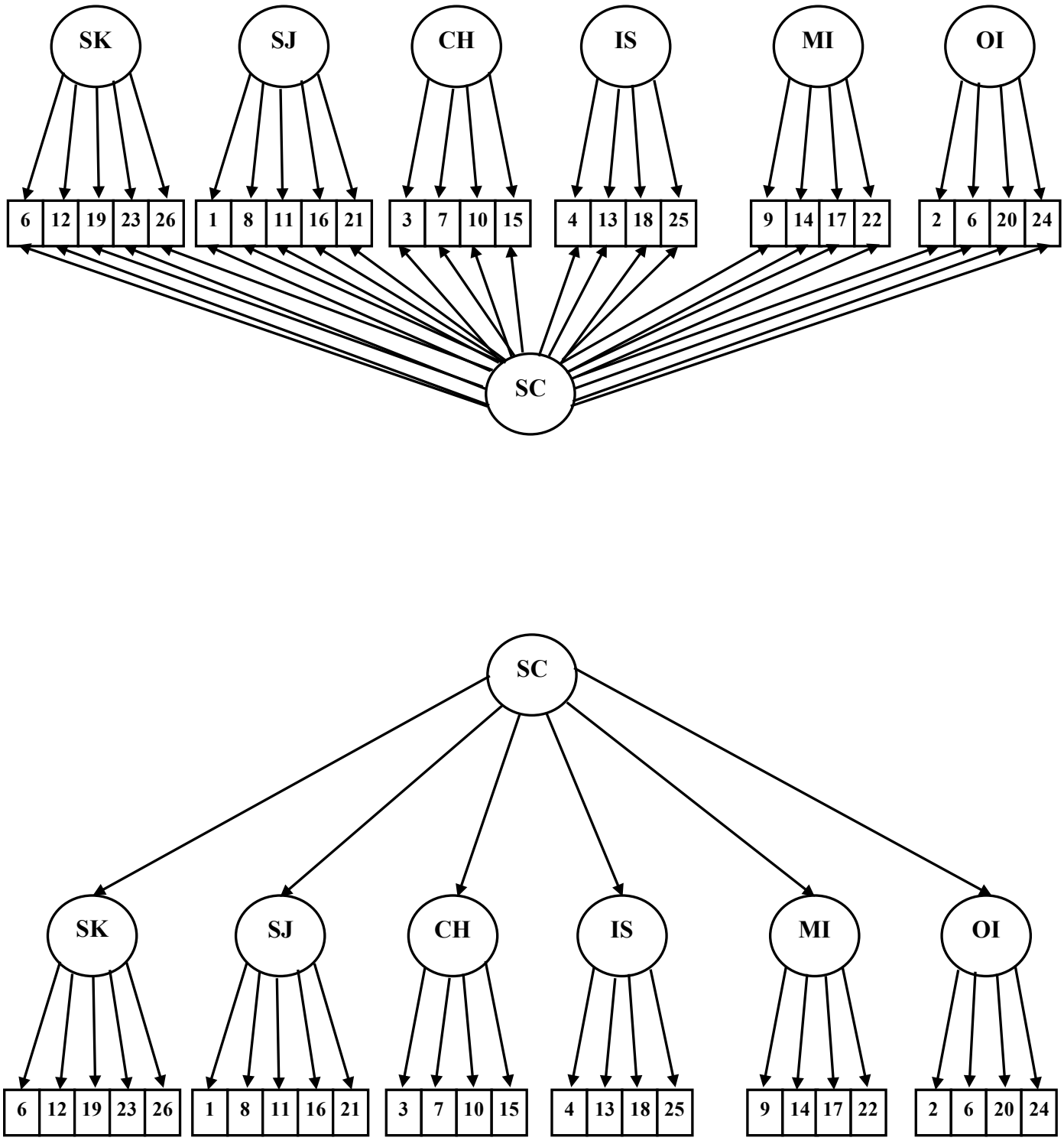

Figure 1. Comparison of a bifactor model (top) and higher-order model (bottom).
SC = Self-Compassion; SK = Self-Kindness; SJ = Self-Judgment; CH = Common Humanity;
IS = Isolation; MI = Mindfulness; OI = Over-Identification.