

# Examining psychometric properties and measurement invariance of a Chinese version of the Self-Compassion Scale – Short Form (SCS-SF) in nursing students and medical workers

This article was published in the following Dove Press journal:  
*Psychology Research and Behavior Management*

Runtang Meng,<sup>1,2,\*</sup> Yong Yu,<sup>2,3,\*</sup> Shouxia Chai,<sup>4,\*</sup> Xiangyu Luo,<sup>5</sup> Boxiong Gong,<sup>6</sup> Bing Liu,<sup>2,3</sup> Ying Hu,<sup>1,7</sup> Yi Luo,<sup>8</sup> Chuanhua Yu<sup>1,7</sup>

<sup>1</sup>Department of Preventive Medicine, School of Health Sciences, Wuhan University, Wuhan 430071, People's Republic of China; <sup>2</sup>Centre of Health Administration and Development Studies, Hubei University of Medicine, Shiyan 442000, People's Republic of China; <sup>3</sup>School of Public Health and Management, Hubei University of Medicine, Shiyan 442000, People's Republic of China; <sup>4</sup>School of Nursing, Hubei University of Medicine, Shiyan 442000, People's Republic of China; <sup>5</sup>Department of Cardiothoracic Surgery, Taihe Hospital, Hubei University of Medicine, Shiyan 442000, People's Republic of China; <sup>6</sup>Department of Oncology, Taihe Hospital, Hubei University of Medicine, Shiyan 442000, People's Republic of China; <sup>7</sup>Global Health Institute, Wuhan University, Wuhan 430072, People's Republic of China; <sup>8</sup>School of Nursing, Ningbo College of Health Sciences, Ningbo 315100, People's Republic of China

\*These authors contributed equally to this work

**Background:** Self-compassion has been regarded as a key psychological construct and a protective factor of mental health status. The focus of the present study was to adapt the Self-Compassion Scale (SCS) into Chinese, assess the validity and reliability of the measure and test measurement invariance (MI) across nursing students and medical workers.

**Methods:** The current study assessed the psychometric properties and invariance of the SCS-Short Form (SCS-SF) in two samples of 2676 from nursing students and medical workers. For construct validity, confirmatory and exploratory factor analyses (CFAs and EFAs) were conducted. Using Perceived Stress Questionnaire, Short Form-8 Health Survey (SF-8) and Goldberg Anxiety and Depression Scale, we evaluated concurrent validity and convergent/divergent validity. For reliability, internal consistency and test-retest analysis were employed. Multi-group analyses were conducted to examine MI of the different SCS-models across populations.

**Results:** CFA showed that the proposed six-factor second-order model could not be replicated and the six-factor first-order model was a reasonable to mediocre fitting model in both samples. EFA supported a three-factor structure which consisted of one positive and two negative factors. CFA confirmed that the hypothesized three-factor structure with 10 items ultimately was considered as the optimal model on the fitted results. The SCS-SF-10 (10 items form) also demonstrated acceptable internal consistency and test-retest reliability, as well as strong concurrent validity with measures of stress perception, health status, anxious and depressive symptoms. Convergent/divergent validity was not satisfactory. Multi-group CFAs provided support for the validity of the established models.

**Conclusion:** The Chinese version of the SCS-SF-10 has sound psychometric properties and can be applied to efficiently assess self-compassion in Chinese-speaking populations. The current study contributes to the identification and measurement of self-compassion after adversities.

**Keywords:** medical workers, nursing students, self-compassion, Chinese, measurement invariance, psychometric assessment

Correspondence: Chuanhua Yu  
Department of Preventive Medicine,  
School of Health Sciences, Wuhan  
University, Wuhan 430071, People's  
Republic of China  
Tel +86 276 875 9299  
Email yuchua@whu.edu.cn

## Introduction

Deriving from various Buddhist psychology thoughts and principles, Neff has innovatively proposed the multifaceted construct of self-compassion which simply represents compassion turned inward and involves approaching oneself perceived failure and inadequacy with kindness or alleviating personal suffering through

acceptance of unfavorable events.<sup>1-3</sup> Self-compassion entails three main theoretical facets, represented by pairs of which have a positive and negative pole that identifies compassionate versus uncompassionate behavior: self-kindness and self-judgment, a sense of common humanity and isolation, mindfulness and over-identification.<sup>3</sup> These elements, indeed, combine and mutually interact to create a self-compassionate frame of mind.<sup>4</sup> Self-compassion offers being benevolent and compassionate toward oneself while confronted with personal shortcomings and challenges, warmth and understanding the ubiquity of suffering, and being attuned to negative emotions without suppressing or exaggerating. An ever-increasing body of research indicates that this construct enables people to suffer less while also helping them to thrive.<sup>4</sup> Hence, the practice of self-compassion has garnered a great deal of attention in the scientific arena, especially on psychological health and psychopathology fields.<sup>5,6</sup>

More systematic studies suggest that self-compassion can be beneficial for target populations and is definitely considered as a protective factor against adverse mental health outcomes. Self-compassion may be a critical quality to cultivate for promoting positive health behaviors such as eating habits, exercise, sleep behaviors, and stress management, due in part to its association with adaptive emotions.<sup>7</sup> Self-compassion can buffer psychological stress and enhance psychological well-being<sup>8</sup>; a meta-analysis revealed a moderate effect size for the association between self-compassion and well-being.<sup>9</sup> A negative correlation existed between self-compassion and affective disorders of depression, anxiety and stress.<sup>5</sup> Another recent meta-analysis study has found that self-compassion is increasingly explored as a protective factor in relation to psychopathology.<sup>6</sup> Empirical studies and study protocols were all focused on specific populations such as medical samples, and found that self-compassion is associated with better medical adherence among people with fibromyalgia, chronic fatigue syndrome, and cancer, due in part to lower stress,<sup>10</sup> and that using controlled clinical trial to evaluate the effect of the interventions, whether mindfulness and self-compassion could reduce work stress and burnout in family and community medicine physicians and nurses.<sup>11</sup>

Alternatively, self-compassion links to the sensitivity to their own experience of suffering and encompasses three basic components—self-kindness, common humanity, and mindfulness. Self-kindness pertains to the support and understanding toward oneself independent of being self-critical; common humanity alludes to

recognizing that everyone fails, makes mistakes, and gets it wrong sometimes; and finally, mindfulness refers to awareness of our negative thoughts and emotions so that they are approached with balance and equanimity.<sup>4</sup> Additionally, it is important to highlight that self-compassion as a total construct is also broader in scope than mindfulness because it includes the additional elements of self-kindness and common humanity: actively soothing and comforting oneself when painful experiences arise, and remembering that such experiences are part of being human.

Self-Compassion Scale (SCS) has emerged as the most common instrument to measure this key psychological construct. According to theoretical analysis in the literature, the developer of the SCS initiative determines self-compassion into these three dimensions. The subscales then are compiled into each dimension and finally can be synthesized as total scale. A report on exploratory factor analysis (EFA) of all items in the SCS was not found in the relevant papers published by the original author. Up to the present, the SCS has been translated into at least 17 languages and was validated independently (or federatively) through their international counterparts (retrieved from: <https://self-compassion.org/self-compassion-scales-for-researchers/>, timestamp: 10/08/2019). There were two alternative forms concerning this scale: SCS – Long Form (SCS-LF or commonly just referred to as SCS)<sup>2</sup> and SCS – Short Form (SCS-SF).<sup>12</sup> The SCS-LF and SCS-SF were, respectively, developed by Neff, Raes and coworkers to assess how people treated themselves in difficult times.<sup>2,12</sup> The original SCS (SCS-LF) includes 26 items, measuring six dimensions of self-compassion (i.e. Self-Kindness = SK, Self-Judgment = SJ, Common Humanity = CH, Isolation = IS, Mindfulness = MI and Over-Identification = OI),<sup>2</sup> whereas the SCS-SF is comprised of a 12-item subset of the SCS. This shorter form of the SCS has been validated over time and its use as a mini-mental health assessment tool.<sup>12</sup>

In the original validation study, the developer identified a six-factor first-order structure with the six intercorrelated factors and a six-factor second-order (the term also known as “higher-order” and can be used interchangeably) model with a general self-compassion factor behind the six components<sup>2</sup> and subsequently chose the latter as a default model. Note that one of the most comprehensive studies to review and examine the SCS factor structure recently found that, using secondary data from 20 diverse samples (N=11,658) of Neff and coworkers, results supported the

use of a total SCS score or six subscale scores.<sup>13</sup> Another study in four distinct populations suggested that the six-factor correlated model demonstrated the best fit across samples and a total SCS score can be used as an overall measure of self-compassion.<sup>14</sup> Several replications have been conducted across the globe, but the findings are not without contradictions.<sup>15</sup> That is, this factor structure has not been consistently confirmed globally. For instance, two studies from Dutch<sup>16</sup> and Portuguese<sup>17</sup> suggested a two-factor solution with one positive factor and one negative factor (i.e., formed by the positively and negatively formulated items, respectively), one study in a Buddhist sample presented a model where the three positive-inter-correlated factors are present and a general negative factor representing all negative aspects,<sup>18</sup> a second study concerning the Chinese revision of SCS supported a three-factor structure.<sup>19</sup> Still another study from Turkish confirmed a one-factor model with a general self-compassion factor.<sup>20</sup> Furthermore, four studies in Greek-speaking populations supported a model that factor analyses established similar factor solutions to the English versions, but noticed that the first sudden change on the scree plot chart emerged after the third factor.<sup>21</sup>

The English version of the SCS-LF, not surprisingly, was judged easy to read and had relatively good validity and high reliability in diverse samples around the world. The SCS-SF is widely applied to different groups of population, especially in children,<sup>22</sup> adolescent,<sup>23,24</sup> and adult.<sup>25–27</sup> The findings from settings in different countries need to be compared and discussed. In Dutch, Spanish, Portuguese and Slovenian, the SCS-SF supported the same six-factor structure as found in the SCS-LF, showing adequate goodness of fit, as well as a single higher-order factor of self-compassion.<sup>12,25–27</sup> Also, the SCS-SF in these four studies demonstrated adequate internal consistency, test–retest reliability and a high correlation with the SCS-LF. In Chinese undergraduate students, the full version SCS, SCS-LF, suggested good reliability encompassing Cronbach's  $\alpha$  coefficient (0.84) and test–retest reliability (0.89, in 2-week interval) and supported a six-factor structure model.<sup>28</sup> However, it should be noted that factor loadings of the items on each factor are not clear, which may be caused by the close relationship between three dimensions (self-kindness, common humanity, and mindfulness) of the SCS, the way of expression of the items and the East-West cultural differences. Hence, this full Chinese version does not name the factors (subscales).<sup>28</sup> Another study in high school students<sup>19</sup> suggested that the full Chinese version SCS finally

retained 12 items (and yet deleted 14 items) with quality criteria as above mentioned in the literature. Note that this shorten version supported a three-factor structure, including the sense of common humanity, self-kindness and mindfulness. It has acceptable internal consistency for the total scale reliability (0.77).

Medical professionals, including nursing, are widely considered as a challenging occupation. Medical workers are particularly vulnerable to stress overload and compassion fatigue due to an emotionally exhausting environment.<sup>29</sup> Nursing students also are confronted with high levels of adverse mental health outcomes, such as stress, anxiety and depression.<sup>30,31</sup> Self-compassion has been recognized as a protective factor against adversities by cultivating and training program (or course) to improve the mental health of medical workers and nursing students.<sup>11,32</sup> Having compassion for others entails self-compassion. Enhancing focus on developing self-compassion holds promise for reducing perceived stress and increasing compassionate clinical care in health professionals.<sup>29</sup> Self-compassion, as one of the feasible intervention strategies, especially is relevant for health professionals. Hence, this study focused on the two key groups, nursing students and medical workers.

The aim of this study was to validate and assess some psychometric properties of the Chinese version of the SCS-SF. The reason for validating only a shorter Chinese version of the SCS was that there had been reported previously in China.<sup>19,28</sup> That is, the two translated versions used an SCS-LF in both of these studies. To investigate the SCS-SF on some psychometric indicators, samples of many different environments thus are being collected and analyzed.

## Materials and methods

### Measuring instruments

#### Self-Compassion Scale (SCS)

Forward–backward or/and counter-translation is surely one of the most commonly used translation process and have specific guidelines.<sup>33,34</sup> We established a triangular group including two PhD/MD students, one master student majoring in translation and interpreting, one master student majoring in nursing psychology and the first author as well as corresponding author who served as moderator. For a start, this procedure concerns a forward translation from the original language (English) to that of the target language (Chinese). Next, the target language (Chinese) is then back–translated into the original language (English) and compared to the original version. Inaccuracies in the

target language are simply identified through differences in meaning that occur in the backward translation. Differences in items were retranslated and modified until full agreement was achieved between the authors and the two groups' independent translators. To make items more easily understandable and avoid potential errors, the Chinese pre-final version was tested in a hospital including nine nursing students (trainee). After further modifications, we finalized the Chinese version SCS-SF (see [Supplementary material](#)). The response scale ranged from 1 (almost never) to 5 (almost always). To calculate a total self-compassion score mean (0–1), i.e. (raw score – the lowest possible score)/(the highest possible score – the lowest possible score), reverse score the negative subscale items,<sup>2</sup> with higher scores denoting greater levels of self-compassion.

### Perceived Stress Questionnaire (PSQ)

The PSQ was developed by Levenstein and colleagues to assess perceived stress in the last month or in the long run.<sup>35</sup> It would require participants to respond to 30 statements based on their subjective experiences (or/and feelings) rather than specific and objective state. Each item is scored from 1 (almost never) to 4 (usually), and eight items are reverse-scored to ensure accuracy of response. The resulting PSQ index similarly is linearly transformed between 0 (lowest possible level of stress) and 1 (highest possible level of stress); index is equal to (raw score – 30)/90.<sup>35</sup> Originally designed in English and Italian, this questionnaire has been translated into multiple languages and validated in different populations. The PSQ's authors granted us permission to introduce this instrument and authorized the final version in China.<sup>36</sup> Psychometric evaluation of a Chinese version of the PSQ (C-PSQ) was recently validated and applied.

### Short Form-8 Health Survey (SF-8)

The SF-8 Health Survey (SF-8), an alternate form to the most widely used SF-36 Health Survey (SF-36), is a member of the assessment of quality of life instruments.<sup>37</sup> It measures eight dimensions of health functioning, which accordingly are used to compute two summary measure scores (physical component summary, PCS and mental component summary, MCS).<sup>37</sup> The SF-8 has been translated into Chinese in prior studies, where Chinese version was validated in some large representative samples and was also available.<sup>38,39</sup> Similarly, the Medical Outcomes Study scoring system was still applied to this study.<sup>40</sup> Total scores are computed as the weighted sum of the scores for all items and the final score ranges between 0 and 100, with higher scores indicating better health status.

### Goldberg Anxiety and Depression Scale (GADS)

The Goldberg Anxiety Scale (GAS) and the Goldberg Depression Scale (GDS) formed from the GADS, each scale provided a count of symptoms consistent with depression and anxiety.<sup>41</sup> The GAS and the GDS ask respondents to report, "Yes" or "No" (respectively, assigning it one or zero point), that if they had felt in the respective nine anxiety symptoms and nine depression symptoms over the past month. Total score of each subscale is summed to reach a maximum of nine. The higher the total score is, the greater the level of symptomatology. Using the translation approaches mentioned above, the Chinese version of the GADS has been reported elsewhere and showed strong correlation with the C-PSQ only in nursing students.<sup>36</sup> The GADS presented a simple, quick and accurate method of screening depression and anxiety in the general population.

### Participants and procedure

A self-administered questionnaire survey method was used to collect data from 2676 valid samples in Ningbo and Shiyang, China. These two samples were, respectively, named NB and SY in the study. The participants were recruited from universities (colleges) and hospitals, which is closely related to medical field (see [Table 1](#)). Two of our papers did offer details on how to perform the sampling process in Ningbo.<sup>36,42</sup> Briefly, using traditional paper-and-pencil and centralized data collection, we tested nursing students from one of the higher vocational colleges (i.e., sample NB) by requesting feedback anonymously and confidentially. This investigation studied sample NB with the class as the unit to enroll. A pretest and a retest were also conducted to randomly select one class (50 students) after a 1-week interval in sample NB. Similarly, this study collected medical workers on questionnaire data with hospital and department as the unit. These participants were derived from three tertiary hospitals in Shiyang (i.e., sample SY). The only difference is that most of the medical workers were busy with work and scattered distribution, thereby developing into decentralized administration among some respondents. Each participant received a small incentive as compensation for his or her time when they completed the questionnaire: a piece of chocolate (or a small notebook and pen) worth 5 Chinese RMB (equal to 0.8 US dollars).

Before data collection, the ethics committee of Wuhan University School of Medicine gave their approval for the study design. The study was conducted in accordance with

**Table 1** Basic statistics on the sample and socio-demographic characteristics of participants

Variable	Total sample	Sample NB	Sample SY
Time range	Nov 2015 to Jan 2016	Nov 2015 to Jan 2016	Dec 2015 to Jan 2016
Location	Two cities	Ningbo	Shiyan
Composition	Students and workers	Junior college student	Medical personnel
Sampling method	Stratified sampling	Stratified random sampling	Stratified random sampling
Response rates	2676/2869 (93.27)	1453/1519 (95.66)	1223/1350 (90.59)
Sex			
Male	355 (13.27)	20 (1.38)	335 (27.39)
Female	2321 (86.73)	1433 (98.62)	888 (72.61)
Age, years	25.03±7.67	19.58±1.09	31.51±7.07
Self-assessment physical health			
Excellent	498 (18.61)	334 (22.99)	164 (13.41)
Good	1231 (46.00)	779 (53.61)	452 (36.96)
Average	836 (31.24)	316 (21.75)	520 (42.52)
Poor	108 (4.04)	24 (1.65)	84 (6.87)
Extremely poor	3 (0.11)	0 (0.00)	3 (0.24)
Self-assessment mental health			
Excellent	617 (23.06)	391 (26.91)	226 (18.48)
Good	1267 (47.35)	737 (50.72)	530 (43.34)
Average	705 (26.35)	300 (20.65)	405 (33.12)
Poor	83 (3.10)	24 (1.65)	59 (4.82)
Extremely poor	4 (0.14)	1 (0.07)	3 (0.24)

**Note:** The above table demonstrated N (%) or mean ± SD.

the declaration of Helsinki (revised form, version 2013).<sup>43</sup> All respondents in our study were informed that participation was voluntary. We obtained the written informed consent of all medical workers and nursing students prior to their participation.

## Analytical procedure

Given cross-cultural equivalence of the two translated versions of an instrument and uncertainty of factor analysis results in previous studies,<sup>12,19,28,33</sup> we performed confirmatory factor analysis (CFA) and EFA to examine factor structure of the SCS-SF, and then tested measurement invariance (MI)<sup>44</sup> across nursing students and medical workers using multi-group confirmatory factor analysis (MGCFA). Considering all that an adequate sample size with at least 200 is probably an appropriate threshold, 500 or more cases are strongly recommended in factor analysis.<sup>45,46</sup> Effective samples of the current study were in compliance with the requirements for CFA and EFA.

First and foremost, to evaluate the fit of the previously proposed SCS models, CFAs were used. We performed

two separate CFAs (one on the nursing students' sample and one on the medical workers' sample) for testing the original first-order and second-order six-factor structure, reflecting the theoretical operationalization/dimensionality of the construct.<sup>12</sup> The CFA-models were fitted for each group separately for nursing students and medical workers, as well as their combined sample. Maximum likelihood (ML) estimation can be applied to CFA because the distributions of the items were multivariate normal distribution.

In model fit of the single and multi-group CFAs, a combination of common indexes was applied to assess the global fit of the models to the data, based on the views of their distinct rules for good fit<sup>47–49</sup>: Normed chi-square (NC), with “good fit” if  $3?>NC>2?$ ,<sup>48</sup> and owing to chi-square test that is sensitive to sample-size<sup>49</sup>; the Tucker–Lewis index (TLI)<sup>50</sup> and the comparative fit index (CFI),<sup>51</sup> with “good fit” if  $>0.90$ ; the root mean square error of approximation (RMSEA) and RMSEA if  $\leq 0.06$  for “good fit,” RMSEA if in the 0.06–0.08 range for “adequate fit”;<sup>52</sup> the P of Close Fit

(PClose), with “good fit” if  $>0.05$ .<sup>53</sup> To compare the goodness-of-fit between the nested MI models, we followed the aforementioned recommendation of using differences in RMSEA, CFI, and TLI. Hereby, models with a change in CFI ( $\Delta\text{CFI}$ )  $\leq 0.010$ , change in RMSEA ( $\Delta\text{RMSEA}$ )  $\leq 0.015$ , and change in TLI ( $\Delta\text{TLI}$ )  $\leq 0.010$  were favored.<sup>54–56</sup> In addition, models were compared with a chi-square difference test. However, the consensus was that this may be an overly stringent criterion since  $\Delta\chi^2$ , in common with  $\chi^2$ , is dependent on sample size with a rejection of models with trivial practical misfit in large samples.<sup>57,58</sup>

Afterward, following suggestions on EFA,<sup>59</sup> the distribution of the items was examined to ensure there were not severe non-normalities. None of the SCS's items showed severe non-normal distribution. Two independent samples should be used when exploring (EFA) and then confirming (CFA) the internal structure of a tool. A series of EFA were conducted to further examine the factor structure of the SCS-SF. In case the CFA might not support the six-factor structure, especially in high-order factor structure, we intended to perform EFAs to explore the SCS-SF factorial structure. Based on the prior literature, EFAs indeed were conducted to investigate some different factor structures, included one, two or three factors as mentioned above.<sup>16,17,19–21</sup> ML method with an oblique rotation (promax, power coefficient = 4) was used since the objective was to identify latent underlying constructs and there were assumptions of the factors as being related in previous studies.<sup>12</sup> Principal component analysis (PCA) is not a factor extraction technique but simply an approach for data reduction, as it uses all the variance in the observed variables without discriminating between shared and unique variance. Actually, the results of factor extraction were the same for both varimax rotation and oblique rotation. We performed in subsequent EFAs on the data of nursing students and medical workers to explore the underlying structure of the SCS-SF. Then, CFAs were performed orderly on the nursing students' sample, the medical workers' sample and the combined sample to verify the identified factor solution.

The number of factors extracted was determined by reviewing the scree plot and considering the following criteria: eigenvalues ( $>1$ ), items content and interpretability and proportion of total variance explained by extracted factors combined (usually  $>60\%$ , at least  $50\%$ ).<sup>47</sup> Of these, adequate item-total correlation should fluctuate between above 0.20 and below 0.80; items with lower correlation should be discarded.<sup>60,61</sup> Factor loadings above

0.45 were considered important, based on 20% overlapping variance (fair) proposed by Comrey and Lee.<sup>45</sup> Due to the large-sample size, loadings above 0.30 (only for sample sizes of 350 or greater) or even 0.20 were considered significant.<sup>47,62</sup> Cross-loadings, items loaded on more than one dimension, should not exceed 0.32, while keeping a gap of at least 0.2 between the target loadings (factor loadings) and each of the cross-loadings. Additionally, all variables with communalities should also be over 0.50 to ensure sufficient explanation.<sup>47</sup>

Finally, MGCFAs were performed to test whether aspects of the factorial structure replicated in a separate sample, that is, across nursing students and medical workers groups for purposes of cross-validation. A two-step procedure, first-order model and second-order model fitting, was used by testing these samples. Three commonly tested levels, that are factor loadings, intercepts, and residual variances, were investigated by conducting multi-group factor models to examine the sensitivity of goodness of fit indexes.<sup>55</sup>

Reliability links to the consistency and precision of a measurement.<sup>63</sup> Consistency was assessed with Cronbach's alpha ( $\alpha$ ), Guttman's lambda-2 ( $\lambda_2$ , a better reliability estimation method<sup>64</sup>), and item-total correlation; test-retest reliability. Both alpha and lambda-2, as the indicators of internal reliability coefficient, were included in the present study. Test-retest reliability (reproducibility) would be exposed through Pearson's correlation and the intraclass correlation coefficient (ICC) at an interval of 7 days. ICC was calculated using the single measures and two-way mixed-effects model with absolute agreement type in view of method and range in collecting retest data.<sup>65,66</sup> For measurement precision evaluation, standard error of measurement also was calculated to quantify the variability of measurement errors.<sup>67</sup> Validity refers to the true value and accuracy that a measure attempts to capture.<sup>63</sup> Validity was measured in terms of construct validity, concurrent validity and convergent/divergent validity. Construct validity and dimensional structure were evaluated through CFAs and EFAs in aspects of exploration, item reduction, confirmation and validation. Concurrent validity and convergent/divergent validity were examined by testing Pearson's correlations of the different forms of SCS-SF with the instruments mentioned above. The mean difference testing (Student's test) was used to compare levels of symptom between key groups. Besides statistical significance, effect size should be also considered for interpreting the magnitude of the

associations with the criterion measures, as well as the differences in mean SCS-SF scores between nursing students and medical workers. The established level of statistical significance was set at 5%.

The author set up the database using EpiData (version 3.1; Jens M. Lauritsen & Michael Bruus, Odense, Denmark) software. All data were analyzed in SPSS/PASW (version 18.0; SPSS Inc., Chicago, IL, USA) and AMOS, as well as AMOS Plugin from Gaskination's StatWiki.

## Results

### Validity analysis

Using a sequence of nursing students, medical workers, combined sample, we ran a series of CFA for the proposed six-factor model (first- and second-order) according to the steps outlined previously. The SCS's proposed second-order six-factor structure was considered to be a poor-fitting model and could not be replicated in this study. Some values of fit indices in the first-order six-factor structure were considered marginal, thereby providing a reason to mediocre fitting model (the first 3 lines, see [Tables 2](#) and [3](#)). For that reason, EFAs were conducted to analyze the factor structure of the SCS-SF in separate samples. Sampling adequacy for factor analysis was preferentially tested for nursing students and medical workers groups. The Kaiser–Meyer–Olkin indexes are equal to 0.778 (nursing students) and 0.770 (medical workers)  $>0.60$  and their Bartlett's test of sphericity were significant ( $P<0.001$ ), indicating that the items could be considered for good factor analyses.<sup>68</sup> Factor analysis in two separately samples extracted almost the same as three factors and the total cumulative was, respectively, 35.63% (nursing students) and 36.82% (medical workers). Inspection of eigenvalues, scree plot and item content and interpretability suggested a three-factor solution. A summary of the EFA and item-total correlations was presented in [Table 4](#), with component loadings (obtained from the pattern matrix) for the entire set of items and selected items shown. The two items, item 2 and item 10, existed with flaws of cross-loadings (item 2 on Factor 2 was  $-0.402$  in medical workers). In addition, item 2 and item 10 with item-total correlations were lower than 0.20 (0.096 and  $-0.009$  as well as 0.025 and 0.093, respectively). Still, we decided to accept this relatively low communality and keep most of items with a view to keeping scales of enough length and in light of its satisfactory other indexes. According to the criteria

mentioned earlier, we tried to retain 10 items, except for items 2 and 10, as the Chinese version of SCS-SF-10 (i.e., 10 items form).

Using these samples, we carried out a series of CFA to validate various factor structures (See [Table 2](#)). Self-compassion or self-compassion subscales were entered as endogenous variables whereas all 12 items (or removed items 2, 10) were considered exogenous variables. Because item 2 loaded on Factor 1 (nursing students) or Factor 2 (medical workers), we in sequence run CFAs to confirm the factor structure of the SCS-SF using nursing students, medical workers, combined sample. Owing to a poor fitting effect in high-order model and reduced results, this test only checked for the first-order model.

Finally, we in turn validated three-factor structure (included 10 items and 12 items form) based on the EFA results. Several indices were often used for assessing goodness of fit. On the whole, three-factor models were a better fit than six-factor models. Further exploration revealed that three-factor model removed item 2, and item 10 was thus a better fit than three-factor model (12 items form). As can be seen, three-factor structure (model) with 10 items ultimately was considered as the optimal model on the fitted results. Of these, these 10 items can be assigned to three factors on a regular basis. We renamed these three factors combined with the literature,<sup>16,17</sup> which consisted of one positive factor (Factor 1: C3, C5, C6, C7), two negative factors (Factor 2: C1, C4, C8, C9; Factor 3: C11, C12). In the meantime, to evaluate concurrent and convergent/divergent validity of the SCS-SF-10, we used the existing gold standard scales, like the Chinese version of the PSQ, SF-8 and GADS. Self-compassion scores were negatively associated with PSQ index ( $r=-0.517$ ,  $P<0.001$ ) and GADS scores ( $r=-0.451$ ,  $P<0.001$ ), positively associated with SF-8 scores ( $r=0.405$ ,  $P<0.001$ ). SCS-SF-10 and SCS-SF as well as its subscales intercorrelations and gold standard scales their reliability results were presented in [Tables 5](#) and [6](#).

In the second phase of the analyses, MI was tested across key groups (nursing students vs. medical workers). [Tables 2](#) and [3](#) present the details of the CFA-models and model fitting results for each level of MI. Overall, some  $\Delta\chi^2$  ( $\Delta df$ ) values were significant, yet difficult to interpret given our large-sample sizes. Although most of  $\chi^2$  and  $\Delta\chi^2$  test were significant (except in the few), other model fit indices ( $\Delta TLI$ ,  $\Delta CFI$ ,  $\Delta RMSEA$ ) in the first- and second-order three-factor model (removed items 2 and 10) did not decrease more than the recommended cut-off values,

Table 2 First-order model fitting results for each measurement invariance level of the SCS-SF across nursing students and medical workers

Groups	Models	NC	TLI	CFI	RMSEA	PClose	$\Delta\chi^2$	$\Delta df$	P	$\Delta TLI$	$\Delta CFI$	$\Delta RMSEA$
Total sample	Six-factor <sup>a</sup>	11.607 (452.660/39)	0.875	0.926	0.063	0.000						
Nursing students	Six-factor <sup>a</sup>	5.090 (198.524/39)	0.908	0.946	0.053	0.234						
Medical workers	Six-factor <sup>a</sup>	7.524 (293.439/39)	0.841	0.906	0.073	0.000						
	M1	6.307 (491.971/78)	0.876	0.927	0.045	0.991	6.947	6	0.326	0.008	-0.001	-0.002
	M2	5.940 (498.919/84)	0.884	0.926	0.043	0.999	44.392	21	0.002	0.018	-0.004	-0.003
	M3	5.174 (543.311/105)	0.902	0.922	0.040	1.000	28.869	12	0.004	0.007	-0.003	-0.002
	M4	4.890 (572.179/117)	0.909	0.919	0.038	1.000						
Total sample	Three-factor <sup>b</sup>	11.610 (592.092/51)	0.875	0.904	0.063	0.000						
Nursing students	Three-factor <sup>b</sup>	5.331 (271.869/51)	0.903	0.925	0.055	0.112						
Medical workers	Three-factor <sup>b</sup>	7.294 (371.991/51)	0.846	0.881	0.072	0.000						
	M1	6.312 (643.870/102)	0.876	0.904	0.045	0.997	20.393	9	0.016	0.007	-0.002	-0.002
	M2	5.984 (664.263/111)	0.883	0.902	0.043	1.000	12.085	6	0.060	0.005	-0.001	-0.001
	M3	5.781 (676.347/117)	0.888	0.901	0.042	1.000	31.765	12	0.002	0.007	-0.004	-0.001
	M4	5.489 (708.113/129)	0.895	0.897	0.041	1.000						
Total sample	Three-factor <sup>c</sup>	14.052 (716.644/51)	0.847	0.881	0.070	0.000						
Nursing students	Three-factor <sup>c</sup>	7.265 (370.521/51)	0.860	0.891	0.066	0.000						
Medical workers	Three-factor <sup>c</sup>	7.748 (395.168/51)	0.835	0.873	0.074	0.000						
	M1	7.507 (765.695/102)	0.848	0.882	0.049	0.625	27.347	9	0.001	0.008	-0.003	-0.001
	M2	7.145 (793.042/111)	0.856	0.879	0.048	0.857	13.042	6	0.042	0.006	-0.001	-0.001
	M3	6.890 (806.084/117)	0.862	0.878	0.047	0.949	29.327	12	0.004	0.010	-0.003	-0.002
	M4	6.476 (835.411/129)	0.872	0.875	0.045	0.996						
Total sample	Three-factor <sup>d</sup>	4.488 (143.614/32)	0.968	0.977	0.036	1.000						
Nursing students	Three-factor <sup>d</sup>	2.640 (84.480/32)	0.971	0.980	0.034	0.999						
Medical workers	Three-factor <sup>d</sup>	2.727 (87.268/32)	0.966	0.976	0.038	0.985						
	M1	2.684 (171.750/64)	0.969	0.978	0.025	1.000	19.309	7	0.007	0.000	-0.003	0.000
	M2	2.691 (191.059/71)	0.969	0.975	0.025	1.000	13.086	6	0.042	0.000	-0.001	0.000
	M3	2.651 (204.145/77)	0.969	0.974	0.025	1.000	27.881	10	0.002	0.000	-0.004	0.000
	M4	2.667 (232.026/87)	0.969	0.970	0.025	1.000						
Cutoff value		<2?-<3?	>0.90	>0.90	<0.06 or <0.08	>0.05			>0.05	$\leq 0.010$	$\leq 0.005$ or $\leq 0.010$	$\leq 0.015$

Notes: <sup>a</sup>Six-factor model and included in all 12 items; <sup>b</sup>three-factor model and included in all 12 items, item 2 loaded on Factor 1; <sup>c</sup>three-factor model and included in all 12 items, item 2 loaded on Factor 2; <sup>d</sup>three-factor model and removed items 2, 10; measurement invariance (MI); M1 Unconstrained, M2 Measurement weights, M3 Structural covariances, M4 Measurement residuals.

Abbreviations: NC, normed Chi-square, CMIN/DF; TLI, Tucker-Lewis index; CFI, comparative fit index; RMSEA, root mean square error of approximation; PClose, the P of close fit.

Table 3 Second-order model fitting results for each measurement invariance level of the SCS-SF across nursing students and medical workers

Groups	Models	NC	TLI	CFI	RMSEA	PClose	$\Delta\chi^2$	$\Delta df$	P	$\Delta TLI$	$\Delta CFI$	$\Delta RMSEA$
Total sample	Six-factor <sup>a</sup>	34.527 (1657.282/48)	0.606	0.713	0.112	0.000						
Nursing students	Six-factor <sup>a</sup>	17.995 (863.763/48)	0.619	0.723	0.108	0.000						
Medical workers	Six-factor <sup>a</sup>	17.685 (848.898/48)	0.592	0.703	0.117	0.000						
	M1	17.840 (1712.684/96)	0.606	0.714	0.079	0.000						
	M2	16.890 (1722.763/102)	0.628	0.713	0.077	0.000	10.079	6	0.121	0.022	-0.001	-0.002
	M3	16.238 (1737.443/107)	0.644	0.711	0.075	0.000	14.679	5	0.012	0.016	-0.002	-0.002
	M4	16.169 (1746.285/108)	0.645	0.710	0.075	0.000	8.842	1	0.003	0.001	-0.001	0.000
	M5	15.395 (1755.055/114)	0.663	0.709	0.073	0.000	8.770	6	0.187	0.018	-0.001	-0.002
	M6	14.149 (1782.807/126)	0.692	0.706	0.070	0.000	27.752	12	0.006	0.029	-0.003	-0.003
Total sample	Three-factor <sup>b</sup>	4.488 (143.614/32)	0.968	0.977	0.036	1.000						
Nursing students	Three-factor <sup>b</sup>	2.640 (84.480/32)	0.971	0.980	0.034	0.999						
Medical workers	Three-factor <sup>b</sup>	2.727 (87.268/32)	0.966	0.976	0.038	0.985						
	M1	2.684 (171.750/64)	0.969	0.978	0.025	1.000						
	M2	2.691 (191.059/71)	0.969	0.975	0.025	1.000	19.309	7	0.007	0.000	-0.003	0.000
	M3	2.638 (195.542/73)	0.970	0.975	0.025	1.000	1.483	2	0.476	0.001	-0.000	0.000
	M4	2.604 (192.718/74)	0.970	0.975	0.024	1.000	0.176	1	0.675	0.000	-0.000	-0.001
	M5	2.651 (204.145/77)	0.969	0.974	0.025	1.000	11.427	3	0.010	-0.001	-0.001	0.001
	M6	2.667 (232.026/87)	0.969	0.970	0.025	1.000	27.881	10	0.002	0.000	-0.004	0.000
Cutoff value		<2? <3?	>0.90	>0.90	<0.06 or <0.08	>0.05			>0.05	≤0.010	≤0.005 or ≤0.010	≤0.015

**Notes:** <sup>a</sup>Six-factor model and included in all 12 items; <sup>b</sup>three-factor model and removed items 2, 10; measurement invariance (MI); M1 Unconstrained, M2 Measurement weights, M3 Structural weights, M4 Structural covariances, M5 Structural residuals, M6 Measurement residuals.

**Abbreviations:** NC, normed Chi-square, CMIN/DF; TLI, Tucker-Lewis index; CFI, comparative fit index; RMSEA, root mean square error of approximation; PClose, the P of close fit.

**Table 4** Summary of exploratory factor analysis with maximum likelihood and promax-rotation of the SCS-SF items

Item	Groups	Factor 1	Factor 2	Factor 3	Com	$r_i(t-i)$	Mean	SD	MC
C1 (R)	Nursing students	−0.044	<b>0.527</b>	0.036	0.293	0.281	2.654	0.928	★
	Medical workers	−0.031	<b>0.564</b>	−0.020	0.305	0.259	2.658	0.944	
C2	Nursing students	0.363	−0.280	0.148	0.179	0.096	3.562	0.785	
	Medical workers	0.382	−0.402	0.084	0.249	−0.009	3.604	0.800	
C3	Nursing students	<b>0.603</b>	0.029	0.050	0.378	0.380	3.681	0.735	★
	Medical workers	<b>0.661</b>	0.028	−0.019	0.441	0.400	3.733	0.777	
C4 (R)	Nursing students	0.038	<b>0.640</b>	−0.010	0.412	0.379	2.714	0.961	★
	Medical workers	0.019	<b>0.563</b>	0.043	0.345	0.326	2.694	0.960	
C5	Nursing students	<b>0.611</b>	−0.025	−0.036	0.368	0.312	3.687	0.728	★
	Medical workers	<b>0.539</b>	−0.075	0.010	0.286	0.301	3.684	0.757	
C6	Nursing students	<b>0.627</b>	0.076	−0.044	0.404	0.364	3.918	0.700	★
	Medical workers	<b>0.641</b>	0.075	0.031	0.435	0.434	3.754	0.797	
C7	Nursing students	<b>0.715</b>	0.083	0.005	0.534	0.427	3.784	0.730	★
	Medical workers	<b>0.761</b>	0.102	−0.001	0.609	0.480	3.728	0.813	
C8 (R)	Nursing students	0.048	<b>0.632</b>	0.038	0.433	0.396	2.887	0.946	★
	Medical workers	−0.002	<b>0.586</b>	0.064	0.381	0.339	2.795	0.945	
C9 (R)	Nursing students	−0.052	<b>0.614</b>	−0.004	0.370	0.315	2.474	0.864	★
	Medical workers	0.080	<b>0.657</b>	−0.033	0.433	0.389	2.550	0.880	
C10	Nursing students	0.321	−0.102	−0.175	0.142	0.025	3.160	0.889	
	Medical workers	0.359	−0.048	−0.188	0.161	0.093	3.386	0.876	
C11 (R)	Nursing students	−0.071	0.013	<b>0.590</b>	0.352	0.198	3.098	0.877	★
	Medical workers	−0.105	−0.013	<b>0.576</b>	0.327	0.121	3.007	0.918	
C12 (R)	Nursing students	0.067	0.037	<b>0.612</b>	0.410	0.320	3.371	0.902	★
	Medical workers	0.059	0.042	<b>0.641</b>	0.447	0.311	3.280	0.908	

**Notes:** The original SCS-SF factor structure: Self-Kindness Items C2, C6; Self-Judgment Items C11, C12; Common Humanity Items C5, C10; Isolation Items C4, C8; Mindfulness Items C3, C7; Over-Identification Items C1, C9; The Chinese SCS-SF structure: Factor 1 to Factor 3; i.e., [target loadings]  $\geq 0.45$ , [cross-loadings]  $< 0.32$ , [target loadings] − [cross-loadings]  $\geq 0.20$ ,  $h^2 \geq 0.50$  and  $r_i(t-i) \geq 0.2$ ; Bolded coefficients were target loadings at 0.45 or higher. The star meant that the items can basically attain Match Condition (MC) and remaining items were marked with an “★”.

**Abbreviations:** Com, communality;  $h^2$ ;  $r_i(t-i)$ , corrected item-total correlation; R, reversed scoring; MC, match condition.

indicating key groups invariance on three levels of factor loadings, intercepts, and residual variances. That is, MI could be achieved in either group. In contrast, some  $\Delta$ TLI values were above the recommended cut-off value in the first- and second-order six-factor model. Further, mainly model fit indices ( $\Delta$ TLI,  $\Delta$ CFI,  $\Delta$ RMSEA) in the first-order three-factor model (retained 12 items) also were up to the standard. In short, MI in three-factor model (removed items 2 and 10) across nursing students and medical workers was the best.

## Reliability analysis

The Chinese version of the SCS-SF consists of 12 items. Cronbach's alpha and Guttman's lambda-2 values of the

SCS-SF were 0.639 and 0.672, which indicates marginally acceptable reliability of this tool in the present study sample. Note that removed items 2 and 10, Cronbach's  $\alpha$  value increased to 0.686 whereas Guttman's lambda-2 value increased to 0.706 with an acceptable reliability value. The reliability coefficients for each subscale separately were 0.741 (Factor 1), 0.696 (Factor 2) and 0.540 (Factor 3) in addition to 0.744 (Factor 1), 0.697 (Factor 2) and 0.540 (Factor 3), respectively, corresponding to Cronbach's alpha and Guttman's lambda-2. Item-total correlation values had increased and strong central tendency when deleted these two items, which can meet quality criteria. In addition, test-retest reliability analysis for the pretest sample of 50 nursing students was 0.617 (Pearson's

**Table 5** SCS-SF-10 subscales intercorrelations and concurrent as well as convergent/divergent validity

Subscales	SCS-SF-10			
	Factor 1	Factor 2	Factor 3	Overall score
SCS-SF-10 <sup>a</sup>				
Factor 1		0.140	0.063	0.624
Factor 2			0.313	0.792
Factor 3				0.565
PSQ index <sup>a</sup>	−0.315	−0.463	−0.205	−0.517
SF-8 <sup>b</sup>	0.247	0.354	0.177	0.405
PCS	0.186	0.292	0.151	0.326
MCS	0.272	0.365	0.179	0.426
GADS <sup>a</sup>	−0.245	−0.415	−0.205	−0.451
GAS	−0.188	−0.358	−0.170	−0.374
GDS	−0.264	−0.407	−0.209	−0.457

**Notes:** All Pearson correlations  $P \leq 0.001$ ; recode reverse-coded items; The description of testing samples: <sup>a</sup>The merged data of nursing students and medical workers; <sup>b</sup>only sample medical workers; PSQ's Guttman's lambda-2: whole scale = 0.938; SF-8's Guttman's lambda-2: PCS = 0.815, MCS = 0.858, whole scale = 0.898; GADS's Guttman's lambda-2: GAS = 0.780, GDS = 0.789, whole scale = 0.870.

**Table 6** SCS-SF subscales intercorrelations and concurrent as well as convergent/divergent validity

Subscales	SCS-SF						
	SK	SJ	CH	IS	MI	OI	Overall score
SCS-SF <sup>a</sup>							
SK		0.025	0.366**	−0.005	0.496**	−0.084**	0.481**
SJ			−0.102**	0.291**	0.068**	0.256**	0.505**
CH				−0.039*	0.397**	−0.051**	0.425**
IS					0.141**	0.530**	0.640**
MI						0.126**	0.638**
OI							0.588**
PSQ index <sup>a</sup>	−0.165**	−0.205**	−0.091**	−0.417**	−0.271**	−0.392**	−0.483**
SF-8 <sup>b</sup>	0.676**	0.546**	0.620**	0.632**	0.629**	0.633**	0.692**
PCS	0.610**	0.460**	0.537**	0.533**	0.530**	0.531**	0.593**
MCS	0.650**	0.555**	0.616**	0.641**	0.640**	0.647**	0.694**
GADS <sup>a</sup>	−0.102**	−0.205**	−0.048*	−0.370**	−0.221**	−0.355**	−0.412**
GAS	−0.062**	−0.170**	−0.019	−0.320**	−0.178**	−0.307**	−0.336**
GDS	−0.127**	−0.209**	−0.069**	−0.363**	−0.230**	−0.349**	−0.423**

**Notes:** \*\* $P < 0.01$ , \* $P < 0.05$ ; The description of testing samples: <sup>a</sup>The merged data of nursing students and medical workers; <sup>b</sup>only sample medical workers.

correlation, 95% CI = 0.436–0.790) and 0.618 (ICC, 95% CI = 0.413–0.764), indicating fair agreement. However, the 12-item form has higher test–retest reliability. In brief, the 10-item form showed acceptable reliability, detailed results are given in Table 7. Given that the SCS-SF-10 (or/and the SCS-SF) is not unidimensional, the reliability coefficients for each subscale separately were provided (see Table 8).

## Levels of self-compassion, perceived stress and anxious and depressive symptoms

The self-compassion level difference was statistical significance ( $t=2.263$ ,  $P=0.024$ ) between medical workers and nursing students, as well as medical workers' self-compassion level ( $0.547 \pm 0.111$ ) was lower than nursing students'

**Table 7** Reliability of the 10 and 12 items form SCS-SF for whole scale in total sample

	Quality criteria	SCS-SF-10	SCS-SF
M ± SD <sup>a</sup>	n/a	0.552±0.110	0.561±0.096
Item-total correlation	0.20–0.80	0.220–0.441	0.046–0.452
Coefficient alpha (Cronbach's $\alpha$ )	0.70–0.95	0.686	0.639
Guttman's lambda-2 ( $\lambda_2$ )	0.70	0.706	0.672
Test–retest correlation <sup>b</sup>	>0.70	0.617 [0.436–0.790]	0.700 [0.558–0.825]
Intraclass correlation coefficient (ICC) for absolute agreement <sup>b</sup>	>0.70 or 0.75	0.618 [0.413–0.764]	0.694 [0.517–0.814]
Standard error of measurement <sup>c</sup>	n/a	0.068	0.053

**Notes:** n/a, not applicable. <sup>a</sup>These have used standardization for comparison, M ± SD = mean ± SD; <sup>b</sup>sample size for test–retest and ICC were 50 (from sample nursing students), of these, 95% CI estimation of test–retest correlation utilized bootstrap, i.e., all  $P < 0.001$ ; <sup>c</sup>standard error of measurement was calculated as  $SD \times \sqrt{1 - ICC}$ .

**Table 8** Reliability of the 10 and 12 items form SCS-SF for each subscale in total sample

	Coefficient alpha (Cronbach's $\alpha$ )	Guttman's lambda-2 ( $\lambda_2$ )
SCS-SF-10		
Factor 1	0.741	0.744
Factor 2	0.696	0.697
Factor 3	0.540	0.540
SCS-SF		
SK	0.291	0.291
SJ	0.540	0.540
CH	0.389	0.389
IS	0.570	0.570
MI	0.667	0.667
OI	0.500	0.500

self-compassion level ( $0.557 \pm 0.108$ ). In addition, independent  $t$ -test revealed that the differences were statistical significance between medical workers and nursing students on levels of perceived stress ( $t = -11.222$ ,  $P < 0.001$ ), anxious symptoms ( $t = -15.057$ ,  $P < 0.001$ ) and depressive symptoms ( $t = -13.805$ ,  $P < 0.001$ ). Medical workers' levels on stress and anxious and depressive symptoms were higher than nursing students' levels (all  $P < 0.001$ ), and yet the former lower than the latter on level of self-compassion ( $P = 0.024$ ). The results of SF-8 for nursing students were not specified because this study not collected the data. Of these, results of Levene's test for equality of variances suggested that the variance homogeneity of stress and depressive symptoms equal variances was not assumed. Additionally, effect sizes of the SCS-SF (10 items and 12 items form) and its subscales were all below 0.10, but effect sizes of the PSQ and the GADS (GAS and GDS) were all above 0.10 and below 0.30. To compare mean scores in different forms of the SCS, detailed results are shown in Table 9.

## Discussion

Regarding its psychometric properties, especially in structure validity, our findings could not replicate the proposed six-factor second-order model (a default model). It seemed to confirm earlier findings that there is still no accordance regarding the SCS's original factor structure.<sup>69</sup> This more comprehensive CFAs and EFAs revealed main indexes of different models concerning six- and three-factor (first- and second-order), indicating a good fit of the three-factor models with the survey data. The best-fitting solution was a three-factor second-order model with 10 items (i.e., the factor structure obtained from EFA was confirmed in CFA), and only at this point we gave our subscales labels, the name one positive factor (Factor 1, positively formulated items: C3, C5, C6, C7), two negative factors (Factor 2, negatively formulated items: C1, C4, C8, C9; Factor 3, negatively formulated items: C11, C12). However, a comparison of the three-factor solution obtained from EFA (Table 4) with the original subscales in the study by Raes et al did not support such labeling.<sup>12</sup> Namely, Factor 3 only consisted of two self-judgment items, while Factor 2 consisted of two over-identification and two isolation items, Factor 1 consisted of two mindfulness items, one common humanity item and one self-kindness item. In a real sense, this best three-factor model with 10 items still included the SCS's six-factor structure main information with at least one each original subscale's item, despite incomplete information. Our shorten version supported a three-factor structure like other shorten version in China, the differences were that their research used the SCS-LF and extracted clear three factors, unfortunately, removed more than half of items.<sup>19</sup> Notably, our best model formed one positive factor and two factors were a bit different from a model with one positive and one negative factor, because the latter was used 24 items (Dutch<sup>16</sup>) and 26 items (Portuguese<sup>17</sup>). To establish an identical set of factors across studies and national settings was a failure, since there were

**Table 9** Nursing students' and medical workers' mean scores in measure instruments

Instruments	Total sample	Nursing students	Medical workers	t	P	Cohen's d	Effect size
SCS-SF-10	0.552±0.110	0.557±0.108	0.547±0.111	2.263	<b>0.024</b>	0.091	0.046
Factor 1	0.687±0.141	0.692±0.135	0.681±0.148	1.931	0.054	0.078	0.069
Factor 2	0.420±0.168	0.421±0.168	0.419±0.168	0.301	0.764	0.012	0.006
Factor 3	0.548±0.186	0.559±0.184	0.536±0.189	3.155	<b>0.002</b>	0.123	0.062
SCS-SF	0.561±0.096	0.562±0.095	0.560±0.097	0.652	0.514	0.021	0.010
SK	0.678±0.148	0.685±0.143	0.670±0.152	2.644	<b>0.008</b>	0.102	0.051
SJ	0.548±0.186	0.559±0.184	0.536±0.189	3.155	<b>0.002</b>	0.123	0.062
CH	0.619±0.161	0.606±0.160	0.634±0.162	-4.454	<b>&lt;0.001</b>	-0.174	-0.087
IS	0.444±0.199	0.450±0.201	0.436±0.197	1.820	0.069	0.070	0.035
MI	0.683±0.165	0.683±0.157	0.683±0.174	0.066	0.948	0.003	0.002
OI	0.396±0.185	0.391±0.183	0.401±0.187	-1.417	0.157	-0.054	-0.027
PSQ index	0.430±0.156	0.399±0.138	0.466±0.168	-11.222	<b>&lt;0.001</b>	-0.436	-0.213
SF-8	–	–	65.255±17.097	–	–	–	–
PCS	–	–	67.145±17.745	–	–	–	–
MCS	–	–	63.364±18.924	–	–	–	–
GADS	9.346±4.714	8.081±4.349	10.850±4.691	-15.725	<b>&lt;0.001</b>	-0.612	-0.293
GAS	5.157±2.551	4.503±2.442	5.935±2.460	-15.057	<b>&lt;0.001</b>	-0.584	-0.280
GDS	4.189±2.561	3.577±2.343	4.915±2.620	-13.805	<b>&lt;0.001</b>	-0.538	-0.260

**Notes:** The SCS-SF's total score mean = (raw score – the lowest possible score)/(the highest possible score – the lowest possible score). In bold were statistically significant between key groups. "–" means not available.

cultural and linguistic diversity, sampling factors, translation of the items and/or other subjective and objective factors.<sup>70,71</sup>

For MI, high levels of MI in the first- and second-order three-factor model were supported across key groups (i.e., invariance on three levels of factor loadings, intercepts, and residual variances). The three-factor structure, especially in high-order model, could be generalized across nursing students and medical workers, which allows the comparison of the groups. Nonetheless, the six-factor structure (grudgingly) supported generalization across subgroups of participants owing to some substandard indexes.

Criterion validity of the measure was evaluated by examining the associations between different factors of the SCS-SF (including 10 and 12 items form) and stress, anxiety and depression. In terms of the criterion validity of the scale, results showed that one positive subscale of self-compassion was negatively associated with stress, anxiety and depression; had a positive link to PCS and MCS. Two negative subscale of self-compassion with the same kind of results was done, due to recoding reverse-coded items. Acceptable psychometric performance implies a Pearson correlation <0.60 (convergent) and a correlation <0.20<sup>72</sup>; convergent/divergent validity of the SCS-SF (including 10 and 12 items form) was not satisfactory, perhaps triggered by an imbalance on the number items per factor. It would mean that the scale

that is being validated is not specific enough to measure the construct of interest in a given population.<sup>73</sup>

With regard to reliability, Neff has repeatedly mentioned that the SCS-SF is not recommended for use in examining the six components separately because subscales have poor reliability<sup>3</sup>; this point also happened to coincide with the current study. There were acceptable reliability coefficients (near-threshold of 0.70) on this adaption. Although the reliability (Cronbach's  $\alpha$ ) was improved for the 10 items form, it is very close to cut-off value and still below that value (0.69). Reliability coefficients of Factors 1 and 2 were above minimum quality criteria. Factor 3 in these dimensions only consisted of two items that led to lower reliability coefficients. Good practice dictates a minimum of three items per factor.<sup>47</sup> It has unacceptable internal consistency for each subscale in six dimensions.

Additionally, key groups-based mean scores differences were also discovered. Level of self-compassion in medical workers was lower than level of self-compassion in nursing students. However, levels of adversities (i.e., stress, anxiety and depression) in medical workers were higher than level of self-compassion in nursing students. Unfortunately, this study did not collected the data of nursing students on quality of life, thereby not compared the level between key groups.

As for levels of self-compassion, perceived stress and anxious and depressive symptoms, a strong link was found between medical workers and nursing students. The findings may provide clues to describe the interaction between self-compassion and the other variables, such as adverse mental health outcomes. Unfortunately, effect sizes of the SCS-SF (10 items and 12 items form) and its subscales were all below 0.10 (i.e., a correlation coefficient of 0.10 is thought to represent a weak or small association<sup>74</sup>), indicating that effect sizes in trials may be statistically significant but not produce clinically important differences in practice settings. Still have, our results supported the use of both the total score and subscale scores of the SCS-SF and/or SCS-SF-10, but demonstrating that some subscales and the scale have lower and still acceptable reliability. The advantage of using the short form but also the disadvantages must be mentioned. The short form has a near-perfect correlation with the long form when examining total scores.<sup>12</sup> We also do not recommend using the short form if you are interested in subscale scores, since they are less reliable with the short form.<sup>3</sup> Despite the fact that the total score with either the long or short form can use for assessing of clients' self-compassion, the short form (SCS-SF) might be as a better, easy-to-use and economical alternative when the aim is to use the total score.

## Implications for practice

Admittedly, there is a surge of interest in the well-being of those same medical professionals, even including medical students or nursing students, and other health care providers who provide treatment.<sup>29</sup> The monitoring of key groups' well-being, especially in medical workers, should be put on the agenda. Cultivating self-compassion may not be easy, but it is no doubt a worthwhile, empowering and liberating way to live your life. Additionally, some interventions could also be proposed, the Mindful Self-Compassion (MSC) training/program: either using informal practices such as the Self-Compassion Break, or formal meditation practices such as Affectionate Breathing (retrieved from: <https://self-compassion.org/tips-for-practice/>).

## Limitations and future directions

The present study is aimed at showing that these two samples from the Chinese population validated, explored and cross-validated the psychometric properties of this adaptation, the Chinese version of the SCS-SF, including 10 and 12 items form. Nevertheless, the authors admit to some weaknesses of the study. First, use of a single self-

report questionnaire to assess self-compassion and its analogous concepts may be lacking. In this connection, it would be useful to further employ more integrated measuring, such as a combination of questionnaires, including mindfulness, self-esteem, self-criticism and/or self-concept. Second, the sample population may not be representative of the Chinese population since the study sample was confined to two areas from medical field, and this limit the generalizability of findings. Although measurement of stability of the instrument over time was used in our study; the sample size of 50 for test-retest reliability seemed to be small. Third, one could argue that not all aspects of validity were analyzed in this study; however, we tried our best to report our measured results and compare with similar studies. Additionally, given that it was an observational study, there existed inevitably various biases too. Only medical workers responded the SF-8 also was one of the perceived drawbacks. Since validation is an ongoing process,<sup>75</sup> further validation study on the Chinese SCS is distinctly needed. In future studies, revisions would be needed if this instrument was tested with large-sample as clinical practice and research settings.

## Conclusion

Despite the limitations of this study, the Chinese version of the SCS-SF was accepted as valid and reliable measures for assessing self-compassion of psychological research areas in Chinese-speaking populations. Briefly, self-compassion scores can be utilized in research to evaluate self-compassion level of some population (adults), such as college students and medical workers.

The short form of the SCS, as the authors who created it suggest, might be particularly useful in settings where time constraints make the use of the long form less feasible or advisable. However, these authors recommend using the full scale if information about subscales is crucial.<sup>12</sup> Further investigation on possible links between SCS-SF and SCS-LF in a large sampling survey is greatly needed but is beyond the scope of the present study. Future research directions should focus on other developmental stages (e.g., children, adolescents or elderly) and using more heterogeneous samples.

## Acknowledgments

We are indebted to Prof. Dr Kristin Neff from University of Texas at Austin for providing this instrument. Special thanks go to our friends: Dr Jingjing Li (holds a PhD, from Department of Behavioral Sciences and Health Education,

Rollins School of Public Health, Emory University), Dr Zhenkun Wang (PhD, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology), Yucong Ma (MTI, he studied at Southeast University-Monash University Joint Graduate School (Suzhou) at that time), Di Zhang (MSN, RN, Quality Control Department, Wuhan Asia General Hospital) and Yongyong Xi (MM, Department of Environment and Occupational Hazard Control, Center for Disease Control and Prevention of Pudong New District) for their valuable assistance at forward and back-translation within the process. The authors express their appreciation to all respondents taking part in the present study and some friends for offering support in collecting data. Again, we would like to thank the anonymous reviewers who offered constructive suggestions for improvement. This research was funded by the National Natural Science Foundation of China (Grant No. 81773552 and Grant No. 81273179) and the National Key Research and Development Program of China (Grant No. 2018YFC1315302, 2017YFC1200502). Additionally, this project was supported by Key Research Center for Humanities and Social Sciences in Hubei Province (Hubei University of Medicine) (Grant No. 2016YB06).

## Author contributions

Chuanhua Yu and Runtang Meng conceived and designed the project. Runtang Meng collected, analyzed and interpreted the data as well as compiled the initial draft of the manuscript. Runtang Meng, Yong Yu and Shouxia Chai jointly edited this manuscript. Xiangyu Luo, Shouxia Chai, Boxiong Gong, Bing Liu and Yi Luo supported in providing the data. Chuanhua Yu and Ying Hu advised on statistical analysis. Runtang Meng directed all facets of the study. Runtang Meng and Chuanhua Yu are responsible for the overall project. All authors contributed to data analysis, drafting or revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

## Disclosure

The authors declare no conflicts of interest in this work.

## References

- Neff KD. Self-compassion: an alternative conceptualization of a healthy attitude toward oneself. *Self Identity*. 2003;2(2):85–101. doi:10.1080/15298860309032
- Neff KD. The development and validation of a scale to measure self-compassion. *Self Identity*. 2003;2(3):223–250. doi:10.1080/15298860309027
- Neff KD. The self-compassion scale is a valid and theoretically coherent measure of self-compassion. *Mindfulness*. 2016;7(1):264–274. doi:10.1007/s12671-015-0479-3
- Neff KD, Dahm KA. Self-compassion: what it is, what it does, and how it relates to mindfulness. In: Ostafin BD, Robinson MD, Meier BP, editors. *Handbook of Mindfulness and Self-Regulation*. New York: Springer New York; 2015:121–137.
- MacBeth A, Gumley A. Exploring compassion: a meta-analysis of the association between self-compassion and psychopathology. *Clin Psychol Rev*. 2012;32(6):545–552. doi:10.1016/j.cpr.2012.06.003
- Muris P, Petrocchi N. Protection or vulnerability? A meta-analysis of the relations between the positive and negative components of self-compassion and psychopathology. *Clin Psychol Psychother*. 2017;24(2):373–383. doi:10.1002/cpp.2005
- Sirois FM, Kitner R, Hirsch JK. Self-compassion, affect, and health-promoting behaviors. *Health Psychol*. 2015;34(6):661. doi:10.1037/hea0000158
- Leary MR, Tate EB, Adams CE, Batts Allen A, Hancock J. Self-compassion and reactions to unpleasant self-relevant events: the implications of treating oneself kindly. *J Pers Soc Psychol*. 2007;92(5):887. doi:10.1037/0022-3514.92.5.887
- Zessin U, Dickhäuser O, Garbade S. The relationship between self-compassion and well-being: a meta-analysis. *Appl Psychol Health Well Being*. 2015;7(3):340–364. doi:10.1111/aphw.12051
- Sirois FM, Hirsch JK. Self-compassion and adherence in five medical samples: the role of stress. *Mindfulness*. 2019;10(1):46–54. doi:10.1007/s12671-018-0945-9
- Pérula-de Torres L-A, Atalaya JCV-M, García-Campayo J, et al. Controlled clinical trial comparing the effectiveness of a mindfulness and self-compassion 4-session programme versus an 8-session programme to reduce work stress and burnout in family and community medicine physicians and nurses: MINDUDD study protocol. *BMC Fam Pract*. 2019;20(1):24. doi:10.1186/s12875-019-0913-z
- Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the self-compassion scale. *Clin Psychol Psychother*. 2011;18(3):250–255. doi:10.1002/cpp.702
- Neff KD, Tóth-Király I, Yarnell LM, et al. Examining the factor structure of the self-compassion scale in 20 diverse samples: support for use of a total score and six subscale scores. *Psychol Assess*. 2019;31(1):27. doi:10.1037/pas0000629
- Neff KD, Whittaker TA, Karl A. Examining the factor structure of the self-compassion scale in four distinct populations: is the use of a total scale score justified? *J Pers Assess*. 2017;99(6):596–607. doi:10.1080/00223891.2016.1269334
- Tóth-Király I, Bóthe B, Orosz G. Exploratory structural equation modeling analysis of the Self-Compassion Scale. *Mindfulness*. 2017;8(4):881–892. doi:10.1007/s12671-016-0662-1
- López A, Sanderman R, Smink A, et al. A reconsideration of the self-compassion scale's total score: self-compassion versus self-criticism. *PLoS One*. 2015;10(7):e0132940. doi:10.1371/journal.pone.0132940
- Costa J, Marôco J, Pinto-Gouveia J, Ferreira C, Castilho P. 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. *Clin Psychol Psychother*. 2016;23(5):460–468. doi:10.1002/cpp.1974
- Zeng X, Wei J, Oei TP, Liu X. The self-compassion scale is not validated in a Buddhist sample. *J Relig Health*. 2016;55(6):1996–2009. doi:10.1007/s10943-016-0205-z
- Gong H, Jia H, Guo T, Zou L. The revision of self-compassion scale and its reliability and validity in adolescents (in Chinese). *Psychol Res*. 2014;7(1):36–40.

20. Deniz M, Ş K, Sümer AS. The validity and reliability of the Turkish version of the self-compassion scale. *Soc Behav Pers.* 2008;36(9):1151–1160. doi:10.2224/sbp.2008.36.9.1151
21. Mantzios M, Wilson JC, Giannou K. Psychometric properties of the Greek versions of the self-compassion and mindful attention and awareness scales. *Mindfulness.* 2015;6(1):123–132. doi:10.1007/s12671-013-0237-3
22. Sutton E, Schonert-Reichl KA, Wu AD, Lawlor MS. Evaluating the reliability and validity of the self-compassion scale short form adapted for children ages 8–12. *Child Indic Res.* 2018;11(4):1217–1236. doi:10.1007/s12187-017-9470-y
23. Marshall SL, Parker PD, Ciarrochi J, Sahdra B, Jackson CJ, Heaven PC. Self-compassion protects against the negative effects of low self-esteem: a longitudinal study in a large adolescent sample. *Pers Individ Dif.* 2015;74:116–121. doi:10.1016/j.paid.2014.09.013
24. Muris P. A protective factor against mental health problems in youths? A critical note on the assessment of self-compassion. *J Child Fam Stud.* 2016;25(5):1461–1465. doi:10.1007/s10826-015-0315-3
25. García-Campayo J, Navarro-Gil M, Andrés E, Montero-Marín J, López-Artal L, Demarzo MMP. Validation of the Spanish versions of the long (26 items) and short (12 items) forms of the Self-Compassion Scale (SCS). *Health Qual Life Outcomes.* 2014;12(1):4. doi:10.1186/1477-7525-12-4
26. Castilho P, Pinto-Gouveia J, Duarte J. Evaluating the multifactor structure of the long and short versions of the self-compassion scale in a clinical sample. *J Clin Psychol.* 2015;71(9):856–870. doi:10.1002/jclp.22187
27. Uršič N, Kocjančič D, Žvelc G. Psychometric properties of the slovenian long and short version of the self-compassion scale. *Psihologija.* 2018;OnlineFirst:1–19.
28. Chen J, Liangshi Yan ZL. Reliability and validity of Chinese version of Self-compassion Scale (in Chinese). *Chin J Clin Psychol.* 2011;19(6):734–736.
29. Raab K. Mindfulness, self-compassion, and empathy among health care professionals: a review of the literature. *J Health Care Chaplain.* 2014;20(3):95–108. doi:10.1080/08854726.2014.913876
30. Cheung T, Wong SY, Wong KY, et al. Depression, anxiety and symptoms of stress among baccalaureate nursing students in Hong Kong: a cross-sectional study. *Int J Environ Res Public Health.* 2016;13(8):779. doi:10.3390/ijerph13121252
31. Zeng Y, Wang G, Xie C, Hu X, Reinhardt JD. Prevalence and correlates of depression, anxiety and symptoms of stress in vocational college nursing students from Sichuan, China: a cross-sectional study. *Psychology, Health & Medicine.* 2019;24(7):798–811. doi: 10.1080/13548506.2019.1574358
32. Pizzutti LT, Carissimi A, Valdivia LJ, et al. Evaluation of Breathworks' Mindfulness for Stress 8-week course: effects on depressive symptoms, psychiatric symptoms, affects, self-compassion, and mindfulness facets in Brazilian health professionals. *J Clin Psychol.* 2019;75(6):970–984. doi:10.1002/jclp.22749
33. Efsthathiou G. Translation, adaptation and validation process of research instruments. In: Suhonen R, Stolt M, Papastavrou E, editors. *Individualized Care: Theory, Measurement, Research and Practice.* Cham: Springer International Publishing; 2019:65–78.
34. Sousa VD, Rojjanasiriat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract.* 2011;17(2):268–274. doi:10.1111/j.1365-2753.2010.01434.x
35. Levenstein S, Prantera C, Varvo V, et al. Development of the perceived stress questionnaire: a new tool for psychosomatic research. *J Psychosom Res.* 1993;37(1):19–32.
36. Luo Y, Gong B, Meng R, et al. Validation and application of the Chinese version of the Perceived Stress Questionnaire (C-PSQ) in nursing students. *PeerJ.* 2018;6:e4503. doi:10.7717/peerj.4503
37. Turner-Bowker DM, Bayliss MS, Ware JE, Kosinski M. Usefulness of the SF-8™ health survey for comparing the impact of migraine and other conditions. *Qual Life Res.* 2003;12(8):1003–1012.
38. Wang S, Luan R, Lei Y, Kuang C, He C, Chen Y. Development and evaluation of Chinese version of Short Form-8 (in Chinese). *Mod Prev Med.* 2007;34(6):1022–1023, 1026.
39. Lang L, Zhang L, Zhang P, Li Q, Bian J, Guo Y. Evaluating the reliability and validity of SF-8 with a large representative sample of urban Chinese. *Health Qual Life Outcomes.* 2018;16(1):55. doi:10.1186/s12955-018-0880-4
40. Ware JE, Kosinski M, Dewey JE, Gandek B. *How to Score and Interpret Single-item Health Status Measures: a Manual for Users of the SF-8 Health Survey (with a Supplement on the SF-6 Health Survey).* Lincoln, RI; Boston, MA: QualityMetric Inc.; Health Assessment Lab; 2001.
41. Goldberg D, Bridges K, Duncan-Jones P, Grayson D. Detecting anxiety and depression in general medical settings. *Bmj.* 1988;297(6653):897–899. doi:10.1136/bmj.297.6653.897
42. Luo Y, Meng R, Li J, Liu B, Cao X, Ge W. Self-compassion may reduce anxiety and depression in nursing students: a pathway through perceived stress. *Public Health.* 2019;174:1–10. doi:10.1016/j.puhe.2019.05.015
43. Association WM. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects World Medical Association Declaration of HelsinkiSpecial Communication. *JAMA.* 2013;310(20):2191–2194. doi:10.1001/jama.2013.281053
44. Meredith W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika.* 1993;58(4):525–543. doi:10.1007/BF02294825
45. Comrey AL, Lee HB. *A First Course in Factor Analysis.* New York, NY: Psychology Press; 2013.
46. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychol Methods.* 1999;4(1):84. doi:10.1037/1082-989X.4.1.84
47. Hair JF, Black WC, Babin BJ, Anderson RE. *Multivariate Data Analysis: Pearson New International Edition.* 7th ed. London: Pearson Higher Education; 2014.
48. Kline RB. *Principles and Practice of Structural Equation Modeling.* 4th ed. New York (NY): Guilford publications; 2016.
49. Marsh HW, Hau KT, Grayson D. Goodness of Fit in Structural Equation Models. In: A. Maydeu-Olivares & J. J. McArdle, editors. *Multivariate applications book series. Contemporary psychometrics: A festschrift for Roderick P. McDonald.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2005:275–340.
50. Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika.* 1973;38(1):1–10. doi:10.1007/BF02291170
51. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull.* 1990;107(2):238. doi:10.1037/0033-2909.107.2.238
52. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling.* 1999;6(1):1–55. doi:10.1080/10705519909540118
53. Joreskog KG, Sorbom D. *LISREL 8: User's Reference Guide.* Chicago: Scientific Software International; 1996.
54. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Modeling.* 2002;9(2):233–255. doi:10.1207/S15328007SEM0902\_5
55. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Modeling.* 2007;14(3):464–504. doi:10.1080/10705510701301834
56. Meade AW, Johnson EC, Braddy PW. Power and sensitivity of alternative fit indices in tests of measurement invariance. *J Appl Psychol.* 2008;93(3):568. doi:10.1037/0021-9010.93.3.568

57. Brannick MT. Critical comments on applying covariance structure modeling. *J Organ Behav.* **1995**;16(3):201–213. doi:10.1002/(ISSN)1099-1379
58. Kelloway EK. Structural equation modelling in perspective. *J Organ Behav.* **1995**;16(3):215–224. doi:10.1002/(ISSN)1099-1379
59. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods.* **1999**;4(3):272. doi:10.1037/1082-989X.4.3.272
60. Everitt BS, Skrondal A. *The Cambridge Dictionary of Statistics.* 4th ed. New York (NY): Cambridge University Press; **2010**.
61. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: a Practical Guide to Their Development and Use.* New York (NY): Oxford University Press, USA; **2015**.
62. Stevens JP. *Applied Multivariate Statistics for the Social Sciences.* 5th ed. New York: Routledge; **2012**.
63. Streiner DL, Norman GR. “Precision” and “accuracy”: two terms that are neither. *J Clin Epidemiol.* **2006**;59(4):327–330. doi:10.1016/j.jclinepi.2005.09.005
64. Sijtsma K, Emons WH. Advice on total-score reliability issues in psychosomatic measurement. *J Psychosom Res.* **2011**;70(6):565–572. doi:10.1016/j.jpsychores.2010.11.002
65. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* **2016**;15(2):155–163. doi:10.1016/j.jcm.2016.02.012
66. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* **1996**;1(1):30. doi:10.1037/1082-989X.1.1.30
67. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res.* **2005**;19(1):231–240.
68. Tabachnick BG, Fidell LS. *Using Multivariate Statistics.* Boston: Pearson Education; **2013**.
69. Muris P, Otgaar H, Petrocchi N. Protection as the mirror image of psychopathology: further critical notes on the self-compassion scale. *Mindfulness.* **2016**;7(3):787–790. doi:10.1007/s12671-016-0509-9
70. Rönnlund M, Vestergren P, Stenling A, Nilsson LG, Bergdahl M, Bergdahl J. Dimensionality of stress experiences: factorial structure of the Perceived Stress Questionnaire (PSQ) in a population-based Swedish sample. *Scand J Psychol.* **2015**;56(5):592–598. doi:10.1111/sjop.12214
71. Davidov E, Meuleman B, Cieciuch J, Schmidt P, Billiet J. Measurement equivalence in cross-national research. *Annu Rev Sociol.* **2014**;40.
72. Luján-Tangarife J, Cardona-Arias J. Construction and validation of measurement scales in health: a review of psychometric properties. *Arch Med.* **2015**;11(3):1–10.
73. Carvajal A, Centeno C, Watson R, Martínez M, Rubiales AS. How is an instrument for measuring health to be validated? *An Sist Sanit Navar.* **2011**;34(1):63–72.
74. Durlak JA. How to select, calculate, and interpret effect sizes. *J Pediatr Psychol.* **2009**;34(9):917–928. doi:10.1093/jpepsy/jsp004
75. Hubley AM, Zumbo BD. Validity and the consequences of test interpretation and use. *Soc Indic Res.* **2011**;103(2):219. doi:10.1007/s11205-011-9843-4

## Psychology Research and Behavior Management

Dovepress

### Publish your work in this journal

Psychology Research and Behavior Management is an international, peer-reviewed, open access journal focusing on the science of psychology and its application in behavior management to develop improved outcomes in the clinical, educational, sports and business arenas. Specific topics covered in the journal include: Neuroscience, memory and decision making; Behavior modification and management; Clinical

applications; Business and sports performance management; Social and developmental studies; Animal studies. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/psychology-research-and-behavior-management-journal>