Assessment
1–13

DOI: 10.1177/1073191116673173
asm.sagepub.com
SAGE

A Brief Stress Diagnostic Tool: The Short Stress Overload Scale

James H. Amirkhan[1]

Abstract

The Stress Overload Scale (SOS) has demonstrated validity in predicting pathological stress reactions; however, at 30 items, it is lengthy for some clinical applications. Here, two studies tested a 10-item SOS–Short (SOS-S). First, the SOS-S was compared with the SOS in a longitudinal community study (n = 391), using indices of pathology as criterion measures. Results showed the SOS-S to be equivalent to the SOS in reliability and concurrent and predictive validity, although not quite as sensitive to somatic symptoms. Second, the SOS-S was compared to the 10-item Perceived Stress Scale in a cross-sectional community study (n = 249), in which symptoms and response biases were also assessed. Results showed both measures to be susceptible to biasing, and the SOS-S to demonstrate superior validity when biases were controlled. The SOS-S appears a viable alternative to the SOS and the 10-item Perceived Stress Scale for assessing stress, and risk for sequelae, across a broad demographic spectrum.

Keywords

stress measure, stress overload, validity, SOS-S, SOS, PSS-10

The Stress Overload Scale (SOS) is a self-report measure that has demonstrated accuracy in identifying people who will develop stress-related pathology in the wake of stressful experiences (Amirkhan, 2012; Amirkhan, Urizar, & Clark, 2015). However, at 30 items, the SOS may be too cumbersome for some research and practice applications. The present studies investigate the validity of a 10-item SOS–Short (SOS-S), both relative to the full SOS and the popular 10-item Perceived Stress Scale (PSS-10; Cohen, Kamarck, & Mermelstein, 1983). The aim was to find a viable diagnostic tool for stress and stress-related disorders across diverse populations, with consideration paid to practicality and respondent burden as well as psychometric strength.

Stress and Illness

The link between stress and illness is complex, in that not every stressful experience results in pathology for all people (Salleh, 2008). This is, in fact, as predicted by stress theories: An early model by Selye (1956) proposed that stress responses are triggered by "adaptational demands," large or small, which perturb homeostasis. Such demands were said to produce feelings of distress but not necessarily dysfunction. If there are adequate resources to counter demands, then homeostasis is reestablished; but with intense or incessant demands, resistance may be overwhelmed, and cellular damage—even death—can ensue. More recent models emphasize different systems but retain the same basic mechanism. For example, McEwen (2000) focuses on the allostatic rather than homeostatic system but maintains the idea that excessive "allostatic load" can overwhelm and dysregulate immunological responses. Others have turned to psychological systems but agree that when demands are appraised as exceeding resources (Lazarus & Folkman, 1984), or triggering excessive resource loss (Hobfoll, 1989), mental or physical illness can result.

In short, theories agree that for stress to become destructive, two conditions must be met: (1) exposure to demanding events, coupled with (2) inadequate resources to meet those demands (Cohen, Kessler, & Gordon, 1995). This pathogenic state has sometimes been called "stress overload" (Amirkhan, 2012; Lunney, 2006) to differentiate it from more transient and less pernicious forms of stress.

Stress Measurement

Stress measures have been repeatedly criticized for their lack of fidelity to stress theory (e.g., Lazarus, 1990). And this is reflected in the failure of most scales to assess both components of stress overload. Some measures, such as life

[1]California State University Long Beach, CA, USA

Corresponding Author:
James H. Amirkhan, Psychology Department, California State University Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840, USA.
Email: james.amirkhan@csulb.edu

event checklists (e.g., Brantley, Jones, & Boudreaux, 1997; Holmes & Rahe, 1967) assess demands but overlook resistive resources. A person faced with many demands, but who has adequate resources to counter them, may be more correctly described as "challenged" (Lazarus & Folkman, 1984) than "stressed." Other measures focus on resources, such as resilience (Windle, Bennett, & Noyes, 2011) or hardiness (Gebhardt, van der Doef, & Paul, 2001) but ignore demand load. A person with greatly depleted resources, but faced with no or few demands, would be more accurately deemed "fragile" (Amirkhan et al., 2015) than "stressed." Feeling challenged or fragile may not be pleasant, but these are states theoretically less likely to yield pathology than true stress overload (Lazarus & Folkman, 1984; Lunney, 2006).

The SOS was designed expressly to capture the overload state (Amirkhan, 2012). In accordance with stress theories, its items were chosen to capture feelings of being overwhelmed. Factor analysis of these items yielded two subscales, which correspond to theory in reflecting both impinging demands (Event Load) and depleted resources (Personal Vulnerability). This two-factor-scale structure differentiates the SOS from other stress measures, as does the fact that it is the only one to have been wholly empirically derived over a sequence of studies using prescribed psychometric procedures (Cronbach & Meehl, 1955; Loevinger, 1957). Its structure also affords a categorical scoring option that is unique among stress measures: By crossing the two scales, a diagnostic matrix can be formed for classifying respondents according to their risk for pathology. Finally, because data from diverse community samples were used in its construction, the SOS is exceptionally well suited to a broad spectrum of people and problems.

Likely due to these attributes, considerable evidence for the validity of the SOS has accrued. It has been shown to predict symptoms and illnesses following real-world stressful life events (Amirkhan, 2012; Amirkhan et al., 2015). It has detected aberrant cortisol responses to laboratory stressors (Amirkhan et al., 2015). And it has accurately differentiated stressed from nonstressed populations (Amirkhan et al., 2015). Nevertheless, the length of the SOS might dissuade researchers and practitioners from its use. At 30 items, it could be too costly to incorporate into large-scale studies, and too burdensome to impose on victims of trauma. For this reason, a prototype short SOS-S is tested here.

Current Studies

The SOS-S was constructed by selecting the strongest of the SOS items. That is, items were first chosen according to the strength of their loadings as markers of the theory-consistent factors underlying the SOS (as per Loevinger, 1957). Then, from among these, items were chosen according to their psychometric strength, in terms of the reliability and validity evidenced on the SOS (as per Cronbach & Meehl, 1955). Finally, from among these, items were chosen according to their demonstrated comprehensibility across a diversity of SOS respondents (as per the requisites of a general-population measure). In recognition of the popularity of the PSS-10 (Lee, 2012), items were limited to the 10 that best satisfied these conditions.

While specifics of this procedure are given later, its general aim was to avoid the "sins" typical of short-form development (Smith, McCarthy, & Anderson, 2000). The choice of a well-validated parent measure and the selection of items to fully represent its content domain were steps in this direction. But the greatest of the sins is assuming that the psychometric properties of the parent measure will be automatically transferred to the short-form, so that further validity testing is unnecessary. The impetus of the present research was, in fact, to demonstrate the criterion validity of the SOS-S, vis-à-vis both its parent measure and a popular peer measure. In Study 1, the ability of the SOS-S to identify current, and predict future, signs of pathology was examined in relation to that of the full SOS. In Study 2, the accuracy of the SOS-S in detecting concurrent pathology was retested but this time relative to the PSS-10. In addition, response biases were evaluated as threats to the validity of both short-format scales.

In this process, other, more specific, sins (Smith et al., 2000) were also addressed. The convergence of the SOS-S and the SOS was tested, both by direct correlation and verification of a shared factor structure. The psychometric properties of SOS-S factor scales were examined, to determine if the reliability and validity of the full scales were retained. Finally, the diagnosticity of the SOS-S was evaluated, to determine if it could differentiate people according to their risk for pathology as accurately as the other measures.

Study 1: SOS-S Versus SOS

To investigate the validity of the SOS-S, methods used in validating the original SOS were re-employed (Amirkhan et al., 2015). This included the use of illness as a criterion (here, as reflected in symptoms and behaviors indicative of psychiatric or somatic disorders). Also, a longitudinal design was utilized to control common-methods effects (Podsakoff, MacKenzie, & Podsakoff, 2012) and to determine the reliability of the criterion measures (a crucial prerequisite of validity tests; Aiken, 2000). Finally, sampling strategies that captured the diversity of the community, in terms of both demographics and stress levels, were again used.

Method

Participants. Two community sites were used to recruit participants. Informed consent was obtained from 440 recruits, but 391 (89%) completed Wave 1 study measures in sufficient detail for analysis. Of these, only 156 (40%) also completed Wave 2 measures 1 week later.

Measures. The full SOS (Amirkhan, 2012) measures perceptions of stress overload occurring in the past week. It consists of 12 Event Load items (which assess perceived demands; e.g., "felt swamped by your responsibilities"), 12 Personal Vulnerability items (which assess perceived debility; e.g., "felt like you couldn't cope"), and six filler items (e.g., "felt generous"). The filler items, along with an innocuous title that does not mention stress ("SOS, A Measure of Day-to-Day Feelings"), are features designed to counter negativity and social desirability response biases. Each item is paired with a 5-point response scale, anchored at the extremes (1 = *Not at all*, 5 = *A lot*). There are also seven demographic items, presented on the last page to avoid priming effects (Steele, 1997).

SOS Event Load (EL) and Personal Vulnerability (PV) scales were derived from an oblique factor solution, and therefore correlate. The scales have shown good internal consistency (α = .94), test–retest stability (r = .75 over 1 week) and validity (Amirkhan, 2012; Amirkhan et al., 2015). Scale scores are summed to yield a single, continuous score (ranging from 24 to 120). Or they may be split at their means and crossed to form a 2 × 2 matrix, which yields categorical scores: "High Risk" (high EL, high PV), "Low Risk" (low EL, low PV), "Challenged" (high EL, low PV), or "Fragile" (low EL, high PV). Categorical scores are not completely redundant of continuous scores: It is possible for a person with a middle-range total score to be assigned to either "High Risk" or another category, depending on the position of their scale scores relative to the scale means. Categorical scores have demonstrated excellent sensitivity (96%) and specificity (100%) in identifying stressed and symptomatic persons in a community sample (Amirkhan et al., 2015).

From the full SOS, 10 items were extracted to form the short SOS-S. In the pursuit of brevity, filler items were dropped. However, other features of the SOS were retained: The format (with identical instructions and prompts), the response scales, the demographic items, and the ambiguous title. And, like the full version, SOS-S scales may be combined into a continuous total score (ranging from 10 to 50), or split at their means to yield four categorical scores.

To construct the SOS-S, findings from eight previous psychometric studies of the SOS (in Amirkhan, 2012, and Amirkhan et al., 2015) were revisited. Each individual SOS item was evaluated using multiple criteria: (1) high loadings on one of the two underlying factors (EL or PV) but not the other; (2) construct validity, in terms of significant convergence with measures of similar constructs (e.g., Daily Hassles for EL items; Hardiness for PV items); (3) test–retest reliability of r > .70 after 1 week; (4) criterion validity, in terms of significant correlations with symptom measures; and (5) cultural sensitivity, with no respondent questions or criticisms in any community study. The five items that best met these criteria for each the EL and the PV scales were chosen to comprise the 10-item SOS-S (see the appendix).

Two indicators of mental and physical illness were employed as criterion measures. A Symptoms checklist that had been constructed for a prior validity test of the SOS (Amirkhan et al., 2015) was reemployed here, for reasons of precedent and because it had proven reliable (α = .91, test–retest r = .85). It was made more exhaustive than most available measures by amassing stress-related symptoms from a variety of sources, including global health measures (e.g., the Cohen-Hoberman Inventory of Physical Symptoms; Cohen & Hoberman, 1983) and internet sites (e.g., "Stress Symptoms," 2013). In this process, care was taken to avoid any that might overlap with SOS items (e.g., "strained"). Its final 35 items describe specific somatic complaints, ranging from "bad breath" to "vomiting"; its response scales, anchored at 1 = *Not at all* and 5 = *A lot*, indicate the extent to which each was experienced in the prior week. Possible Symptoms totals range from 35 to 175.

A Behaviors checklist was created here to parallel the Symptoms measure. Also exhaustive, it assesses behavioral responses indicative of stress-related disorders, gleaned from the literature (e.g., Krueger & Chang, 2008) and Internet sites (e.g., "Behavioural Stress Symptoms," 2009). The measure lists 35 specific acts, such as "cancelled appointments" and "lost temper," but none that overlap with SOS items (e.g., "couldn't cope"). Respondents indicate which occurred in the past week by means of 5-point response scales identical to those on the Symptoms scale. Possible Behaviors scores range from 35 to 175.

Procedure

Two community sites, which had yielded a gamut of demographic profiles and stress levels in a previous study (Amirkhan et al., 2015), were revisited to obtain the present sample. A county courthouse is a venue peopled by a diverse public, yet one that likely over-represents stressed people. Here, recruitment took place early on weekday mornings, as participants (n = 197) reported for trial or jury duty. A city aquarium also draws a diverse crowd but in general a more relaxed one. Here, recruitment occurred at midday on weekends, as participants (n = 194) arrived for the opening of the tourist attraction.

A convenience sampling strategy was used. Persons who agreed to participate and who met the selection criteria (over 18 and English-literate) filled out Informed Consent and Contact Information forms before receiving the Wave 1 survey packet. This packet contained *either* the full SOS or the short SOS-S, and both the Symptoms and Behaviors measures, in counterbalanced orders. Participants completed the packet on site, using only an assigned code as an

Table 1. Demographic Composition of Study Samples.

	Study 1			
	Wave 1	Wave 2	Study 2	U.S. Census
Sample type	Community		Community	Community
Size (*n*)	391	155	249	
Gender, *n* (%)				
Male	151 (55%)	81 (53%)	123 (49%)	49%
Female	140 (45%)	73 (47%)	119 (48%)	51%
Age (years), *n* (%)				
18-24	59 (20%)	13 (8%)	72 (29%)	16%
25-34	76 (26%)	37 (24%)	74 (29%)	20%
35-49	74 (26%)	38 (25%)	49 (20%)	29%
50-65	69 (24%)	59 (36%)	38 (15%)	23%
>65	12 (4%)	10 (7%)	7 (3%)	12%
Ethnicity, *n* (%)				
African American	31 (9%)	13 (9%)	26 (10%)	9%
Asian American	31 (9%)	17 (11%)	17 (7%)	14%
Hispanic American	92 (26%)	34 (22%)	81 (32%)	
Caucasian	163 (45%)	75 (49%)	97 (39%)	50%
Other/mixed	42 (11%)	14 (9%)	19 (8%)	9%
Education, *n* (%)				
High school or less	102 (28%)	28 (18%)	58 (23%)	44%
Some college	142 (40%)	65 (42%)	107 (43%)	22%
College degree	66 (18%)	33 (21%)	47 (19%)	19%
Advanced degree	49 (14%)	28 (18%)	27 (11%)	10%
Income ($) (household)				
<25,000	108 (31%)	31 (23%)	74 (29%)	23%
25,000-39,000	64 (18%)	22 (15%)	42 (17%)	14%
40,000-59,000	47 (14%)	26 (18%)	24 (10%)	15%
60,000-99,000	74 (21%)	33 (23%)	43 (17%)	22%
100,000-149,000	56 (16%)	31 (21%)	31 (12%)	14%
>150,000	0 (0%)	0 (0%)	17 (7%)	12%

Note. Percentages exclude missing values. Census numbers are from 2010 Census and 2013 American Community Survey for persons over 18 years old in Los Angeles County. "Asian American" includes Pacific Islanders; "Hispanic American" is not a distinct category in Census data.

identifier, sealed it into an envelope, and dropped it into a locked box to ensure confidentiality.

At this point, participants received the Wave 2 packet, marked with a matching code and the due date for its return. This packet included the short SOS-S, the Symptoms and Behaviors checklists, a prepared return envelope, and a $1 State Lottery incentive. Instructed to wait 1 week before responding, and to avoid the use of identifiers, participants sealed their completed measures into the provided envelopes for return by post. All participants received reminder e-mails or phone calls prior to the due date, and some received a second reminder if their packet had not been received 3 days following the due date.

Results

Sample Characteristics. The sampling strategy was effective in procuring a diverse Wave 1 sample in terms of gender, age, ethnicity, income, and level of education (see Table 1). Sample proportions generally approximated U.S. Census figures for the region, except for deviations in education, $\chi^2(3) = 9.35$, $p = .025$, and income, $\chi^2(5) = 13.87$, $p = .016$. Specifically, the very least educated and the very wealthiest brackets were underrepresented. The Wave 2 sample was similar, despite the considerable attrition between waves. It differed from the Wave 1 sample only in terms of age, $\chi^2(4) = 17.35$, $p = .002$, with older participants more likely to return the follow-up surveys.

Scale Characteristics. To test whether the SOS-S replicated the factor structure of the full SOS, an exploratory factor analysis was conducted. Wave 1 data were used because only this sample met the recommended minimum size for this analysis (Comrey, 1973). Extraction by means of principal axis factoring extraction revealed two primary factors, each having eigenvalues >1, and together explaining 76%

Table 2. Descriptive Statistics and Correlations for Study 1 Measures.

Variable	*M*	*SD*	Range	α	Full SOS, Wave 1	SOS-S, Wave 1	SOS-S, Wave 2
Demographics							
Age	39.09	14.92	18-84		−.23*	−.29*	−.09
Gender					.11	.20	.10
Education					−.02	−.12	.07
Income					−.29*	−.28*	−.22*
Pathology (Wave 1)							
Symptoms	51.49	17.50	35-140	.94	.60**	.50**	
Behaviors	60.70	25.04	35-169	.96	.81**	.78**	
Pathology (Wave 2)							
Symptoms	49.27	17.10	35-137	.94	.77**	.48**	.61**
Behaviors	54.94	23.42	35-166	.96	.71**	.66**	.71**
Full SOS (Wave 1)							
Personal Vulnerability	24.36	12.39	12-60	.91	.95**		.76**
Event Load	29.25	13.57	12-60	.94	.96**		.77**
Total	53.56	24.85	24-120	.96			.81**
Short SOS-S (Wave 1)							
Personal Vulnerability	9.12	5.79	5-25	.91		.94**	.75**
Event Load	11.28	6.38	5-25	.91		.95**	.72**
Total	20.51	11.60	10-50	.94			.75**
Short SOS-S (Wave 2)							
Personal Vulnerability	8.50	5.31	5-25	.92			.94**
Event Load	10.92	6.34	5-25	.94			.96**
Total	19.44	11.08	10-50	.95			

Note. SOS-S = Short Stress Overload Scale. Zero-order correlations are shown. Higher "Gender" scores indicate female.
*$p < .01$. **$p < .001$.

of the variance (subsequent factors explained 5% or less). Rotation by means of Oblimin to an oblique solution showed these factors to indeed correspond to Event Load and Personal Vulnerability (see the appendix). The interfactor correlation ($r = .62$) was greater than that reported for the full scale ($r = .48$; Amirkhan, 2012).

Evidence for the reliability of the SOS-S was found. As shown in Table 2, internal consistency—of the measure as a whole and its subscales—was good, and approximated levels previously reported for the full SOS (Amirkhan et al., 2015). Across the 1-week span between Waves, the SOS-S ($r = .75$) and its subscales (PV, $r = .71$; EL, $r = .73$) demonstrated adequate test–retest stability, on par with values for the SOS (Amirkhan, 2012). In addition, both the SOS-S and SOS demonstrated good variability of response, as evidenced by wide ranges, mid-range means, and large standard deviations (see Table 2). There were no ceiling or basement effects to compromise correlational analyses.

The criterion measures demonstrated the requisite reliability for validity tests. As may be seen in Table 2, internal consistency for the Symptoms and Behaviors checklists was good at both Waves 1 and 2. In addition, test–retest stability for the measures was adequate, despite the fact that they assess transitory phenomena ($r = .80$, and $r = .75$, respectively). Both criterion measures also exhibited good variability of response.

Possible Confounds. Correlational analyses were used for the primary validity tests, with a conservative $p = .01$ to adjust for multiple tests. First, zero-order correlations between the demographic items and the stress-overload measures were calculated. These revealed a potential problem: Income was significantly associated with SOS and SOS-S scores (see Table 2), and also with the criterion measures of Symptoms ($r = -.21$ at Wave 1, $r = -.33$ at Wave 2, $ps < .001$) and Behaviors ($r = -.23$ at Wave 1, $r = -.29$ at Wave 2, $ps < .001$). Therefore, Income constituted a possible third-variable confound in validity tests with the illness criteria.

Validity Tests. Zero-order correlations showed the short SOS-S to relate strongly to the full SOS ($r = .81$, $p < .001$, 95% confidence interval [CI .75, .86]). The strength of this association is noteworthy given the week-long interval between scores. That is, owing to the study design, this figure represents the correlation between the Wave 1 SOS and the Wave 2 SOS-S.

Partial correlations, controlling for Income, were used to test the associations between the stress-overload measures and the criterion indices. These showed the Wave 1

Table 3. Means of Criterion Variables Predicted by Short SOS-S Versus Full SOS Categories in Study 1.

	Short SOS-S Categories (Wave 1)				Full SOS Categories (Wave 1)			
			Personal Vulnerability				Personal Vulnerability	
			Low	High			Low	High
Pathology (Wave 1)								
Symptoms	Event Load	Low	46.25_{a}	47.38_{a}	Event Load	Low	43.57_{a}	51.59_{b}
		High	50.95_{b}	65.79_{c}		High	52.56_{b}	63.53_{c}
Behaviors	Event Load	Low	48.99_{a}	58.88_{b}	Event Load	Low	45.30_{a}	65.56_{b}
		High	64.50_{b}	90.13_{c}		High	58.41_{b}	84.42_{c}
Pathology (Wave 2)								
Symptoms	Event Load	Low	42.86_{a}	48.50_{b}	Event Load	Low	42.92_{a}	56.00_{b}
		High	49.00_{b}	60.71_{c}		High	57.38_{b}	73.86_{c}
Behaviors	Event Load	Low	44.58_{a}	64.50_{b}	Event Load	Low	47.08_{a}	69.00_{b}
		High	56.73_{b}	77.35_{c}		High	67.88_{b}	80.00_{c}

Note. SOS-S = Short Stress Overload Scale. Within each factorial, means with different subscripts differ at least at the $p < .05$ level.

SOS-S to relate to concurrent Symptoms (partial $r = .40$, $p < .001$, 95% CI [.31, .48]) and to subsequent Symptoms (partial $r = .68$, $p < .001$, 95% CI [.59, .76]). While these relationships were significant, they were lower that the corresponding values found between the full SOS and Wave 1 Symptoms (partial $r = .59$, $p < .001$, 95% CI [.52, .65]) and Wave 2 Symptoms (partial $r = .76$, $p < .001$, 95% CI [.69, .82]). A test for comparing the magnitude of coefficients (McNemar, 1975) indeed showed the SOS-S to be a poorer predictor of Symptoms than the SOS at both Wave 1, $t(150) = 4.56$, $p < .001$, and Wave 2, $t(150) = 6.15$, $p < .001$.

Partial correlations also showed the Wave 1 SOS-S to predict both concomitant (partial $r = .75$, $p < .001$, 95% CI [.70, .79]) and future Behaviors (partial $r = .68$, $p < .001$, 95% CI [.59, .76]). These coefficients were commensurate with those found between the SOS and Wave 1 Behaviors (partial $r = .79$, $p < .001$, 95% CI [.75, .82]) and Wave 2 Behaviors (partial $r = .68$, $p < .001$, 95% CI [.59-.76]). In sum, correlational analyses indicated that the SOS-S paralleled the full SOS in its ability to predict signs of illness, albeit not as well in regard to symptoms.

Diagnosticity Tests. Categorical scores from the SOS-S and SOS were also compared. According to stress theories, persons in the High Risk category (high PV, high EL) should exhibit more illness than those in the Low Risk (low PV, low EL) or even those in the Fragile (high PV, low EL) or Challenged (low PV, high EL) categories. Using Wave 1 data, 2 × 2 categorical matrices were constructed for each the SOS-S and the SOS. These showed that the SOS-S categorized 24% of its respondents as High Risk and 62% as Low Risk, while the SOS classified 30% and 53% of its respondents accordingly. A chi-square test indicated no significant difference in the proportions assigned to the four categories, $\chi^2(3) = 3.23$, $p = .357$.

The matrices were then compared for their accuracy in differentiating unhealthy from healthy people. Owing to the unequal sizes of the categories, generalized linear model (GLM) analysis of variance procedures were used, with each Symptoms and Behaviors at each Wave 1 and Wave 2 as dependent variables. Because only four tests were conducted, and the direction of the findings was anticipated, no adjustments were made to the p value for significance. In regard to the SOS-S, significant main effects for PV and EL, but not their interaction, were found in every analysis; this underscores the equivalent (i.e., parallel) importance of both scales in discriminating the ill from the well. As may be seen in Table 3, mean differences among the four cells were all in the direction predicted by theory. Post hoc contrasts showed High Risk means to be greater than Low Risk, Fragile, or Challenged means for both Symptoms and Behaviors at both Waves 1 and 2 (all ps $< .001$). This pattern of results mirrored precisely that found for the full SOS: Main, but no interaction effects, with the High Risk means exceeding those for other categories in every analysis (all ps $< .05$). In sum, categorical scores derived from the SOS-S and the full SOS worked equally well in identifying people who exhibited signs of ill health, as well as those who would later exhibit such signs.

Discussion

Present findings provided considerable evidence to show that the sins typical of short-form development (Smith et al., 2000) had been avoided. First, convergence of the short and full versions of the SOS was verified: Not only did the SOS-S correlate strongly with the SOS (despite the fact that the measures were administered a week apart), but

it also evidenced the same underlying factor structure that reflected the same theoretical constructs. Second, the SOS-S and its component scales were found to approximate the reliability of the SOS, both in terms of internal consistency and test–retest stability. Finally, and most important, the SOS-S and its scales demonstrated criterion validity, at a level comparable to that of the full SOS. However, there were indications that the SOS-S was not quite as sensitive in detecting somatic symptoms.

The validity analyses permitted evaluation of the SOS-S' diagnosticity. Concurrent validity tests showed that its categorical scores were able to distinguish sicker from healthier persons at the time of the assessment. Predictive validity tests showed that the categories were able to isolate those who would exhibit future signs of illness. In both regards, the accuracy of SOS-S categories paralleled those of the full SOS.

The use of pathology as a criterion for validating stress measures is a widespread practice, but one that is problematic for a number of reasons (Amirkhan et al., 2015). Corrective steps were taken here, but these may not have been wholly adequate. First, to avoid spurious correlations, items on the criterion measures were carefully selected to be dissimilar to stress-overload items. But creating these measures meant that, although face valid and reliable in the current data, they were largely of unknown psychometric strength. To confirm present findings, it would be necessary to repeat tests with an established and well-validated indicator of physical and psychiatric symptomatology. Second, to minimize criterion contamination, a 1-week buffer period was imposed between assessments. However, this period may not have been long enough to mitigate more trait-like response biases, such as social desirability or negative affectivity. To insure that current validity correlations were not artifactual, it would be prudent to assess and statistically control these biases in subsequent tests. Third, to obtain a diverse sample, recruitment took place in the community. Still, participants were all volunteers, and differed somewhat from general population demographics. To insure representativeness in future, the use of more sophisticated sampling techniques would be advisable.

Overall, the first study showed the SOS-S to closely approximate the diagnostic abilities of the SOS, despite being one-third its length. However, the study's limitations, as well as a desire to compare the SOS-S to a peer measure, prompted a second investigation.

Study 2: SOS-S Versus PSS-10

This study matched the SOS-S against the most widely used of brief stress measures (Cole, 1999), the 10-item Perceived Stress Scale (PSS-10; Cohen et al., 1983). Validation strategies similar to those of the first study were re-employed, to adhere to recommendations for avoiding the sins of short-form development (Smith et al., 2000). However, the methodology was somewhat adjusted to address new "sins" emerging from the prior study: First, a published and validated symptoms questionnaire was selected as the criterion measure. Second, the response biases most likely to contaminate stress scales were measured and used as statistical controls in analyses. And third, sampling strategies were altered to improve representativeness.

Method

Participants. Of 260 community residents who consented to participate, 249 (96%) returned surveys complete enough for analysis.

Measures. The 10-item SOS-S, as previously described, was used. In the prior study, it was seen that this short version demonstrated internal ($\alpha = .94$) and test–retest reliability ($r = .75$, over one week), and validity in terms of significant covariation with the parent SOS ($r = .81$).

The PSS-10 (Cohen et al., 1983) assesses the degree to which people perceived their life as stressful in the prior month. It consists of four negative items (e.g., "felt nervous and stressed") and six that are reverse-keyed (e.g., "felt you were on top of things"). A review of psychometric studies of the PSS concluded that the 10-item version is superior to both the full 14-item and the abbreviated 4-item versions (Lee, 2012). Those studies indicated good internal consistency (mean $\alpha = .83$), test–retest reliability (mean $r = .79$, over 1 to 2 weeks), and validity (correlations with physical health, affective disorders, and life events). Item response scales for PSS-10 range from 0 = *Never* to 4 = *Very often*, yielding continuous total scores from 0 to 40.

The PSS-10 does not provide guidelines for classifying stressed versus nonstressed respondents, so a categorical scoring scheme was improvised here. Recent confirmatory factor analysis of the measure (Taylor, 2015) revealed a two-oblique factor structure like that of the SOS-S. However, the factors, identified as perceived Helplessness and Self-Efficacy, do not map onto the constructs identified in stress theories. Moreover, the measure's authors have dismissed differentiation of these factors as "irrelevant" (Cohen & Williamson, 1988, p. 43), and scores based on all 10 items have proven to be valid (Taylor, 2015). For these reasons, it was decided not to construct a 2 × 2 categorical matrix similar to that of SOS-S. Instead, a simple median split of PSS-10 scores was employed to form "High Stress" and "Low Stress" categories. While such methods have been criticized (Preacher, Rucker, MacCallum, & Nicewander, 2005), it was assumed to be the one most likely used by researchers and clinicians.

As the criterion measure of pathology, the 15-item Patient Health Questionnaire (PHQ-15: Kroenke, Spitzer, Williams, & Lowe, 2010) was chosen for a number

of reasons. First, it was "neutral," having not been used in previous validation studies of the SOS or PSS; second, there was no overlap between its items and those on either stress scale; and third, there is considerable evidence of its psychometric strength. In normative samples of 6000 patients (Kroenke, Spitzer, & Williams, 2002), it demonstrated good internal consistency (α = .80), and validity (correlations with indices of functionality, healthcare utilization, disability days, and difficulty in activities or relationships). The PHQ-15 assesses somatic symptoms (e.g., "headaches") that are common among patients suffering either physical or mental disorders. Respondents rate the severity of these complaints by means of scales ranging from 0 = *Not bothered at all* to 2 = *Bothered a lot*. Possible PHQ-15 scores range from 0 to 30.

Negative affectivity (NA) is a disposition of particular concern for stress measures: It can affect responses to both stress and symptom scales, artificially inflating the correlation between them (Watson & Pennebaker, 1989). A Type D Personality test (Denollet, 2005), which assesses negative affectivity and social inhibition as conjoint predictors of cardiovascular morbidity, provided the NA scale used here. It consists of seven items (e.g., "I am often in a bad mood") that are rated on a 5-point response scale ranging from 0 = *False* to 4 = *True*. It was selected because it has demonstrated good internal consistency (α = .88), test–retest reliability (r = .76, over a 3-month interval), and convergent validity with a neuroticism scale. Possible NA scores range from 0 to 28.

Social desirability (SD) is a response bias of concern for sensitive questions (King & Bruner, 2000), such as those found on stress scales. However, there has been debate about the nature of SD and the extent of its impact. Research has shown that SD is not a unitary construct, but an amalgam of two independent factors, Impression Management and Positive Self-Deception (Tracey, 2015). Moreover, it has been argued that the likelihood of self-report biasing is small, and the widespread use of bias indicators is unwarranted (McGrath, Mitchell, Kim, & Hough, 2010; Tracey, 2015). In regard to stress measures, however, differentiating the components of SD seems irrelevant: Whether one is trying to hide weakness from others, or deny distress to themselves, the effect on scores would be the same. And in fact, there is evidence to show that SD does affect mood measures in general (Soubelet & Salthouse, 2011), and stress scales specifically (Wiechman, Smith, Smoll, & Placek, 2000). The Marlow-Crowne Social Desirability Scale is a standard for assessing SD, and it taps both SD components (Tracey, 2015). However, because the full scale is lengthy, a 13-item short form (Reynolds, 1982) was chosen for this study. Unlike other short versions, this one has shown acceptable reliability (K-R_{20} = .76) and validity (r = .93 with the full scale). Here, items (e.g., "I always admit it when I make a mistake") were paired with 5-point response scales to match those used with the NA items. Possible SD scores ranged from 0 to 52.

Procedure. The sample was drawn from same two community sites that provided a good range of stress and demographic characteristics in the first study. As before, the courthouse (n = 124) and aquarium (n = 125) samples were of comparable size. However, this time a quota-sampling strategy was employed, in which demographic profiles guided recruitment efforts. These profiles (e.g., "Female, African-American, Elderly, Wealthy") were created to insure a match to U.S. Census proportions for the region. Nevertheless, there was room for error in the recruiters' judgments; moreover, if no one appeared to match a given profile within a 15-minute period, recruiters were instructed to move on to the next profile. Persons who did fit a profile, who satisfied the age and literacy selection criteria, and who provided consent, were handed survey packets with clipboards and pencils. The packet contained all five measures, in counterbalanced orders, a large envelope, and a $1 lottery ticket as an incentive. With detailed instructions to protect privacy and insure anonymity, participants completed the surveys on site, sealed them into the envelopes, and dropped them into locked collection boxes.

Results

Sample Characteristics. The community sites again yielded a diverse sample, with demographic proportions generally matching Census figures for the region (see Table 1). However, despite refinements to the sampling strategy, there were discrepancies in terms of age, $\chi^2(4) = 14.07$, $p = .007$, and education, $\chi^2(3) = 13.41$, $p = .004$, with older and less educated segments of the population underrepresented. On the other hand, there was no significant difference in income, and persons in the higher brackets were better represented this time.

Scale Characteristics. As shown in Table 4, the internal reliability of the study measures ranged from good (SOS-S, NA, PHQ-15) to adequate (PSS-10, SD). It may also be seen that all measures elicited good variability of response, with no evidence of ceiling or basement effects.

Confirmatory Factor Analysis was used to verify the two-oblique-factor structure of the SOS-S. Although a chi-square test suggested a poor fit to the data, $\chi^2(34) = 55.59$, $p < .01$, this is typical with sample sizes over 200 (Moss, 2016). In contrast, comparative-fit indices indicated a good fit for the model: Goodness of fit index (GFI) = .956, comparative fit index (CFI) = .982, root mean square error of approximation (RMSEA) = .053. Moreover, it was a much better fit than either a one-factor model, $\chi^2(35) = 165.52$, $p < .0001$, GFI = .844, CFI = .904, RMSEA = .126, or a

Table 4. Descriptive Statistics and Correlations for Study 2 Measures.

Variable	*M*	*SD*	Range	α	PSS-10	SOS-S
Demographics						
Age	34.72	14.09	18-76		−.08	−.07
Gender					.13	−.07
Education					−.13	−.18
Income					−.29**	−.21*
Response biases						
Negative Affectivity	10.35	6.85	0-28	.90	.50**	.49**
Social Desirability	29.90	10.70	1-47	.70	−.20*	−.19*
Symptoms						
PHQ-15	7.75	5.68	0-28	.89	.32**	.57**
Perceived Stress Scale-10						
Total	16.12	6.59	1-35	.76		.45**
Short SOS-S						
Personal Vulnerability	10.77	4.97	5-25	.87	.49**	.89**
Event Load	14.52	5.21	5-25	.87	.36**	.90**
Total	25.29	9.06	10-50	.91	.45**	

Note. SOS-S = Short Stress Overload Scale; PSS-10 = 10-item Perceived Stress Scale. Zero-order correlations are shown. Higher "Gender" scores indicate female.
*$p < .01$. **$p < .001$.

two-orthogonal-factor model, $\chi^2(35) = 267.72$, $p < .0001$, GFI = .856, CFI = .829, RMSEA = .169.

Possible Confounds. For correlational validity analyses, a conservative $p = .01$ threshold was used to compensate for the number of tests. Zero-order correlations showed Income to covary with the SOS-S and PSS-10 (see Table 4) as well as the PHQ-15 symptom scale ($r = -.225$, $p < .001$), again making it a potential confound in correlations among these variables.

Response Biases. Both the NA and SD response bias scales correlated with the SOS-S and PSS-10 stress scales (Table 4). Moreover, NA scores ($r = .43$, $p < .001$, 95% CI [.32, .53]) and SD scores ($r = -.18$, $p = .004$, 95% CI [−.29, .06]) also correlated with PHQ-15 symptoms. In conceptual terms, these findings indicate that stress and symptom levels were likely exaggerated by the more negative participants, and downplayed by the participants desiring to appear normal. In psychometric terms, the findings indicate that correlations among the study measures were likely inflated, owing to an artificial consistency in scores imposed by the biases.

Validity Tests. Zero-order correlations provided multiple indications of the validity of both stress measures (see Table 4). The SOS-S and the PSS-10 intercorrelated, indicating their construct validity. And each correlated significantly with the PHQ-15 criterion, indicating their concurrent validity.

To procure more accurate estimates of the strength of association among the measures, partial correlations were used to control for possible Income, NA and SD confounds. These yielded coefficients of considerably lower magnitude than those shown in Table 4. The SOS-S and the PSS-10 were still significantly correlated (partial $r = .26$, $p < .001$, 95% CI [.14, .37]). And the SOS-S was still related to PHQ-15 symptoms (partial $r = .45$, $p < .001$, 95% CI [.35, .54]). But the PSS-10 was no longer significantly associated with the PHQ-15 (partial $r = .12$, $p = .061$, 95% CI [.00, .24]). A test for differences in the magnitude of correlations (McNemar, 1975) confirmed that the SOS-S was a stronger predictor of symptoms than the PSS-10, $t(241) = 2.56$, $p = .02$. In short, reanalysis preserved evidence for the criterion validity of the SOS-S but not the PSS-10.

Diagnosticity Tests. Risk categories derived from the SOS-S and the PSS-10 were compared in terms of their ability to identify symptomatic participants. For the SOS-S, the four-category diagnostic matrix was again formed using mean splits of its PV and EL subscales. For the PSS-10, the median score ($md = 16$, which approximated the mean, $M = 16.12$) was used to divide respondents into High Stress and Low Stress categories.

Owing to their differing sizes, a GLM procedure was used to conduct two analyses of variance on the SOS-S and PSS-10 categories, with PHQ-15 symptoms as the dependent variable. Because the number of tests was limited, and the direction of results anticipated, no adjustments were made to p values. In regard to the SOS-S, previous findings were replicated: Significant main effects were found for both PV, $F(1, 245) = 23.19$, $p < .001$, and EL, $F(1, 245) = 15.58$, $p < .001$, but not their interaction. Post hoc comparisons of SOS-S risk categories again showed the High Risk

(M = 11.68) group to report the most symptoms, significantly more than the Low Risk (M = 5.11), Fragile (M = 7.30), or Challenged (M = 6.65) groups (all ps < .01). With regard to the PSS-10, there was a significant difference between the High Stress (M = 9.09) and the Low Stress (M = 6.44) groups, $F(1, 247) = 14.31$, $p < .001$.

However, these results were possibly compromised by response biases, which had been found associated with both stress and symptoms scores. By impacting these scores, the biases might have caused some participants to be incorrectly categorized, and even the correctly categorized participants to give inaccurate symptom reports. To determine if results would be the same had there been no biasing, GLM was used to conduct two analysis of covariance analyses with NA and SD as covariates. For the SOS-S, results did not change: There were still significant main effects for PV, $F(1, 242) = 11.91$, $p < .001$, and EL, $F(1, 242) = 11.64$, $p < .001$, and still no significant interaction. Examining the estimated means, the High Risk group (est. M = 10.76) still differed from the Low Risk (est. M = 5.64), the Fragile (est. M = 7.24), and the Challenged groups (est. M = 6.93), and these differences were all significant (all ps < .05). But for the PSS-10, reanalysis changed the results: The difference between the High Stress group (est. M = 8.06) and the Low Stress group (est. M = 7.38) was in the right direction, but no longer significant, $F(1,244) = 0.86$, $p = .355$. In short, after correcting for response biases, SOS-S categories still differentiated symptomatic from healthy people, but the ersatz PSS-10 categories did not.

Discussion

Addressing the methodological flaws specified in the literature (Smith et al., 2000) as well as those introduced in the first study, this investigation still provided evidence for the viability of a short-form SOS. That is, prior findings regarding the SOS-S' internal reliability and criterion validity were replicated. Yet, this study also pointed out a weakness in the SOS-S, its vulnerability to social desirability and negative affectivity biases. In contrast, the full SOS had proven unaffected by such response tendencies (Amirkhan, 2012). It may be that the vulnerability of the shorter version is due to its loss of the filler items built into the full scale.

Initial findings for the PSS-10 paralleled those for the SOS-S. It, too, demonstrated significant associations with symptoms, and it, too, proved vulnerable to response biases. But the picture changed once the response biases were statistically controlled. Continuous PSS-10 scores were no longer correlated with symptom counts, while SOS-S scores were. In terms of diagnosticity, categorical PSS-10 scores no longer differentiated people according to symptom levels, while SOS-S categories did. The latter comparison, however, may be unfair. Like most stress measures, the PSS-10 does not provide a rubric for forming risk categories, and the one improvised here may not have been the best.

Other methodological problems should also be noted. Because the longitudinal aspect of the first study was dropped, there was no assessment of subsequent illness; thus, whether the SOS-S or the PSS-10 is a better predictor of stress sequelae is unknown. And, despite the efforts taken, the present sample was still not completely representative of the general population.

General Discussion

Adhering to recommendations for short-form development (Smith et al., 2000), two studies examined the viability of a 10-item version of the Stress Overload Scale. Results showed the new SOS-S to mirror its parent measure in several ways: It duplicated the factor structure, justifying two identical subscales. It achieved comparable reliability, with good internal consistency and adequate score stability, as a whole and for each subscale. Most important, it demonstrated validity, both in its own right and relative to its parent and a popular peer measure. SOS-S continuous scores proved capable of identifying signs of pathology, nearly as well as the SOS and better than the PSS-10. Its categorical scores demonstrated diagnostic accuracy in differentiating symptomatic from healthy people, as well as the SOS and better than the PSS-10. In short, SOS-S appears a viable alternative to both the full SOS and the popular PSS-10 for use in clinical research and practice, and a good choice where there are concerns about time, cost, or respondent fatigue.

However, there are caveats: First, the SOS-S was less sensitive to somatic symptoms than the full SOS. In applications seeking to capture a wide range of stress sequelae, such as studies of the effects of urban stressors (Shmool et al., 2015), the full SOS would be the better choice. Second, the SOS-S was tested on samples drawn from public spaces, so persons with debilitating physical or psychological conditions were not represented. The accuracy of the SOS-S in identifying signs of stress in patient populations, such as those suffering chronic illnesses (Zautra, 1996), is as yet untested. Third, unlike the full SOS (Amirkhan, 2012), the SOS-S was found vulnerable to response biases. This was attributed to the absence of filler items, a protective feature built into the full scale and lost in the short one. But, the PSS-10 proved equally vulnerable to social desirability and negative affectivity biases, a finding in accord with prior warnings in the literature (Cole, 1999; Gitchel, Roessler, & Turner, 2011). It may be that scores on *all* short stress measures are subject to such biasing, simply because those scores are based on fewer items. This does not completely negate the utility of the SOS-S, which still predicted symptoms after the most pernicious biases had been statistically controlled. But it does indicate that it would not be a good choice for precisely weighing the

contribution of stress to the etiology of a specific disorder, such as addictive behavior (Bergevin, Gupta, Derevensky, & Kaufman, 2006). The PSS-10 might not be the best choice, either, for after controlling response biases, it no longer related to symptoms here. In fairness, the PSS-10 has been tested on thousands of people, accruing considerable evidence for its association with pathology (Cohen & Williamson, 1988). Current findings were based on only a few hundred people, and should be taken as no more than a cautionary tale: The possibility of specious correlations due to biased responding should be considered in the choice of any brief stress measure.

Limitations to the present studies have been noted. Some were common to both studies, such as difficulties in obtaining a wholly representative sample. But other weaknesses were endemic to one study and offset by strengths in the other, such as longitudinal versus cross-sectional designs, exhaustive versus validated criterion measures, and unmeasured versus controlled response biases. And across these methodological variations, evidence for the validity of the SOS-S was consistent. Also of note is that Income emerged across studies as a covariate of SOS-S scores.

Income, as well as Social Desirability, was treated as a nuisance variable in the present analyses. However, their associations with the SOS-S suggest directions for substantive inquiry. In regard to Income, there is a large literature showing disproportionately high levels of stress and illness among the poor (Santiago, Wadsworth, & Stump, 2011). But there is debate regarding the causes–whether poverty causes dysfunction or whether the dysfunctional drift downwards economically (Hudson, 2005). The SOS-S, by virtue of measuring both the environmental and personal components of pathogenic stress, might be a useful tool in this debate. In regard to Social Desirability, there is speculation that it is a viable personality trait, one associated with healthy functioning (Tracey, 2015). And evidence shows it to predict differential physiological reactions to stressful stimuli (Tomaka, Blascovich, & Kelsey, 1992). Its association with the SOS-S might therefore have positive implications, suggesting that the scale could be as useful in detecting health-prone personalities as it is in identifying people at high risk for disease.

Other possible research applications include epidemiological investigations. One such study used a life events measure in a nationwide survey of 34,653 people to see if stress played a role in affective disorders (McLaughlin, Conron, Koenen, & Gilman, 2010). The SOS-S might have been a better choice, both because it is dedicated to measuring the entirety of pathogenic stress (debility as well as demands), and could have done so at a greatly reduced cost (10 items vs. 25 items). Another application could be program evaluation research. One community-based program taught skills for effective coping to ethnically diverse, low-income couples (Wadsworth et al., 2010). The SOS-S would be an apt choice for gauging the efficacy of such interventions, both because of its broad demographic fit and the minimal burden it imposes on already distressed people.

With regard to practice, the SOS-S would be an aid to diagnosis, owing to its unique categorical scoring scheme. For example, it could be helpful in identifying acute stress disorders, thereby improving the accuracy of predicting subsequent PTSD (Bryant, Friedman, Spiegel, Ursano, & Strain, 2011). It could also be useful in emergency situations. Following a fireworks accident that affected more than 2,000 people, differing levels of distress were noted among the survivors (Smid et al., 2012). The SOS-S, by virtue of its ability to accurately assign risk levels and thereby better target psychological services, might prove a valuable triage tool in such disasters.

Appendix

SOS-S item	Item type	Factor loading[a]	
IN THE PAST WEEK, have you felt:			
1. . . . inadequate?	Personal Vulnerability	**.690**	.045
2. . . . swamped by your responsibilities?	Event Load	−.037	**.898**
3. . . . that the odds were against you?	Personal Vulnerability	**.746**	.152
4. . . . that there wasn't enough time to get to everything?	Event Load	−.027	**.868**
5. . . . like nothing was going right?	Personal Vulnerability	**.976**	−.086
6. . . . like you were rushed?	Event Load	.085	**.821**
7. . . . like there was no escape?	Personal Vulnerability	**.648**	.263
8. . . . like things kept piling up?	Event Load	−.021	**.866**
9. . . . like just giving up?	Personal Vulnerability	**.823**	−.045
10. . . . like you were carrying a heavy load?	Event Load	.177	**.726**

Note. Each item is paired with a 5-point response scale anchored at *Not At All* and *A Lot*.
[a]Following Principal Axis Factoring extraction and Oblimin rotation, using SOS-S responses from Study 1, Wave 1 (n = 391).

Acknowledgment

The author expresses gratitude to his research team for its hard work in conducting these studies.

Author's Note

For the formatted SOS-S or inquiries about this research, please contact the author.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially supported by a California State University Long Beach RSCA Award.

References

Aiken, L. R. (2000). *Psychological testing and assessment*. Boston, MA: Allyn & Bacon.

Amirkhan, J. (2012). Stress overload: A new approach to the assessment of stress. *American Journal of Community Psychology*, *49*, 55-71. doi:10.1007/s10464-011-9438-x

Amirkhan, J., Urizar, G., & Clark, S. (2015). Criterion validation of a stress measure: The Stress Overload Scale. *Psychological Assessment*, *27*, 985-996. doi:10.1037/pas0000081

Behavioural Stress Symptoms. (2009). Retrieved from http://www.stresskey.com/stress-symptoms/behavioural.html

Bergevin, T., Gupta, R., Derevensky, J., & Kaufman, F. (2006). Adolescent gambling: Understanding the role of stress and coping. *Journal of Gambling Studies*, *22*, 195-208. doi:10.1007/s10899-006-9010-z

Brantley, P., Jones, G., & Boudreaux, E. (1997). Weekly Stress Inventory. In C. Zalaquett & R. Wood (Eds.), *Evaluating stress: A book of resources* (pp. 405-420). Lanham, MD: Scarecrow.

Bryant, R., Friedman, M., Spiegel, D., Ursano, R., & Strain, J. (2011). A review of acute stress disorder in DSM-5. *Depression and Anxiety*, *28*, 802-817. doi:10.1002/da.20737

Cohen, S., & Hoberman, H. (1983). Positive events and social supports as buffers of life change stress. *Journal of Applied Social Psychology*, *13*, 99-125. doi:10.13072midss.462

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*, 385-396. doi:10.2307/2136404

Cohen, S., Kessler, R., & Gordon, L. (1995). Strategies for measuring stress in studies of psychiatric and physical disorders. In S. Cohen, R. Kessler & L. Gordon (Eds.), *Measuring stress* (pp. 148-171). New York, NY: Oxford University Press.

Cohen, S., & Williamson, G. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health* (pp. 31-67). Newbury Park, CA: Sage.

Cole, S. (1999). Assessment of differential item functioning in the Perceived Stress Scale-10. *Journal of Epidemiology and Community Health*, *52*, 319-320. doi:10.1136/jech.53.5.319

Comrey, A. (1973). *A first course in factor analysis*. New York, NY: Academic Press.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302. doi:10.1037/h0040957

Denollet, J. (2005). DS14: Standard assessment of negative affectivity, social inhibition and Type D personality. *Psychosomatic Medicine*, *67*, 89-97. doi: 10.1097/01.psy.0000149256.81953.49

Gebhardt, W., van der Doef, M., & Paul, L. (2001). The Revised Health Hardiness Inventory (RHHI-24): Psychometric properties and relationship with self-reported health and health behavior in two Dutch samples. *Health Education Research*, *16*, 579-592. doi:10.1093/her/16.5.579

Gitchel, W., Roessler, R., & Turner, R. (2011). Gender effect according to item directionality on the Perceived Stress Scale for adults with multiple sclerosis. *Rehabilitation Counseling Bulletin*, *55*, 20-28. doi:10.1177/0034355211404567

Hobfoll, S. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, *44*, 513-524. doi:10.1037/0003-066X.44.3.513

Holmes, T., & Rahe, R. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research*, *11*, 213-218. doi:10.1016/0022-3999(67)90010-4

Hudson, C. (2005). Socioeconomic status and mental illness: Tests of the social causation and selection hypotheses. *American Journal of Orthopsychiatry*, *75*, 3-18. doi:10.1037/0002-9432.75.1.3

King, M., & Bruner, G. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology and Marketing*, *17*, 79-103. doi:10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0

Kroenke, K., Spitzer, R., & Williams, J. (2002). The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic Medicine*, *64*, 258-266. PMID: 11914441

Kroenke, K., Spitzer, R., Williams, J., & Lowe, B. (2010). The Patient Health Questionnaire Somatic, Anxiety and Depressive Symptom Scales: A systematic review. *General Hospital Psychiatry*, *32*, 345-359. doi:10.1016/j.genhosppsych.2010.03.006

Krueger, P., & Chang, V. (2008). Being poor and coping with stress: Health behaviors and the risk of death. *American Journal of Public Health*, *98*, 889-896. doi:10.2105/AJPH.2007.114454

Lazarus, R. (1990). Theory-based stress measurement. *Psychological Inquiry*, *1*, 3-13. doi:10.1207/s15327965pli0101_1

Lazarus, R., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York, NY: Springer.

Lee, E. (2012). Review of the psychometric properties of the Perceived Stress Scale. *Asian Nursing Research*, *6*, 121-127. doi:10.1016/j.anr.2012.08.004

Lunney, M. (2006). Stress overload: A new diagnosis. *International Journal of Nursing Terminologies and Classifications*, *17*, 165-175. doi:10.1111/j.1744-618X.2006.00035.x

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635-694. doi:10.2466/PR0.3.7.635-694

McEwen, B. (2000). The neurobiology of stress: From serendipity to clinical relevance. *Brain Research*, *886*, 172-189. doi:10.1016/S0006-8993(00)02950-4

McGrath, R., Mitchell, M., Kim, B., & Hough, L. (2010). Evidence for a response bias as a source of error variance in applied assessment. *Psychological Bulletin*, *136*, 450-470. doi:10.1037/a0019216

McLaughlin, K., Conron, K., Koenen, K., & Gilman, S. (2010). Childhood adversity, adult stressful life events, and risk of past-year psychiatric disorder: A test of the stress sensitization hypothesis in a population-based sample of adults. *Psychological Medicine*, *40*, 1647-1658. doi:10.1017/S0033291709992121

McNemar, Q. (1975). *Psychological statistics* (5th ed.). New York, NY: Wiley.

Moss, S. (2016, June 27). *Fit indices for structural equation modeling*. Retrieved from http://www.sicotests.com/psyarticle.asp?id=277

Podsakoff, P., MacKenzie, S., & Podsakoff, N. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539-569. doi:10.1146/annurev-psych-120710-100452

Preacher, K., Rucker, D., MacCallum, R., & Nicewander, W. (2005). Use of extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods*, *10*, 178-192. doi:10.1037/1082-989X.10.2.178

Reynolds, W. (1982). Development of reliable and valid short forms of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *38*, 119-125. doi:10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I

Salleh, M. R. (2008). Life event, stress and illness. *Malaysian Journal of Medical Sciences*, *15*, 9-18. PMID: 22589633

Santiago, C., Wadsworth, M., & Stump, J. (2011). Socioeconomic status, neighborhood disadvantage, and poverty-related stress: Prospective effects on psychological syndromes among diverse low-income families. *Journal of Economic Psychology*, *32*, 218-230. doi:10.1016/j.joep.2009.10.008

Shmool, J., Yonas, M., Newman, O., Kubzansky, L., Joseph, E., Parks, A., . . . Clougherty, J. (2015). Identifying perceived neighborhood stressors across diverse communities in New York City. *American Journal of Community Psychology*, *56*, 145-155. doi:10.1007/s10464-015-9736-9

Soubelet, A., & Salthouse, T. (2011). Influence of social desirability on age differences in self-reports of mood and personality. *Journal of Personality*, *79*, 741-762. doi:10.1111/j.1467-6494.2011.00700.x

Selye, H. (1956). *The stress of life*. New York, NY: McGraw-Hill.

Smid, G., van der Velden, P., Lensvelt-Mulders, G., Knipsheer, J., Gersons, B., & Kleber, R. (2012). Stress sensitization following a disaster: A prospective study. *Psychological Medicine*, *42*, 1675-1686. doi:10.1017/S0033291711002765

Smith, G., McCarthy, D., & Anderson, K. (2000). On the sins of short-form development. *Psychological Assessment*, *12*, 102-111. doi:10.1037//1040-3590.12.1.102

Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613-629. doi:10.1037//0003-066X.52.6.613

Stress Symptoms. (2013). Retrieved from http://www.mayoclinic.org/healthy-living/stress-management/in-depth/stress-symptoms/art-20050987

Taylor, J. (2015). Psychometric analysis of the ten-item Perceived Stress Scale. *Psychological Assessment*, *27*, 90-101. doi:10.1037/a0038100

Tomaka, J., Blascovich, J., & Kelsey, R. (1992). Effects of self-deception, social desirability, and repressive coping on psychophysiological reactivity to stress. *Personality and Social Psychology Bulletin*, *18*, 616-624. doi:10.1177/0146167292185012

Tracey, T. (2015). A note on socially desirable responding. *Journal of Counseling Psychology*, *63*, 224-232. doi:10.1037/cou0000135

Wadsworth, M., Santiago, C., Einhorn, L., Etter, E., Rienks, S., & Markman, H. (2010). Preliminary efficacy of an intervention to reduce psychosocial stress and improve coping in low-income families. *American Journal of Community Psychology*, *48*, 257-271. doi:10.1007/s10464-010-9384-z

Watson, D., & Pennebaker, J. (1989). Health complaints, stress, distress: Exploring the central role of negative affectivity. *Psychological Review*, *96*, 234-254. doi:10.1037/0033-295X.96.2.234

Wiechman, S., Smith, R., Smoll, F., & Placek, J. (2000). Masking effects of social desirability response set on relations between psychosocial factors and sport injuries: A methodological note. *Journal of Science and Medicine in Sport*, *3*, 194-202. doi:10.1016/S1440-2440(00)80081-X

Windle, G., Bennett, K., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and Quality of Life Outcomes*, *9*, 8. doi:10.1186/1477-7525-9-8

Zautra, A. J. (1996). Investigations of the ongoing stressful situations among those with chronic illness. *American Journal of Community Psychology*, *24*, 697-717. doi:10.1007/BF02511031