Psychological Assessment

Criterion Validation of a Stress Measure: The Stress Overload Scale

James H. Amirkhan, Guido G. Urizar, Jr., and Sarah Clark

Online First Publication, February 2, 2015. http://dx.doi.org/10.1037/pas0000081

CITATION

Amirkhan, J. H., Urizar, G. G., Jr., & Clark, S. (2015, February 2). Criterion Validation of a Stress Measure: The Stress Overload Scale. *Psychological Assessment*. Advance online publication. http://dx.doi.org/10.1037/pas0000081

Psychological Assessment
© 2015 American Psychological Association
1040-3590/15/$12.00 http://dx.doi.org/10.1037/pas0000081

Criterion Validation of a Stress Measure: The Stress Overload Scale

James H. Amirkhan, Guido G. Urizar Jr., and Sarah Clark
California State University Long Beach

Validating stress scales poses problems beyond those of other psychological measures. Here, 3 studies were conducted to address those problems and assess the criterion validity of scores from a new theory-derived measure, the Stress Overload Scale (SOS; Amirkhan, 2012). In Study 1, the SOS was tested for its ability to predict postsemester illness in a sample of college students ($n = 127$). Even with precautions to minimize criterion contamination, scores were found to predict health problems in the month following a final exam on all of 5 different criteria. In Study 2, a community sample ($n = 231$) was used to test the SOS' ability to differentiate people in stressful circumstances from those in more relaxed contexts. SOS scores demonstrated excellent sensitivity (96%) and specificity (100%) in this general population application. In Study 3, the SOS was tested for its ability to differentiate salivary cortisol responses to a laboratory stressor in a group of pregnant women ($n = 40$). High scores were found to be associated with a blunted cortisol response, which is indicative of HPA-axis overload and typical of persons suffering chronic stress and stress-related pathology. Across all 3 studies, despite variations in the stressor, criterion, population, and methods, SOS scores emerged as valid indicators of stress. However, each study also introduced new problems that beg additional corrective steps in future stress-scale validity tests. These strategies, and the SOS' utility as a research and diagnostic tool in varied applications and populations, are discussed.

Keywords: Stress measure, criterion validity, SOS, stress overload, validation methods

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Validation of psychological measures is a process fraught with difficulties. Choosing an appropriate criterion can be tricky, and once chosen, finding a reliable measure of that criterion is challenging (Cronbach & Meehl, 1955). Moreover, the criterion measure can overlap with the measure being validated, resulting in criterion contamination (Lehman, 1978). The validation of stress measures is particularly problematic, for these reasons and more. In this study, we examined the criterion validity of scores from a new stress measure, the Stress Overload Scale (SOS; Amirkhan, 2012), using a variety of methodological strategies to address these problems.

The Stress Construct(s)

Classical theories of stress date from decades ago yet still inform contemporary conceptions of stress (Wheaton & Montazer, 2010). They attempt to explain both stress itself and its relationship to illness. Some are biological in nature: Stress is defined as environmental demands, which if prolonged or frequent, exceed the body's adaptive abilities and open the door to pathology (McEwen, 2000, 2004; Selye, 1956). Others are psychological in focus: Stress is an appraisal that one's coping resources are inadequate in relation to the level of demands, a perception that prompts emotional, physiological, and behavioral changes that ultimately endanger well-being (Lazarus & Folkman, 1984). And yet other theories are economic in essence: Stress results from any expenditure of resources, so that repeated or continuing demands prompt a downward spiral that eventually renders a person vulnerable to illness (Hobfoll, 1989).

Despite their differences, stress theories have a common denominator: Stress is seen as the product of two constructs, impinging demands and compromised resources, which conjoin to produce somatic and mental changes that put people at risk for pathology (Cohen et al., 1995).

Stress Measures

Although theories portray stress as arising from two constructs, stress measures have typically assessed only one—and this single construct only sometimes corresponds to those identified by theory. Some measures focus on demands, assessing major and minor life events (e.g., the Weekly Stress Inventory, Brantley et al., 1997), but overlook the resources brought to bear on these demands. Other measures focus on resources, such as assessing vulnerability to prototypical stressors (e.g., the Perceived Stress Reactivity Scale, Schlotz, Yim, Zoccola, Jansen, & Schulz, 2011), but fail to determine the extent of actual demands. In addition, many measures ignore both demands and resources, focusing on symptoms of stress (e.g., the Index of Clinical Stress, Abell, 1991). That most do not correspond to stress theory is, in fact, an oft-

James H. Amirkhan, Guido G. Urizar, and Sarah Clark, Psychology Department, California State University Long Beach.

This research was partially supported by California State University Long Beach Faculty Research, Scholarship and Creative Activities (RSCA) Awards and by a National Institute of Mental Health (NIMH) R24 MH073882 Sub-Project Award. We express gratitude to our respective research teams for their efforts in conducting these studies.

Correspondence concerning this article should be addressed to James H. Amirkhan, Psychology Department, California State University Long Beach, 1250 Bellflower Boulevard, Long Beach, CA 90840. E-mail: james.amirkhan@csulb.edu

repeated criticism of stress measures (e.g., Derogatis & Fleming, 1997; Hobfoll et al., 1998; Lazarus, 1990).

Atheoretical approaches to stress assessment run the risk of producing measures of compromised validity. That event checklists fail to consider individual differences in resources, for example, means that their tallies will overestimate the stress level of some respondents and underestimate it for others. An implicit acknowledgment of this flaw is that many checklists now incorporate subjective "impact" or "stressfulness" ratings (e.g., Brantley et al., 1997; Kanner et al., 1981). Whether this strategy has been effective, however, requires accurate gauging of the measure's validity, which is difficult to achieve.

Problems in Validating Stress Measures

Health as Criterion

Stress theories explain how it is linked to pathology, and stress measures are often purposed toward determining risk for morbidity and mortality (e.g., Nielsen, Kristensen, Schnor, & Grønbaek, 2008). Therefore, it is logical that illness would be chosen as a criterion for validating stress scales, typically using somatic or psychiatric symptom checklists as criterion measures (e.g., Amirkhan, 2012; Brantley & Jones, 1989; Brantley et al., 1997; Cohen et al., 1983; Holmes & Rahe, 1967; Kanner et al., 1981; Levenstein et al., 1993; Sheridan & Smith, 1987).

However, there are problems with the choice of illness as a criterion. First, stress theories imply a time lag between the experience of stress and the onset of illness, as bodily and mental changes accrue to the threshold of susceptibility. But empirical studies show the length of this interval to vary greatly, from less than a month (Kendler et al., 1998) to more than 5 years (Caspi et al., 2003) for major depression. Therefore, the optimal time to administer the criterion measure after the stress measure is problematic; too soon and symptoms may not yet have developed, too late and they may have abated. Second, criterion contamination is a problem. With symptom-focused stress measures, there is often item overlap with criterion measures of both somatic (e.g., "sweaty palms") and psychiatric (e.g., "feeling overwhelmed") symptoms. In effect, the respondent is answering the same question on both the stress and the criterion measures, artificially inflating the validity correlation. Third, even without item duplication, there are response sets that can bias answers in the same direction on both measures. Negative affectivity, for example, can affect responses to both stress and symptom measures, yielding overestimates of their true level of covariation (Watson & Pennebaker, 1989).

Life Events as Criterion

The sine qua non of validity criteria for stress measures is exposure to real-world stressors. Life-event checklists have been used as criterion indices in the validation of virtually every popular stress scale (Abell, 1991; Amirkhan, 2012; Brantley & Jones, 1989; Brantley et al., 1997; Cohen et al., 1983; Derogatis & Fleming, 1997; Kanner et al., 1981; Levenstein et al., 1993; Radmacher & Sheridan, 1989). However, owing to different resources, two people facing the same life event might experience it in quite different ways— one as a threat, another as a mere challenge (Lazarus & Folkman, 1984). In addition, there are timing issues once again. The stressfulness of a life event diminishes with time (van Eck et al., 1996), but many checklists assess events over an extended period—from a week (Brantley et al., 1997) to a year (Holmes & Rahe, 1967). Therefore, two respondents might check off the same event, but for one it is a fresh wound whereas for the other it is "old news." In short, if not reflecting true stress levels, event checklists are questionable choices as criterion indices for the validation of stress measures.

Other Problems

The assessment of self-reported stress is subject to social desirability biases (Stone, 1995). Respondents may be unwilling to admit true levels of stress, for fear of appearing weak or inadequate. In addition, there may be unconscious reporting errors because of memory distortions (Stone, 1995) or defense mechanisms such as denial or repression.

Problem-Resolution Strategies

The aforementioned problems are not insurmountable, and strategies may be devised to minimize or circumvent them in validating stress measures. For fidelity with both stress theory and empirical precedent, illness may be used as a criterion but with some methodological adjustments. First, it would be helpful to control the timing as well as the type of stressor, so that all participants experience the same demand at the same point in time. This would minimize the unmeasured variations in stressor intensity that can compromise correlations with illness criteria. Second, symptoms should be assessed over a protracted period, to capture variations in the lag between stress and illness onset. Third, the stress and criterion measures should be checked for item overlap and similar items eliminated to prevent criterion contamination. Fourth, the stress and criterion measures should be administered at different time points, with enough of an intervening interval to minimize short-term response biases (such as mood or recall of prior answers). Fifth, steps should be taken to disguise the purpose and content of the measures. Ambiguous titles and filler items might offset the more persistent response biases (such as social desirability and negative affectivity).

A different strategy would be to avoid illness or any other criterion measure altogether. This would obviate all of the problems mentioned above, as well as the requirement of criterion reliability demanded by classic psychometrics (e.g., Aiken, 2000). This can be achieved by determining concurrent reliability, or a measure's ability to differentiate populations known a priori to differ in the construct of interest. For a stress scale, this would entail examining the measure's ability to discriminate between stressed and nonstressed samples.

Another strategy would be to find a criterion measure not susceptible to self-report problems, such as some irrefutable biomarker of stress. Cortisol, released by adrenal glands in response to environmental demands, holds promise as a stress marker (Austin & Leader, 2000).

The Current Studies

All of these strategies were used here in determining the criterion validity of the SOS. The SOS was chosen because it is the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

most recently published general stress measure and is argued to represent improvements over previous scales (Amirkhan, 2012).

First and foremost, the SOS was constructed to reflect the commonalities of stress theories. Potential items were selected for the SOS because they described a state of *overload,* in which life demands overwhelm one's resources to meet those demands. Subsequent analyses confirmed that these items indeed reflected two distinct constructs consistent with those identified by theory: event load and personal vulnerability.

Second, the SOS was wholly empirically derived. Using procedures prescribed by Loevinger (1957), exploratory and confirmatory factor analyses identified which among the potential items were the best markers of theoretical constructs. The remaining items were then subjected to classic psychometric tests (Cronbach & Meehl, 1955) to determine which evidenced the greatest test–retest reliability and construct validity.

Third, the SOS was constructed entirely within community samples, matched to census proportions, and diverse in terms of age, gender, ethnicity, and socioeconomic status (SES). Comprehensibility across this broad demographic spectrum constituted a third criterion for item selection.

In sum, the SOS is the end product of an evolutionary process, consisting of only those items that survived sequenced tests of theory match, psychometric strength, and demographic fit (Amirkhan, 2012). It assesses stress overload using two subscales, Event Load and Personal Vulnerability. Like other stress measures, the SOS yields continuous scores; but unlike others, its subscales may be crossed to form a diagnostic grid that assigns categorical risk scores.

Whether these features render the SOS a valid measure of stress is the focus of the present studies, each designed to correct for the problems endemic to stress-scale validation and each approved by the university's institutional review board.

Study 1: Predictive Criterion Validity

The first study made use of the traditional stress-scale criterion of illness. Students often complain that they get sick following the strain of the final weeks of the semester (e.g., Anderson, 2007). These complaints appear more than anecdotal—research has indeed tied declines in immune functioning to the stress of final exams (Uchakin et al., 2001). If SOS scores could correctly identify those students most overwhelmed by the exams, and therefore most likely to succumb to illness, this would provide evidence of their predictive validity.

Countermeasures to validation problems were built into the study's design. The type and timing of the stressor was controlled by virtue that all students experienced final exams in the same week. Symptoms were assessed over an entire month following the exams to cast a wide net in catching possible stress-related sequelae. This long span also minimized criterion contamination, because the more temporal biases (e.g., bad mood) fade with time. To counter more persistent biases (e.g., negative affectivity), the criterion measure was constructed to avoid item overlap with the SOS, and the SOS itself was disguised in ways to offset its negative tone. Finally, an embedded study was conducted to verify the reliability of the criterion measure, a critical prerequisite for determination of validity (e.g., Aiken, 2000).

Method

Participants. An upper division psychology class of 149 undergraduates was used, of whom 127 (85%) completed the study. A subset of 66 participants also completed a follow-up study to determine the reliability of the criterion health measure.

Measures. The full SOS was used; it consists of 30 items, six of which are filler items (e.g., "calm") intended to offset the generally negative tone of stress questionnaires. Each item is preceded by a prompt, "In the past week, have you felt . . ." and followed by a 5-point rating scale, ranging from 1 (*not at all*) to 5 (*a lot*). The innocuous title of "A Measure of Day-to-Day Feelings" masks the SOS' true purpose, in hopes of curtailing social desirability and negativity biases; in addition, instructions guarantee anonymity and encourage honest responding. Basic demographic questions are placed at the end to avoid possible "priming effects" (Steele, 1997).

Twelve even-numbered items on the SOS comprise the Event Load (EL) subscale, which reflects perceived demands (e.g., ". . . felt swamped by your responsibilities"). Twelve odd-numbered items comprise the Personal Vulnerability (PV) subscale, which reflects perceived inability to deal with those demands (e.g., ". . . felt like you couldn't cope"). These subscales were derived from an oblique factor solution and thus are distinct but correlated (Amirkhan, 2012). They may be summed to provide a continuous SOS total score, or they can be split at their means and crossed to provide 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).

A health survey was constructed as the criterion measure. Resembling the intake questionnaires used at doctors' offices, it consisted of a list of physical ailments and symptoms gleaned from health measures and Internet sites. In creating this list, steps were taken to eliminate any items similar to those on the SOS. The remaining items were divided into two subsections: *Illnesses,* which assessed the frequency of 10 distinct disorders (ranging from allergies to viral infections) in the preceding month, and *Symptoms,* which assessed the severity of 30 specific symptoms (from bad breath to vomiting) in the prior month. Each item was paired with a 5-point response scale, yielding possible scores of 10 to 50 for Illnesses and 30 to 150 for Symptoms. As behavioral indicators of health, two open-ended items asked for estimates of the number of *Sick Days* (days of not feeling well) and *Missed Days* (days missed at work or school) in the preceding month. Finally, a *Self-Health* rating, using a 10-point scale ranging from 1 (*very poor*) to 10 (*very good*), provided a subjective evaluation of general health. As the psychometric properties of the health survey were unknown, its test–retest reliability was determined by readministering the survey within 2 weeks to the subset of participants.

Procedure. The study began on the day of the course final exam. All students were invited to participate for extra course credit with the sole restriction that they were in good health (an alternative task was provided for symptomatic students). Of the 149 students in the class, 142 (95%) began the study by completing informed-consent forms and the SOS on site, immediately following the exam. When done, they handed their SOS (identified only by a random code) and signed consent forms to the experimenter. They then received the health survey (marked with a code match-

ing their SOS), instructions, and a preaddressed and stamped return envelope.

Instructions directed participants to wait 1 month before taking the health survey and mailing it back. The majority ($n = 127$) returned their surveys within 1 week of the deadline and were included in the study. Of these, more than half ($n = 72$) indicated that they were willing to participate in a follow-up study and provided their contact information.

On receipt of the mailings, research assistants used code numbers to pair SOS and health-survey responses for analysis. They also addressed envelopes to the follow-up study volunteers, which contained a second health survey marked with the code number. The follow-up packet was mailed within 1 day of receipt of the first envelope.

Participants in the follow-up study were instructed to complete the second health survey 1 week after the first, to write the date of completion on the measure, and to return it in a provided envelope. Completed health surveys were received from most of the volunteers ($n = 66$) within 3 days of the 1-week deadline.

Results

Sample characteristics. The sample was composed largely of younger ($M = 23.7$ years) and female students. However, owing to the diversity of the campus, it was heterogeneous in terms of ethnicity (57% non-White) and SES (representing annual incomes from less than $25,000 to more than $100,000). Although typical of the student body, these demographics do not reflect the composition of the surrounding community as may be seen in Table 1.

Study variables. Three independent variables were derived from the SOS: scores for the Personal Vulnerability and the Event Load scale, as well as their sum, the SOS total score. As seen in Table 2, all three scores showed good variability of response; that is, there was no ceiling effect despite that the measure was administered at a time assumed to be stressful.

The health survey yielded five dependent variables. Illness frequency responses were added, as were Symptom severity responses, to form two summative scale scores. The number of Sick Days, Missed Days, and the Self-Health ratings filled in by respondents constituted the other three variables. With the exception of Missed Days, which showed a basement effect, these health criteria demonstrated good variability (see Table 2).

In addition, the follow-up study provided evidence that these health scores were reliable. The test–retest intervals varied between 5 and 10 days ($M = 6.9$), yet the correlations were all significant. The Illness items demonstrated good test–retest reliability, $r = .91$, $p < .0001$. However, their internal consistency was low ($\alpha = .68$), likely because illnesses as distinct as toothaches and stomachaches do not necessarily covary. Symptom

Table 1
Demographic Composition of Study Samples

	Study 1	Study 2	Study 3	U.S. Census
Sample type	Student	Community	Community	Community
Size (*n*)	127	231	40	
Gender				
Male	27 (21%)	120 (52%)	0 (0%)	49%
Female	100 (79%)	109 (47%)	40 (100%)	51%
Age (*years*)				
18–24	98 (77%)	55 (24%)	19 (47.5%)	14%
25–34	19 (15%)	45 (19%)	19 (47.5%)	19%
35–49	6 (5%)	74 (32%)	2 (5%)	30%
50–65	4 (3%)	44 (19%)	0 (0%)	23%
>65	0 (0%)	13 (6%)	0 (0%)	14%
Ethnicity				
African American	10 (8%)	16 (7%)	12 (30%)	7%
Asian American	28 (22%)	28 (12%)	4 (10%)	15%
Latin American	24 (19%)	56 (24%)	18 (45%)	
White	54 (43%)	104 (45%)	5 (12.5%)	53%
Other–mixed	10 (8%)	20 (9%)	1 (2.5%)	25%
Education				
High school or less	19 (15%)	57 (25%)	20 (50%)	
Some college	104 (82%)	98 (42%)	10 (25%)	
College degree	4 (3%)	41 (18%)	7 (17.5%)	
Advanced degree	0 (0%)	32 (14%)	3 (7.5%)	
Income (household)				
<$25,000	56 (44%)	52 (23%)	28 (70%)	
$25,000–39,000	16 (13%)	50 (22%)	5 (12.5%)	
$40,000–$59,000	14 (11%)	25 (11%)	7 (17.5%)	
$60,000–$99,000	34 (27%)	53 (23%)	0 (0%)	
$100,000–$149,000	2 (1%)	27 (12%)	0 (0%)	
$150,000–$250,000	1 (<1%)	3 (1%)	0 (0%)	
>$250,000	1 (<1%)	4 (2%)	0 (0%)	

Note. Census numbers are 2010 figures averaged across Los Angeles and Orange Counties, California. Census age percentages are based on those more than 18 years old only. Asian American includes Pacific Islanders; Latin American is not a distinct category in Census data.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Correlations Between Stress Overload Scale (SOS) Scores and Other Variables in Study 1

Correlate	*M* (*SD*)	Range	Time 1 SOS		
			PV scale score	EL scale score	Total score
Demographics					
Age	23.7 (6.42)	18–52	−.17*	−.11	−16
Gender			.19*	.27**	.26**
Education			−.32***	−.39****	−.41****
Income			−.03	−.03	−.02
Time 1 Stress Overload Scale					
PV	23.89 (9.96)	17–53			
EL	43.15 (10.70)	16–60	.62****		
Total score	67.10 (18.62)	28–113	.89****	.91****	
Time 2 health measures					
Illnesses	6.67 (5.18)	0–30	.31***	.18*	.27**
Symptoms	23.05 (14.06)	0–77	.44****	.31***	.41****
Sick days	5.63 (5.45)	0–30	.33****	.25**	.32***
Missed days	0.54 (1.16)	0–5	.18*	.05	.13
Self-health rating	6.75 (2.06)	1–10	−.32***	−.21*	−.29**

Note. Higher gender scores indicate more female. PV = Personal Vulnerability; EL = Event Load.
* $p < .05$. ** $p < .01$. *** $p < .001$. **** $p < .0001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

items showed both good test–retest, $r = .85$, $p < .0001$, and good internal reliability ($\alpha = .91$). The single-item variables, Sick Days, $r = .84$, $p < .0001$; Missed Days, $r = .94$, $p < .0001$; and Self-Health ratings, $r = .80$, $p < .0001$, all evidenced good test–retest stability.

Continuous score tests. First, zero-order correlations were calculated among all study variables, including the demographic items (see Table 2). The SOS scales were found to be intercorrelated, as anticipated. Gender and education were associated with SOS scores, with women and lower division students having higher SOS total and subscale scores. Age and income were not consistently associated with SOS scores.

Relevant to the study's purpose, correlations showed Time 1 SOS scores to be associated with Time 2 health scores (see Table 2). The Personal Vulnerability scale correlated with all five health criteria, although only weakly with missed days. Event load and SOS total scores were related to every criterion except missed days. The general weakness in predicting days spent home from work or school may have been because of the restricted range of the missed days scores.

Although Personal Vulnerability might appear to be a stronger predictor of subsequent health than Event Load in Table 2, a series of *t* tests (McNemar, 1975) showed no significant difference between the subscales in the magnitude of their correlations to each health criterion.

Categorical score tests. Mean splits of the SOS scales were used to divide participants into the four risk categories, with persons above the mean on both scales considered to be at high risk, those below the mean on both scales at low risk, and those in off-diagonal categories (fragile or challenged) also at lower risk for subsequent illness.

Univariate analyses of variance (ANOVAs) were conducted on this 2 × 2 grid, examining each dependent variable in turn. Because the number of participants was not equal across cells, a least-squares method (GLM) was used. For nearly every health outcome, findings were as anticipated. For Illnesses, significant main effects were found for both Personal Vulnerability, $F(1, 125) = 5.47$, $p < .02$, and Event Load, $F(1, 125) = 5.85$, $p < .02$, but not for their interaction, $F(1, 123) = 1.96$, *ns*. Still, when *t* tests were used to compare the means of each quadrant, the High Risk group showed significantly more illnesses than the Low Risk, Fragile, or Challenged groups (see Table 3). The same pattern emerged for symptoms, with main effects for Personal Vulnerability, $F(1, 125) = 13.185$, $p < .001$, and Event Load, $F(1, 125) =$

Table 3
Means of Health Variables Predicted by Stress Overload Scale (SOS) Categories in Study 1

Time 2 health variable	Time 1 SOS categories	
	PV	
	Low	High
Illnesses		
EL		
Low	5.44_a	5.45_a
High	6.50_a	9.14_b
Symptoms		
EL		
Low	17.39_a	22.90_a
High	23.50_a	30.56_b
Sick days		
EL		
Low	3.92_a	5.25_a
High	5.63_a	8.25_b
Missed days		
EL		
Low	0.32_a	0.60
High	0.23_a	0.95_b
Self-health rating		
EL		
Low	7.40_a	6.85
High	6.80	5.84_b

Note. Cells have the respective sizes of low Personal Vulnerability (PV)–low Event Load (EL), $n = 48$; low PV–high EL, $n = 22$; high PV–low EL, $n = 20$; high PV–high EL, $n = 36$. Within each health variable, means with different subscripts differ at the $p < .05$ level or lower.

7.53, $p < .01$, but no significant interaction, $F(1, 123) = 0.10$, *ns*; and again for Sick Days, with significant Personal Vulnerability, $F(1, 125) = 8.47$, $p < .01$, and Event Load, $F(1, 125) = 5.48$, $p < .02$, effects and no interaction, $F(1, 123) = 0.42$, *ns*. Participants in the High Risk group reported significantly more symptoms and sick days than those in any other group (see Table 3).

The results for the other two health criteria followed this general pattern, with slight discrepancies. As above, Self-Health ratings showed main effects for both Personal Vulnerability, $F(1, 124) = 8.43$, $p < .01$, and Event Load, $F(1, 124) = 4.38$, $p < .05$, with no significant interaction. However, those in the High Risk group differed in Self-Health ratings from those in only one other group (Low Risk). For Missed Days, a significant effect was found for only Personal Vulnerability, $F(1, 125) = 6.88$, $p < .01$, and those identified as High Risk reported more missed days than those in only two of the other groups (Low Risk and Challenged). This latter deviation from the general pattern might be methodological in origin because of the restricted range of the Missed Days responses or it may reflect a real-world phenomenon; namely, that all but the sickest people drag themselves to work and school.

Discussion

Using a traditional criterion in the validation of stress measures, SOS scores were found capable of predicting poststressor illness as measured by five different indices. Moreover, this finding emerged even after steps were taken to correct stress-scale validation problems.

The categorical scoring option, by which the SOS sorts respondents into risk groups, yielded additional insights. The significant differences found between the highest and the lowest risk group on all five illness indices suggests good sensitivity for SOS categorical scores. That the high-risk group differed from all other groups on three of the five indices suggests specificity as well. Consistent with theories that say stress arises from the interplay of life demands and an inability to meet those demands (Cohen et al., 1995), current results showed both the Event Load and Personal Vulnerability subscales to play equally important roles in predicting illness.

Even with the corrective steps taken, this study had limitations. First, the sample consisted entirely of students and was not representative of the general population. Final exams were also not typical of the stressors that most of the population faces. Second, the stress and criterion measures all relied on self-report, therefore the possibility remains that shared-method variance accounted for the correlations among them, despite the precautions against criterion contamination. Finally, the criterion measures, although demonstrating good reliability, were of unknown validity. The missed day count was of particular concern: It did not generate the variability of the other criterion indices and perhaps should be avoided as an illness marker.

Study 2: Concurrent Criterion Validity

The second study was designed to avoid the validation problems inherent in the use of a criterion measure and to compensate for some of the shortcomings of the first study. It did so by determining whether the SOS could differentiate between people in stressed versus relaxed circumstances. Two samples were drawn from the general community: One, a group of litigants, defendants, jurors, and lawyers at a courthouse on an early weekday; the other, a group of vacationers, sightseers, and families at an aquarium at midday on a weekend. If the SOS could discriminate those at the contentious legal setting from those at the tranquil tourist attraction, this would provide evidence of concurrent validity. In addition, being community-based, this study procured a wider spectrum of stressors and demographics than that afforded by the prior college sample.

Method

Participants. Two hundred fifty residents were sampled from two Southern California counties, a region that is diverse in terms of ethnic background and SES. Of these, 231 (92%) returned surveys complete enough for analysis.

Measure. The full, 30-item SOS was again used for this study. The precautions to minimize response biases (benign title, filler items) and to encourage honest responding (anonymity) were retained, despite that there was no criterion measure.

Procedure. To obtain a stressed group, participants were recruited in front of a county courthouse on weekdays at 7:00 a.m. as they reported for trial or jury duty. For a nonstressed group, recruitment took place in front of a renowned aquarium at noon on weekends as the attraction opened. In addition to their potential to yield criterion groups, these sites were chosen for their promise in procuring a wide demographic spectrum.

Convenience sampling was used, with both passive recruitment (banners displayed to attract potential participants) and active recruitment (research assistants approaching people and requesting participation). All parties recruited by either method were screened for age (>18 years) and English fluency.

Twelve research assistants, six at each site, conducted the research. Although some were engaged in recruitment, others staffed tables. Here, they conducted the screening and informed consent processes. They then handed out the survey materials to participants who completed these processes; materials that included questionnaires, envelopes, clipboards, and pencils. They instructed participants to complete the questionnaires on site, out of view of the researchers or others, and then seal their responses into the envelopes and return them to the table. At this point, they told participants to deposit their sealed envelopes into a locked collection box, and handed them their incentive, a one-dollar lottery scratcher ticket.

Results

Sample characteristics. Because no data were obtained from nonparticipants, it is not certain whether those who self-selected into the study were representative of the general population. However, in comparing participant demographics to U.S. Census figures, it appears that the sample paralleled the diversity of the region (see Table 1). The number of men (52%) and women (47%) was equitable, ages ranged from 18 to 85 years old ($M = 37.9$), and multiple ethnicities were represented in proportions comparable to census norms. Education levels (25% with a high-school diploma or less, 32% with college or advanced degrees) and income levels (23% living in near poverty, 14% earning more than $100,000 per year) varied widely, indicating the representation of a broad socioeconomic spectrum.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The courthouse and aquarium groups were compared in terms of demographic composition. Tests showed significant between-groups differences in age, $t(226) = 2.27$, $p < .03$, with older participants in the aquarium group. However, age did not correlate significantly with SOS scores, $r = -.12$, $p = .10$, therefore, it was not controlled as a potential confound in subsequent analyses. Chi-square tests showed no significant differences between groups in the categorical variables of gender, ethnicity, education, or income.

Study variable. The SOS produced good variability of response, across and within groups (see Table 4). Overall, scores ranged from 24 (the lowest possible value) to 116 (near the highest possible value of 120).

Continuous score tests. A preliminary indication of discriminative ability was obtained by correlating SOS scores with group, coded such that the more stressful site was assigned a higher value (courthouse = 2, aquarium = 1). As seen in Table 4, SOS total and subscale scores all correlated significantly, in the expected direction, with group membership. Moreover, tests of the magnitude of the correlations obtained (McNemar, 1975) showed no significant differences between the Personal Vulnerability and Event Load subscales.

More formal analysis of differences between the courthouse ($n = 121$) and aquarium groups ($n = 113$) made use of independent sample t tests. These analyses revealed significant group differences in SOS total, $t(232) = 22.01$, $p < .0001$, and subscale scores: For Event Load, $t(232) = 16.14$, $p < .0001$; for Personal Vulnerability, $t(232) = 19.21$, $p < .0001$. All differences were in the expected direction, with the courthouse group obtaining higher scores than the aquarium group.

Categorical score tests. SOS categorical scores were assigned to members of both the courthouse and the aquarium groups. That is, group means on the EL and PV subscales were used to divide that group's members into the four diagnostic categories.

As is apparent in Table 5, the courthouse group had a greater proportion of its members in the High Risk category and a lesser proportion in the Low Risk category than the aquarium group. To test this difference, participants were first assigned a rank score according to the category into which they fell: 1 for Low Risk, 3 for High Risk, and 2 for the off-diagonal (Challenged or Fragile) categories. Then, a Mann–Whitney test was used to determine whether there were significant differences between the stressed and nonstressed groups in these rankings, which proved to be the case ($U = 652.00$, $p < .0001$).

The diagnostic grid in Table 5 was also used to estimate the sensitivity and specificity of the SOS. If it is assumed that everyone at the courthouse was stressed, then the SOS categorized 75 people correctly (true positives) and 3 incorrectly (false negatives) as high risk participants. Assuming everyone at the aquarium was nonstressed, the SOS identified 90 correctly (true negatives) and 0 incorrectly (false positives) as low risk participants. This yielded a sensitivity of 96.15% (95% confidence interval [CI] [89.15, 99.16]) and a specificity of 100% (95% CI [95.94, 100]).

Discussion

Designed to avoid the problems associated with criterion measurement and to use a sample more representative of the general population, the second study verified the concurrent validity of SOS items. Results showed that SOS scores could successfully discriminate between respondents at a courthouse (assumed to be stressed) and those at a tourist attraction (assumed to be calm). Moreover, the SOS diagnostic grid demonstrated excellent sensitivity and specificity in identifying which respondents were from which site. Whether other stress measures have comparable sensitivity and specificity is unknown because they do not provide a rubric for categorizing respondents into risk groups.

Although sidestepping some issues of the first study, this second study had potential shortcomings of its own. First, the sample was self-selected, not only for attendance at a site, but additionally for participation in the study. Therefore, although demographically diverse, it is unknown whether study participants were representative of the general population in all ways. Second, there was no means of verifying that the courthouse did indeed draw more highly stressed people than the aquarium. The persons sampled from these sites may have differed in ways other than stress level (e.g., mood) that the SOS erroneously detected.

Table 4
Correlations Between Stress Overload Scale (SOS) Scores and Other Variables in Study 2

	M (*SD*)			SOS		
Correlate	Aquarium	Courthouse	Range	PV scale score	EL scale score	Total score
Demographics						
Age	38.67 (14.21)	36.96 (15.74)	18–85	−.13	−.07	−.12
Gender				−.04	.04	−.01
Education				−.18*	−.00	−.09
Income				−.22*	−.01	−.07
Stress Overload scale						
PV	19.02 (5.56)	38.51 (9.35)	12–58			
EL	28.61 (8.88)	45.88 (7.47)	12–60	.69****		
Total score	47.63 (11.59)	84.39 (13.78)	24–116	.92****	.92****	
Participant group						
Site				.45****	.39****	.45****

Note. Higher gender scores indicate more female. Higher site scores indicate courthouse. PV = Personal Vulnerability; EL = Event Load.
* $p < .05$. **** $p < .0001$.

Table 5
Frequencies in Stress Overload Scale (SOS) Categories by Group in Study 2

	SOS categories	
	PV	
Participant group	Low	High
Courthouse		
EL		
Low	3 (3%)	16 (13%)
High	25 (21%)	75 (63%)
Aquarium		
EL		
Low	90 (80%)	2 (2%)
High	20 (18%)	0 (0%)

Note. PV = Personal Vulnerability; EL = Event Load.

Study 3: Concurrent Criterion Validity

The third study was designed to avoid the self-report problems associated with use of a criterion measure and also to correct for the ambiguities of the prior study. It did so by using a biomarker of stress as the criterion, assessing salivary cortisol in a group of pregnant women.

Although cortisol is a documented marker of stress (Austin & Leader, 2000), results from studies using a variety of stressors and participant samples have yielded inconsistent and largely nonsignificant relationships between absolute cortisol levels and self-reported stress (Hellhammer et al., 2009). This is likely because cortisol is routinely released by the HPA axis over the course of the day and in response to passing demands. It is the frequent or chronic activation of the HPA axis that has been associated with subjective stress and several negative health outcomes (including depression, cognitive decline, and cardiovascular disease; see Adam & Kumari, 2009). In regard to this sample, persistent activation of the HPA axis also negatively affects pregnancy, having been linked to adverse outcomes such as low infant birth weight and long-term developmental problems (Gunnar, 1998; Kurstjens & Wolke, 2001). For these reasons, researchers have sought means of assessing chronic HPA activation, and several have shown abnormal cortisol responses to laboratory-induced stressors to be a reliable indicator (Campbell & Ehlert, 2012; Vedhara et al., 2003).

The Trier Social Stress Test (TSST) is a laboratory procedure that has been used to induce changes in salivary cortisol (both in general population and pregnant samples) through the use of standardized stress-generating tasks (De Weerth, Wied, Jansen, & Buitelaar, 2007; Kirschbaum, Pirke, & Hellhammer, 1993). The TSST also has advantages that make it suitable for stress-scale validation: It avoids reliance on self-report, controls for the type and timing of the stressor, and allows for a prolonged reading of physiological sequelae, assessing salivary cortisol production for 60 min after the stressor. Typically, it is not high cortisol levels that are indicative of HPA-axis overload. Rather, a blunted cortisol response, reflecting compromised cortisol production, is often observed among those who experience chronic stress and depression (Burke et al., 2005). For this reason, this study examined whether the SOS, a measure of psychological overload, could predict blunted cortisol reactions to the TSST, an indication of physiological overload.

Method

Participants. A subset of 40 pregnant women was recruited from community sites in Southern California as part of a larger study (n = 100) examining depression risk and health during pregnancy (Urizar, 2012). This subset provided data for the current analyses.

Measures. The full, 30-item SOS was used to examine psychological stress for this study. Physiological stress levels were assessed through six salivary cortisol samples that were collected during the TSST: One at baseline, one immediately after the TSST tasks, and others at 15, 30, 45, and 60 min after the TSST tasks. Salivary cortisol samples were frozen and stored until radioimmunoassay. The resulting readings were then logarithmically transformed (base 10, converted from nmol/L) because salivary cortisol values are typically skewed. Three dependent variables were formed: Total cortisol output during the TSST (i.e., area under the curve [AUC]), differences between baseline and 15-min post-TSST values (cortisol reactivity), and differences between 15- and 60-min post-TSST values (cortisol recovery). For all dependent variables, lower values indicate a blunted cortisol response to the laboratory stressor.

Procedure. Women were eligible for the study if they were 18 years of age or older, between 8 and 25 weeks pregnant, fluent in either English or Spanish, and free of any major medical or psychiatric disorders. A quota-sampling technique was used to obtain equal numbers of low-income Latinas, low-income African American women, and middle- to high-income women (regardless of ethnicity). Income status was defined as having public (low income) versus private health insurance (middle to high income).

Women were recruited from several local prenatal clinics and community centers for a study on depression risk and health during pregnancy. Those who were eligible took part in a 2-hr, clinic-based TSST, scheduled between 2 p.m. and 6 p.m. to minimize normal diurnal cortisol fluctuations (salivary cortisol levels are most stable later in the day; Kirschbaum et al., 1993). Free transportation and childcare were offered on the day of the clinic visit.

Participants were instructed to refrain from eating, drinking, or smoking 1 hour before their arrival to their appointment to minimize factors known to interfere with salivary cortisol samples. At the beginning of the clinic visit (during the initial 20-min rest and acclimation period of the TSST), participants completed the SOS and other psychosocial questionnaires and provided a baseline salivary cortisol sample. This required the participants to chew on a cotton swab for 1 min or until the cotton swab became soaked with saliva, which was then placed into a plastic tube and stored in a −20 °C freezer until analysis. Next, in the reactivity phase of the TSST, participants were led to a second room to perform public speaking and mental arithmetic tasks (the stressors) in front of two judges who videotaped the performances. These tasks lasted a total of 15 min. Then, participants exited the room and immediately provided a second saliva sample to assess for cortisol reactivity to the TSST. For the final recovery phase, participants provided additional saliva samples at 15, 30, 45, and 60 min following the TSST. During the recovery phase, women participated in an open-ended qualitative interview that assessed their coping style and

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

their knowledge of available prenatal resources. Such tasks have been shown to not affect cortisol levels during the TSST recovery period (Kirschbaum et al., 1993). Participants were compensated for their time and effort with a $25 gift card.

Results

Sample characteristics. The purposeful sampling strategy divided participants evenly among the targeted ethnicity–SES categories: 14 were low-income Latina women, 13 were low-income African American women, and 13 were middle- to high-income women of varying ethnicities. Their ages ranged from 18 to 39 years ($M = 26$), and their gestation periods varied between 8 and 25 weeks of pregnancy ($M = 17$). Most had a husband or partner (63%), a high school education (50%), and an annual household income under $25,000 (70%). These demographics were not representative of the general community (see Table 1). Because none of these demographic variables covaried with SOS scores (see Table 6), they were not treated as potential confounds in the analyses.

Study Variables

SOS scores. The SOS produced good variability of response, across and within groups (see Table 6). Overall, scores ranged from 27 (near the lowest possible value of 24) to 100 (near the highest possible value of 120).

Because norms for the SOS were derived from general population samples (Amirkhan, 2012), Cronbach's alpha coefficients for the measure were calculated for this specialized sample. These showed good internal consistency for the SOS as a whole (.89), as well as for the Personal Vulnerability (.89) and Event Load (.91) subscales.

Salivary cortisol. Pairwise *t* tests revealed that the TSST tasks did produce changes in cortisol levels. Cortisol reactivity, the difference between baseline ($M = 8.26$ nmol/L) and 15-min post-TSST ($M = 11.03$ nmol/L) cortisol levels, significantly increased as expected, $t(38) = -2.99$, $p < .01$. Cortisol recovery, the difference between 15- ($M = 11.03$ nmol/L) and 60-min post-TSST ($M = 7.85$ nmol/L) cortisol levels, significantly decreased over time as expected, $t(37) = 5.15$, $p < .001$.

Continuous Score Tests

Correlations were used to determine whether the SOS was associated with overall cortisol output (AUC), cortisol reactivity, or cortisol recovery. These showed SOS total and subscale scores all to be significantly and negatively associated with AUC cortisol readings, such that higher SOS scores were associated with lower cortisol output during the TSST (see Table 6). No significant associations were found between the SOS and the reactivity or recovery variables.

Categorical Score Tests

The categorical scoring option of the SOS was used to divide participants into risk categories, using group means as dividing points on the Personal Vulnerability and Event Load subscales. Most women fell into the Low Risk ($n = 17$) or High Risk quadrant ($n = 14$), with relatively few in the off-diagonal categories (Fragile, $n = 2$; Challenged, $n = 7$).

Because of the low number of participants in the off-diagonal quadrants, a one-way rather than a factorial ANOVA was conducted on the cortisol scores. Specifically, a general linear model (GLM) was used to examine whether three SOS categories (High Risk, Low Risk, and combined Fragile/Challenged) differed in AUC, cortisol reactivity, or cortisol recovery, adjusting for gestational age. Results showed a significant result for AUC, $F(2, 37) = 5.84$, $p < .05$, such that the low risk participants had the highest overall cortisol output ($M = 73.26$ nmol/L, $SD = 3.38$), followed by those in the off-diagonal groups ($M = 70.02$ nmol/L, $SD = 5.02$), with the high risk participants having the lowest output ($M = 67.42$ nmol/L, $SD = 3.51$). This indicates an overall blunted cortisol response among members of the High Risk group; no

Table 6
Correlations Between Stress Overload Scale (SOS) Scores and Other Variables in Study 3

Correlate	*M* (*SD*)	Range	SOS		
			PV scale score	EL scale score	Total score
Demographics					
Age	25.53 (5.39)	18–39	−.21	−.16	−.20
Gestational age	17.30 (4.53)	8–25	−.16	−.27	−.24
No. of children	0.95 (1.19)	0–4	.24	.12	.19
Education			−.23	.03	−.09
Income			−.09	.03	−.02
Stress Overload scale					
PV	24.08 (9.26)	12–54			
EL	33.48 (12.27)	12–56	.69***		
Total score	57.55 (19.80)	27–100	.89***	.94***	
Salivary cortisol					
Cortisol reactivity	2.80 (5.86)	−5.00–26.49	−.05	−.14	−.11
Cortisol recovery	3.18 (3.81)	−6.01–17.16	−.04	−.01	−.03
AUC	734.59 (427.30)	246.76–2561.13	−.32*	−.38*	−.38*

Note. Salivary cortisol values are presented in nmol/L; however, correlation analyses were conducted with the log scores of these variables. PV = Personal Vulnerability; EL = Event Load; AUC = area under the curve.
* $p < .05$. *** $p < .001$.

significant group differences were found for the reactivity or recovery variables.

Discussion

Avoiding the problems inherent to self-report criterion measures, this study used a documented biomarker of stress (Burke et al., 2005) to test the SOS validity. Results showed that pregnant women who had higher SOS scores (either continuous or categorical) exhibited a blunted biological response to laboratory-induced stress, secreting significantly less cortisol over the course of the laboratory trial. However, high- and low-scorers showed no significant differences in either short-term cortisol reactivity or in cortisol recovery. This may signal a limit to the SOS predictive power, namely that it can detect only broad rather than moment-by-moment stress reactions. Yet broad deficits in physiological response may be most indicative of a general state of overload, the very state that the SOS was designed to assess.

Any such interpretations must be made with caution given the constraints of the study. First, the sample consisted entirely of pregnant women, and it is possible that their reactions were not typical of a more general population. Second, this sample size was small, which may have further compromised its representativeness and did constrain data analyses. Finally, salivary cortisol levels can be affected by a number of factors, including eating, drinking, smoking, medications, time of saliva collection, and gestational age (Hellhammer, Wüst, & Kudielka, 2009). Although precautions were taken to minimize the impact of such extraneous factors in this study, they may nevertheless have introduced noise into the cortisol readings, error variance into the analyses, and imprecision into the resultant validity estimates.

General Discussion

Multiple strategies were used here to address the problems inherent to the validation of stress measures. The type and timing of stressors was fixed, wide nets were cast to catch stress sequelae, and varied safeguards against criterion contamination were employed. The SOS (Amirkhan, 2012) was the focus of this endeavor because it had been empirically constructed for consistency with stress theories and thereby held the greatest promise for detecting actual stress phenomena. Indeed, across all of the current studies—despite variations in validation methods (predictive and concurrent), sampled populations (student, community, pregnant), stressors (exams, life events, lab tasks), and stress criteria (health-related, contextual, hormonal)—SOS scores demonstrated good criterion validity.

Further Problem-Resolution Strategies

Although designed to address methodological weaknesses, these studies had shortcomings of their own, indicating the need for additional corrective strategies in future research. The criterion measures constructed for the first study, although shown to be reliable, had no more than face validity. It is not known whether such symptom and illness lists actually capture the full range of stress sequelae or even reflect true states of pathology. A nonconfounding, complete and valid inventory of stress-related disorders would greatly benefit future studies. Alternatively, the identification of reliable illness markers (either behavioral, such as medical bills, or somatic, such as blood pressure) would allow future researchers to evade self-report and shared-method issues entirely. It should be noted that current findings indicate that days missed from work or school may not be a good marker of stress sequelae.

The concurrent validity test used in the second study showed the SOS to differentiate groups presumed to be stressed versus relaxed. However, it is not known if these groups differed in other ways or even if they actually differed in stress level. In the future, it would be useful to use groups that leave little doubt as to stress level, sampled perhaps from high-demand occupations (e.g., air-traffic controllers, combat soldiers) or trying circumstances (e.g., high-crime neighborhoods, prisons). Additionally, larger samples should be used, both to ensure representativeness and provide a broader test of the measure's sensitivity. The use of multiple stress measures could verify group differences, provide cross-validation, and indicate the discriminative ability of the new scale relative to that of existing ones.

The use of a stress biomarker as the criterion in the third study avoided self-report problems but required invasive and expensive laboratory procedures. More problematic was that salivary cortisol readings are easily perturbed by momentary factors, both intended (induced stressors) and unintended (such as caffeine or nicotine consumption). If future researchers are willing to bear the expense of biochemical assays, a biomarker unique to prolonged stress overload would be a preferable criterion. Markers of chronic inflammation (such as glucocorticoid receptor resistance; Cohen et al., 2012) show promise in this regard.

It is unlikely that any one study, however, can ever address all the problems endemic to stress-scale validation. Multiple studies offer a solution. Here, the strengths of one study compensated for the limitations of another: Restricted samples were augmented with a general population sample, self-report criterion measures were counterbalanced with objective validity criteria, potentially confounded criteria were offset by established biomarkers, and so on.

Stress Overload Scale

The evidence presented here seems to verify the SOS' place in the field of stress assessment. However, it should not imply that the SOS is the best stress measure for all purposes. First, current findings showed the measure is not sensitive to moment-by-moment fluctuations in stress level. Therefore, the SOS would not be appropriate to tracking changes over the course of any single stressful episode, a purpose perhaps better served by physiological readings. Second, the SOS is a subjective measure, therefore it would not be useful in identifying the type or frequency of specific stressors in a person's life; such purposes beg the use of objective event inventories. Finally, the 30-item SOS may impose too great a respondent load for some purposes. In trauma contexts, a brief stress scale would be the better choice (e.g., Lee, 2012).

Still, SOS continuous scores were found to covary with a variety of stress criteria, from traditional (illness) to cutting-edge (cortisol levels), which indicates its potential as a research tool. Additionally, SOS categorical scores demonstrated good sensitivity and specificity, suggesting its use as a diagnostic tool. For example, the SOS could be used both for exploring links among prolonged stress, chronic inflammation, and disease genesis (Cohen et al.,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

2012; Hunter, 2012), and for identifying the persons at greatest risk for those diseases. In such applications, owing to evidence of its validity across all current samples, the SOS promises to be useful with diverse populations in varied contexts suffering a gamut of life challenges.

References

Abell, N. (1991). The Index of Clinical Stress: A brief measure of subjective stress for practice and research. *Social Work Abstracts and Research, 27,* 12–15. http://dx.doi.org/10.1093/swra/27.2.12

Adam, E. K., & Kumari, M. (2009). Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology, 34,* 1423–1436. http://dx.doi.org/10.1016/j.psyneuen.2009.06.011

Aiken, L. R. (2000). *Psychological testing and assessment* (10th ed.). Boston, MA: Allyn & Bacon.

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

Anderson, A. (2007, January 8). Why do I always get sick after final exams? Retrieved from http://scienceline.org/2007/01/ask-anderson-finalscough/

Austin, M. P., & Leader, L. (2000). Maternal stress and obstetric and infant outcomes: Epidemiological findings and neuroendocrine mechanisms. *Australian and New Zealand Journal of Obstetrics and Gynaecology, 40,* 331–337. http://dx.doi.org/10.1111/j.1479-828X.2000.tb03344.x

Brantley, P., & Jones, G. (1989). *The Daily Stress Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.

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

Burke, H. M., Davis, M. C., Otte, C., & Mohr, D. C. (2005). Depression and cortisol responses to psychological stress: A meta-analysis. *Psychoneuroendocrinology, 30,* 846–856. http://dx.doi.org/10.1016/j.psyneuen.2005.02.010

Campbell, J., & Ehlert, U. (2012). Acute psychosocial stress: Does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology, 37,* 1111–1134. http://dx.doi.org/10.1016/j.psyneuen.2011.12.010

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., . . . Poulton, R. (2003, July 18). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science, 301,* 386–389. http://dx.doi.org/10.1126/science.1083968

Cohen, S., Janicki-Deverts, D., Doyle, W. J., Miller, G. E., Frank, E., Rabin, B. S., & Turner, R. B. (2012). Chronic stress, glucocorticoid receptor resistance, inflammation, and disease risk. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 5995–5999. http://dx.doi.org/10.1073/pnas.1118355109

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

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. http://dx.doi.org/10.1037/h0040957

Derogatis, L., & Fleming, M. (1997). The Derogatis Stress Profile (DSP). In C. Zalaquett & R. Wood (Eds.), *Evaluating stress* (pp. 113–140). Lanham, MD: Scarecrow.

De Weerth, C., Wied, G. D., Jansen, L. M., & Buitelaar, J. K. (2007). Cardiovascular and cortisol responses to a psychological stressor during pregnancy. *Acta Obstetricia et Gynecologica Scandinavica, 86,* 1181–1192. http://dx.doi.org/10.1080/00016340701547442

Gunnar, M. R. (1998). Quality of early care and buffering of neuroendocrine stress reactions: Potential effects on the developing human brain. *Preventive Medicine, 27,* 208–211. http://dx.doi.org/10.1006/pmed.1998.0276

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology, 34,* 163–171. http://dx.doi.org/10.1016/j.psyneuen.2008.10.026

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

Hobfoll, S., Schwarzer, R., & Chon, K. (1998). Disentangling the stress labyrinth: Interpreting the meaning of the term stress as it is studied in the health context. *Anxiety, Stress & Coping, 11,* 181–212. http://dx.doi.org/10.1080/10615809808248311

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

Hunter, P. (2012). The inflammation theory of disease: The growing realization that chronic inflammation is crucial in many diseases opens new avenues for treatment. *European Molecular Biology Organization Reports, 13,* 968–970. http://dx.doi.org/10.1038/embor.2012.142

Kanner, A. D., Coyne, J. C., Schaefer, C., & Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of Behavioral Medicine, 4,* 1–39. http://dx.doi.org/10.1007/BF00844845

Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1998). Stressful life events and major depression: Risk period, long-term contextual threat, and diagnostic specificity. *Journal of Nervous and Mental Disease, 186,* 661–669. http://dx.doi.org/10.1097/00005053-199811000-00001

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28,* 76–81. http://dx.doi.org/10.1159/000119004

Kurstjens, S., & Wolke, D. (2001). Effects of maternal depression on cognitive development of children over the first 7 years of life. *Journal of Child Psychology and Psychiatry, 42,* 623–636. http://dx.doi.org/10.1111/1469-7610.00758

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

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

Lee, E. H. (2012). Review of the psychometric evidence of the perceived stress scale. *Asian Nursing Research, 6,* 121–127. http://dx.doi.org/10.1016/j.anr.2012.08.004

Lehman, R. (1978). Symptom contamination of the Schedule of Recent Events. *Journal of Consulting and Clinical Psychology, 46,* 1564–1565. http://dx.doi.org/10.1037/0022-006X.46.6.1564

Levenstein, S., Prantera, C., Varvo, V., Scribano, M. L., Berto, E., Luzi, C., & Andreoli, A. (1993). Development of the Perceived Stress Questionnaire: A new tool for psychosomatic research. *Journal of Psychometric Research, 37,* 19–32. http://dx.doi.org/10.1016/0022-3999(93)90120-5

Loevinger, J. (1957). Objective tests as instruments of psychological theory: Monograph Supplement 9. *Psychological Reports, 3,* 635–694. http://dx.doi.org/10.2466/pr0.1957.3.3.635

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

McEwen, B. S. (2004). Protective and damaging effects of the mediators of stress and adaptation: Allostasis and allostatic load. In J. Schulkin (Ed.), *Allostasis, homeostasis and the costs of physiological adaptation* (pp. 65–98). New York, NY: Cambridge University Press.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

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

Nielsen, N. R., Kristensen, T. S., Schnohr, P., & Grønbaek, M. (2008). Perceived stress and cause-specific mortality among men and women: Results from a prospective cohort study. *American Journal of Epidemiology, 168,* 481–491. http://dx.doi.org/10.1093/aje/kwn157

Radmacher, S., & Sheridan, C. (1989). The Global Inventory of Stress: A comprehensive approach to stress assessment. *Medical Psychotherapy: An International Journal, 2,* 75–80.

Schlotz, W., Yim, I. S., Zoccola, P. M., Jansen, L., & Schulz, P. (2011). The Perceived Stress Reactivity Scale: Measurement invariance, stability, and validity in three countries. *Psychological Assessment, 23,* 80–94. http://dx.doi.org/10.1037/a0021148

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

Sheridan, C. L., & Smith, L. K. (1987). Toward a comprehensive scale of stress assessment: Development, norms and reliability. *International Journal of Psychosomatics: Official Publication of the International Psychosomatics Institute, 34,* 48–54.

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

Stone, A. (1995). Measurement of affective response. In S. Cohen, R. Kessler, & L. Gordon (Eds.), *Measuring stress* (pp. 148–171). New York, NY: Oxford University Press.

Uchakin, P. N., Tobin, B., Cubbage, M., Marshall, G., Jr., & Sams, C. (2001). Immune responsiveness following academic stress in first-year medical students. *Journal of Interferon & Cytokine Research, 21,* 687–694. http://dx.doi.org/10.1089/107999001753124426

Urizar, G. (2012). Socio-demographic differences on coping styles and cortisol patterns during pregnancy. *Annals of Behavioral Medicine, 43,* S125.

van Eck, M., Berkhof, H., Nicolson, N., & Sulon, J. (1996). The effects of perceived stress, traits, mood states, and stressful daily events on salivary cortisol. *Psychosomatic Medicine, 58,* 447–458. http://dx.doi.org/10.1097/00006842-199609000-00007

Vedhara, K., Miles, J., Bennett, P., Plummer, S., Tallon, D., Brooks, E., . . . Farndon, J. (2003). An investigation into the relationship between salivary cortisol, stress, anxiety and depression. *Biological Psychology, 62,* 89–96. http://dx.doi.org/10.1016/S0301-0511(02)00128-X

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

Wheaton, B., & Montazer, S. (2010). Stressors, illness, and distress. In T. Scheid & T. Brown (Eds.), *A handbook for the study of mental health* (pp. 171–199). New York, NY: Cambridge University Press.

Received December 6, 2013
Revision received November 24, 2014
Accepted December 3, 2014 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.