

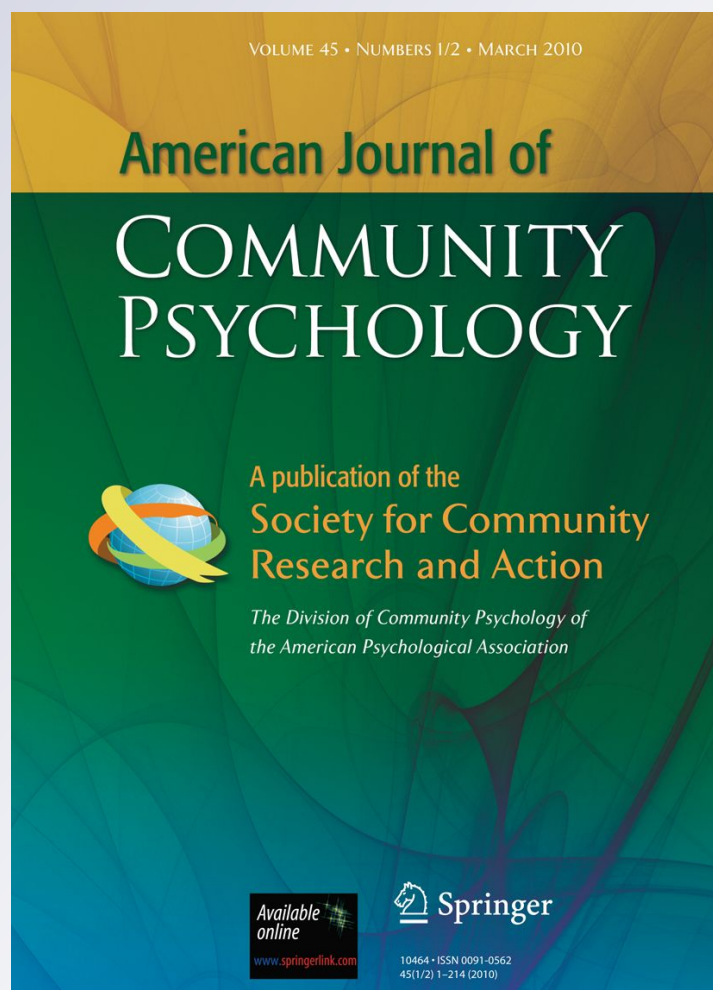
Stress Overload: A New Approach to the Assessment of Stress

James H. Amirkhan

American Journal of Community Psychology

ISSN 0091-0562
Volume 49
Combined 1-2

Am J Community Psychol (2012)
49:55-71
DOI 10.1007/s10464-011-9438-x



Your article is protected by copyright and all rights are held exclusively by Society for Community Research and Action. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Stress Overload: A New Approach to the Assessment of Stress

James H. Amirkhan

Published online: 3 May 2011

© Society for Community Research and Action 2011

Abstract The link between stress and health has not received strong empirical support, possibly due to problems in the stress measures used. Here, the first wholly empirical development of a new “Stress Overload Scale” is described. A pool of 150 items was formed to reflect “overload”, a common denominator in stress theories. Then, the results of five sequenced studies, conducted in heterogeneous community samples, were used to pare the item pool. Exploratory ($n = 431$) and confirmatory ($n = 433$) analyses revealed two factors (Event Load and Personal Vulnerability) corresponding to theoretical constructs; only the best factor markers were submitted to further construct validity ($n = 310$) and reliability tests ($n = 342$). The 24 strongest items were selected for the SOS, which demonstrated criterion validity in predicting who ($n = 285$) would become sick following a common stressor. The SOS is (1) psychometrically strong, especially in its validity relative to popular measures; (2) appropriate to community research, due to its brevity and fit to a broad demographic spectrum; (3) unique in its ability to cross-section individuals into risk categories. It should prove useful to community psychologists in determining etiology, diagnosing risk for pathology, and evaluating the efficacy of interventions.

Keywords Stress · Measure · Health · Risk factor · Self-report · Assessment

A fundamental assumption of community psychology is that stress affects health (Dohrenwend 2002). Much of

community research (such as studies seeking to identify at-risk groups; e.g., Nielsen et al. 2008) and community practice (such as crisis intervention programs; e.g., Auerbach and Stolberg 1986) is based on this premise. But, although there is empirical evidence of a link between stress and both mental (e.g., Turner and Lloyd 1995) and physical health (e.g., Bernard and Krupat 1994), the correlations are typically of low magnitude (Dohrenwend and Dohrenwend 1981, 1984; Gentry and Kobasa 1984; Lin and Ensel 1989; Sarason and Sarason 1984) and often do not reach significance (e.g., Bernard and Krupat 1994). Moreover, even the significant relationships are suspected of being artifactual (Dohrenwend and Dohrenwend 1984; Watson and Pennebaker 1989).

Among the myriad explanations offered for these unexpectedly weak findings, one is that current stress measures are inadequate (Dohrenwend 2006; Hobfoll et al. 1998; Lazarus 1990; Turner and Lloyd 1995). Available instruments have been criticized for being atheoretical (Derogatis and Fleming 1997; Hobfoll et al. 1998; Lazarus 1990), psychometrically unsound and impractical (Dohrenwend 2006). Here, the construction of a new questionnaire, using empirical methods in large and heterogeneous community samples, is described. The Stress Overload Scale (SOS) is believed to represent an improvement over existing scales (1) *Conceptually*, in that it is derived from constructs shared by stress theories, (2) *Psychometrically*, in that it offers both reliability and superior validity, and (3) *Practically*, in that it poses a small respondent burden and is appropriate to diverse population groups.

Conceptual Basis

The classic theories of stress both defined it and explicated the mechanism by which it impacts health. Selye's (1956)

J. H. Amirkhan (✉)

Psychology Department, California State University Long Beach, Long Beach, CA 90840, USA
e-mail: jamirkha@csulb.edu

General Adaptation Syndrome is a seminal model. Biological in nature, it essentially defines stress as any threat to homeostasis. In most cases, stress is resolved when such threats are met and countered by the body's resistive resources. However, when adaptational demands exceed the body's ability to resist, the result is ill health or even death.

Selye's model has been refined to fit more recent physiological evidence. McEwen (2000, 2004) argued that homeostasis should not be the focal point of a stress definition, for homeostatic mechanisms (e.g., Ph balance, body temperature) are few, and do not fluctuate in response to environmental challenges. Most body systems are allostatic (e.g., blood pressure, immune response), and vary to meet adaptational demands. A more precise definition of stress should focus on these mechanisms, which promote adaptation in the short run but which can have damaging effects over time. Specifically, it is "allostatic load", the result of prolonged or repeated demands and/or a compromised response to these demands, that causes allostasis to become dysregulated and pathogenic. This is "the price the body pays for being forced to adapt" (2000, p. 174).

In another departure from Selye's model, which emphasized physical demands, McEwen (2000) cited evidence that psychological stressors have a greater impact on functioning. In this regard, he echoed another seminal theory, that of Lazarus (Lazarus and Folkman 1984). It was Lazarus' position that stress is a wholly psychological rather than physiological phenomenon, constructed from an appraisal of the balance between perceived demands and perceived resources. Specifically, stress and ultimately pathology result when "a particular relationship between the person and the environment is appraised by the person as taxing or exceeding his or her resources and endangering his or her well-being" (p. 19).

Conservation of Resources theory (Hobfoll 1989; Hobfoll et al. 1998) proposed an economy that weighs both the perceived and the objective. In essence, stress is the loss of resources, real or threatened, over and above any resource gain. In this calculus, the subjective weight attached to resource loss is greater than that for resource gain. Moreover, loss can initiate a "spiraling sequence" that ultimately renders a person vulnerable to illness: "When resources are lost people are at increased vulnerability both because they have lost resources and because they thereby have a weaker resource reservoir to call on to meet future demand" (Hobfoll et al. 1998, p. 191).

Despite their differences, a common theme may be discerned among these and other (e.g., Masten 1994) stress theories: "They all share... a process in which environmental demands tax or exceed the adaptive capacity of an organism, resulting in psychological and biological changes that may place persons at risk for disease" (Cohen et al.

1995, p. 3). In short, theories converge on the idea of *overload*, which derives from the interplay of two constructs, (1) demands and (2) resources. Moreover, theories agree that these constructs must pair in a specific manner—high demands meeting low resources—for stress and subsequent illness to occur. Theoretically, other pairings should not produce stress, a conception consistent with a large literature that shows that enhanced resources ("hardiness", Kobasa et al. 1982; "resilience", Masten 1994; Rutter 1985) can render people resistant to the impact of demands.

Stress measures have been criticized for being "detached from theoretical underpinnings" (Derogatis and Fleming 1997, p. 114). To address this criticism, the optimal starting point might be to focus on the common ground shared by theories. First and foremost, stress measures should contain items specific to the experience of overload. However, with few exceptions (e.g., "difficulties piling up so high that you could not overcome them", Cohen et al. 1983), most current items assess either symptoms (e.g., "changes in appetite", Levenstein et al. 1993) or stressors (e.g., "death of a spouse", Holmes and Rahe 1967). Second, because the overload experience is built upon two underlying constructs, stress measures might well employ a two-scale structure. But most available measures consist of only one scale, focused on either demands (e.g., Holmes and Rahe 1967) or the person (e.g., Abell 1991) and not their interplay.

In a landmark assessment treatise, Loevinger (1957) outlined the process for creating a theory-consistent measure. These steps were followed here to construct a new questionnaire, consistent in both content and structure with the concept of overload shared by stress theories.

Psychometric Properties

To compare self-report stress measures, a useful taxonomy (Cohen et al. 1983) divides them into two broad categories: The *objective*, focused on environmental events (e.g., "Death of spouse"), and the *subjective*, focused on personal reactions (e.g., "How often have you felt nervous and 'stressed'"). This distinction is blurred by some measures that pair "objective" events with "subjective" ratings of their impact, called *hybrids* here. Examples of popular measures of all three types, along with their psychometric properties, are shown in Table 1.

In regards to internal consistency, subjective scales are generally superior. This is logical: Objective scales list life events, and there is little reason for such items to intercorrelate (the "death of a spouse" does not necessarily coincide with a "jail term"). On the other hand, subjective items should correlate ("nervous" and "upset" do

Table 1 Key characteristics of some popular self-report stress measures

Name	Format	Reliability	Validity
<i>Objective measures:</i>			
Social Readjustment Rating Scale (Holmes and Rahe 1967)	43 items, Checklist, e.g., “Death of a spouse”	$\alpha = .72-.80$ Test–retest: $r = .87-.90$ (one week) ^a $r = .64-.74$ (6–12 months) ^b	Those scoring in high range have greater likelihood of developing illness (80%) than those in moderate (50%) or low (30%) ranges
<i>Hybrid measures:</i>			
Daily Hassles (Kanner et al. 1981)	117 items, Checklist w/3-point “severity” ratings, e.g., “Concerns over weight”	Test–retest (1 month): Mean $r = .79$ (Hassles), Mean $r = .48$ (Severity)	$r = .21$ w/life events $r = .34$ w/negative affect $r = .50-.60$ w/physical symptoms $r = .30-.40$ w/self-rated health ^c
Daily Stress Inventory (Brantley and Jones 1989)	58 items, Checklist w/7-point “impact” ratings, e.g., “Heard some bad news”	$\alpha = .83$ (Events), $.87$ (Impact) ^d Test–retest (28 days) ^d : $r = .72$ for Events $r = .41$ for Impact	$r = 0$ (Events), $.36$ (Impact) w/anxiety $r = .33$ (Events), $.57$ (Impact) w/hassles Total scores correlate with pain, incidence of asthma and migraines ^c
Weekly Stress Inventory (Brantley et al. 1997)	87 items, Checklist w/7-point “stressfulness” ratings, e.g., “Argued with a co-worker”	$\alpha = .92-.96$ (Events), $\alpha = .93-.97$ (Stressfulness) Test–retest (1–3 weeks): Mean $r = .74$ for Events Mean $r = .63$ for Stressfulness	$r = .45-.53$ w/life events $r = .57$ (Evts), $.54$ (Strs) w/stress $r = .29$ (Evts), $.34$ (Strs) w/depression $r = .24$ (Evts), $.30$ (Strs) w/physical symptoms
<i>Subjective measures:</i>			
Derogatis Stress Profile (Derogatis 1984)	77 items, reflecting 3 theoretical dimensions: Environment, Emotional Responses, Personality Mediators. 5-point “true of me” ratings, e.g., “I get easily annoyed or irritated”	$\alpha = .83$ (Emotion), $\alpha = .85$ (Environment) $\alpha = .89$ (Personality) ^f Test–retest (7 days) ^f : $r = .83$ for Emotion $r = .83$ for Environment $r = .89$ for Personality $r = .90$ for Total Score	$r = .32$ w/hassles ^f $r = .48$ w/life events ^f $r = .56-.72$ w/subjective stress indices ^f
Global Inventory of Stress (Sheridan and Smith 1987)	22 items, reflecting 3 theoretical dimensions: Stressors, Resources, Strain 5-point frequency ratings, e.g., “...how frequently you are stressed lately”	$\alpha = .86$ ^g Test–retest (2 weeks) ^g : $r = .89$	$r = .36$ w/life events ^g $r = .56$ w/anxiety ^g $r = .54$ w/physical symptoms ^h $r = .47$ w/health status ^h
Index of Clinical Stress (Abell 1991)	25 items, 5-point “how much of the time” ratings, e.g., “I feel wound like a coiled spring”	$\alpha = .96$	$r = .24$ w/life events $r = .75$ w/contentment
Perceived Stress Scale (Cohen et al. 1983)	14 items, 5-point “how often” ratings, e.g., “how often have you felt nervous and ‘stressed’”	$\alpha = .84-.86$ Test–retest: $r = .85$ (2 days) $r = .55$ (6 weeks)	$r = .20$ w/life events $r = .35$ w/life events impact $r = .70$ w/depression $r = .61$ w/physical symptoms

Table 1 continued

Name	Format	Reliability	Validity
Recent Perceived Stress Questionnaire (Levenstein et al. 1993)	30 items, 4-point “how usual” ratings, e.g., “Changes in appetite”	$\alpha = .90$ Test–retest: $r = .84$	$r = .67$ w/life events $r = .56$ w/self-rated stress $r = .75$ w/trait anxiety $r = .56$ w/depression $r = .58$ w/physical symptoms

^a Lei and Skinner 1980, ^b Rahe et al. 1974, ^c DeLongis et al. 1982, ^d Brantley et al. 1987, ^e Brantley et al. 1998, ^f Derogatis and Fleming 1997, ^g Radmacher and Sheridan 1989, ^h Sheridan and Radmacher 1998

coincide). This difference may not be striking in Table 1, because comparisons are hampered by the fact many stress scales were not derived in the general population. The use of homogeneous samples (e.g., college students, Radmacher and Sheridan 1989, or patients, Abell 1991) affects response variance and ultimately inter-item correlations.

In terms of test–retest reliability, objective scales are generally superior, likely because it is easier to recall whether an event occurred than the intensity of the feelings associated with it. But here scale comparisons are complicated both by differences in normative samples and test–retest intervals (which range from as little as 1 h, Brantley et al. 1998, to as long as 12 months, Rahe et al. 1974). Still, by comparing subscales within the hybrid measures in Table 1 and thereby holding methods constant, objective (e.g., “Hassles” or “Events”) scores can be seen to be more stable than subjective ones (e.g., “Severity” or “Impact”).

In terms of validity, both subjective and objective scales have problems. The flaw in subjective measures is that people are often unwilling or unable to provide accurate information, due to social desirability pressures or to memory distortions (Stone 1995). The flaw in objective measures may be even more profound: By eliminating subjectivity altogether, they do not allow that a given event might have different impacts on different people (Dohrenwend 2006; Hough et al. 1976). Comparing the validity of measures is hampered not only by differences in norm groups but also in the validation criteria employed. If only common criteria are considered, Table 1 shows subjective scales to be superior, having stronger associations with negative affect, physical symptoms, self-rated health, and life-event tallies.

In sum, subjective stress scales hold an overall psychometric advantage, with generally better internal consistency and validity. And there are strong arguments in the literature against the use of objective formats for measuring stress (e.g., Lazarus and Folkman 1989). But it cannot be ignored that objective items, which are concrete and specific, have generally better test–retest reliability than subjective ones, which can be ambiguous and general (e.g., “How stressed do you feel today?”; Goldman et al. 1996).

Clearly, item format as well as item content must be considered in constructing a stress measure. Here, it was decided to use a subjective format in order to capitalize on its psychometric strengths, but an attempt was made to word the items as concretely as objective ones (e.g., “felt like you were carrying a heavy load”) in hopes of also maximizing score stability. In this manner, a large pool of potential items was written, of which only those that proved strongest were selected for the SOS.

Practical Considerations

Interview methods, such as the Life Events and Difficulty Schedule (LEDS; Brown and Harris 1978), have proven good at predicting stress-related illness (Brown and Harris 1989); but these methods have not come into general use due to their toll on both the researcher and respondent (Dohrenwend 2006). Self-report measures are less expensive for the researcher, but can still be quite taxing for the respondent. Objective measures are typically lengthier (Table 1), because they must cover a range of probable events; some (e.g., the 117-item Daily Hassles scale; Kanner et al. 1981) can require as much time to complete as an interview. Subjective scales are typically shorter, but not all are less onerous (e.g., the 77-item Stress Profile; Derogatis 1984). The burden imposed by such measures can be deleterious to community research: It can deter participation, affecting sample size and representativeness, and also diminish data quality, owing to respondent fatigue.

In response to such concerns, some researchers have gone to the extreme of advocating single-item stress measures (Littman et al. 2006). But single items have been shown to correlate poorly with full and psychometrically sound stress scales (Sagrestano et al. 2006). The present goal was to maintain validity (with enough items to adequately represent theoretical constructs) and reliability (with enough items for adequate internal consistency), but otherwise trim the item pool to the fewest possible items.

In addition, a closed-ended format was used to facilitate responding and scoring. In these ways, the SOS' burden on both respondent and researcher would be minimized.

Another practical consideration was the measure's comprehensibility across different cultures. To assure the broad applicability of the SOS, each in the progression of psychometric tests was conducted in a different, large and diverse general population sample. Only items that were consistently understood across this wide socioeconomic and ethnic spectrum (which included non-native English speakers) were chosen for the SOS.

The Current Research

With the goal of achieving a more accurate reading of stress, and thereby better predictability of stress-induced pathology, a new self-report measure was developed here. The SOS is the first stress measure constructed by entirely empirical methods, using community samples matched to U.S. Census demographic proportions. It is the end-point of a prescribed (Loevinger 1957) series of tests aimed at producing a theory-consistent and psychometrically strong but also practical measure.

Initially, a large pool of potential items was formed. These addressed the common denominator of stress theories by describing the state of overload (e.g., "overextended", "overcommitted") and its components, excessive demands (e.g., "swamped by your responsibilities") and personal vulnerability (e.g., "you couldn't cope"). They employed a subjective (e.g., "felt there was too much to do and too little time") rather than objective format (e.g., "impending deadline at work"), to attain the internal consistency and validity typical of subjective measures. But their wording was concrete and specific ("felt like you had to make quick decisions"), in hopes of capturing the test-retest stability of objective measures. Items that described physical symptoms ("sweaty palms") were excluded, to prevent spurious correlations with illness measures. And idiomatic items, potentially problematic for any population subgroup, were avoided to minimize cultural biases.

First, this item pool was subjected to multiple factor analyses (Phase 1) to determine the correspondence of its underlying structure to that implied by theory. The items that proved to best fit this structure were then tested for construct validity and reliability (Phase 2). Those that survived these psychometric tests, and also proved consistently comprehensible across the tested populations, formed the SOS. Lastly, this finalized measure was tested for its ability to actually predict health and health-related behavior following a stressful event (Phase 3).

Phase 1: Structural Analyses

Exploratory and confirmatory analyses, conducted in independent community samples, were used to determine the factor structure underlying overload items. Results were examined for their congruence to theory, and used to pare the item pool to only the best factor-markers.

Method

Participants

For the exploratory study, 600 community residents were recruited, of whom 435 agreed to participate and 431 (72%) completed all items. For the confirmatory study, another 600 people were solicited, of whom 433 (72%) completed the questionnaire. The quota-sampling method was successful in capturing the diversity of the region, with few departures from Census proportions (see Table 2).

Measure

Over 500 potential items were gathered from a variety of sources, including journals, existing scales, our own qualitative data archives, and student focus groups. Independent judges decided whether these reflected overload and also eliminated items that were redundant or symptoms of physical or mental illness (average pair-wise agreement = .96). The remaining 150 items were put into a uniform format ("In the past week, have you felt..."), paired with 5-point rating scales anchored at "Not at All" and "A Lot", and prefaced with instructions (which explained rating scales, encouraged honesty, and promised confidentiality). Demographic items were added at the end to avoid priming effects (Steele 1997). This prototype, entitled "S.O.S./A Measure of Day-to-Day Feelings", was used in the exploratory study; the SOS used in the confirmatory study was shorter, containing only 55 items that had proven good factor-markers.

Procedure

Samples for these (and all subsequent) studies were drawn from Southern California, which is a demographically diverse region. For a substantial portion, English is a second language (including immigrants from Mexico and Central America, as well as Cambodia, Vietnam and Korea). U. S. Census figures for the region were used to set sampling quotas. The quotas were used to create recruitment lists, each a random series of demographic profiles (e.g., "female, 18–24 years old, African-American"). Concentrating on sites where people were likely to be bored (e.g., laundromats, commuter trains, government agencies, shopping malls and public parks), researchers

Table 2 Demographic composition of study samples

	Phase 1		Phase 2		Phase 3	
	Exploratory FA	Confirm. CFA	Construct validity	Reliability	Criterion validity	Gen. pop. norms
Type	Quota	Quota	Quota	Quota	Convenience	Census
Size (<i>n</i>)	431	433	310	344	285	
Gender						
Male	177 (41%)	161 (38%)	155 (50%)	159 (46%)	146 (51%)	49%
Female	252 (59%)	270 (62%)	152 (50%)	185 (54%)	139 (49%)	51%
Age						
18–24 years	96 (22%)	60 (14%)	55 (18%)	86 (24%)	51 (18%)	14%
25–44 years	197 (46%)	204 (48%)	154 (50%)	136 (41%)	140 (49%)	45%
45–64 years	107 (25%)	116 (27%)	82 (26%)	102 (29%)	81 (28%)	27%
>65 years	31 (7%)	46 (11%)	19 (6%)	19 (6%)	13 (5%)	14%
Ethnicity						
African-Am.	26 (6%)	36 (8%)	25 (8%)	35 (10%)	22 (8%)	2%
Asian-Am.	66 (15%)	58 (14%)	41 (13%)	50 (15%)	40 (14%)	13%
Hispanic-Am.	54 (13%)	99 (23%)	64 (21%)	86 (25%)	55 (19%)	28%
Caucasian	251 (59%)	200 (47%)	154 (50%)	155 (45%)	143 (50%)	56%
Other	32 (7%)	36 (8%)	26 (8%)	18 (5%)	25 (9%)	1%
Education						
High sch.	91 (21%)	98 (23%)	60 (14%)	71 (21%)	54 (19%)	
Some coll.	180 (42%)	205 (48%)	142 (46%)	150 (44%)	133 (47%)	
Coll. deg.	109 (26%)	92 (21%)	65 (21%)	80 (23%)	59 (21%)	
Grad deg.	48 (11%)	36 (8%)	39 (13%)	43 (12%)	38 (13%)	
Income						
<\$25 K	105 (25%)	125 (30%)	73 (24%)	82 (24%)	63 (22%)	
\$25–39 K	106 (26%)	88 (22%)	66 (21%)	71 (21%)	62 (22%)	
\$40–\$59 K	155 (35%)	149 (36%)	122 (39%)	65 (19%)	55 (19%)	
>\$60 K	57 (14%)	48 (12%)	42 (14%)	125 (36%)	100 (35%)	

Census figures are averaged 2008 estimates for portions of Los Angeles and Orange Counties, CA, and for persons over 18 years old only. "Asian-Am." includes Pacific Islanders. "High Sch." signifies high-school degree or less

used the lists to recruit people who appeared to fit the desired profile. If refused, they tried twice again before moving on to the next profile in the series. If accepted, they ascertained whether the person was over 18 years old and spoke English, but did not try to verify the profile. After informed consent procedures, they instructed participants to complete the SOS on site, and to critique-or simply cross out- any items found to be difficult. If at any point a participant asked for clarification, the assistant made note of the point of confusion. Participants dropped completed protocols into locked collection boxes to ensure anonymity of response.

Results

Exploratory Analyses

Comments by participants and researchers were used to identify items that were ambiguous, incomprehensible, or

culturally insensitive. For example, "I felt as if I was just spinning my wheels" was found to be problematic for respondents who were not native English-speakers. Items that received at least two negative comments were excluded from data analysis, leaving a total of 80 of the original 150 items.

Using the principal-factors extraction method, two factors were found to account for most of the variance (eigenvalues of 31.342 and 5.086 before rotation). Scree analysis showed eigenvalues to level at three factors; but closer inspection of the third factor showed it to be a measurement artifact, comprised solely of the reverse-keyed items. Therefore it was decided that a two-factor solution was optimal. Using Varimax, the two factors were rotated to an orthogonal solution, which proved less than adequate: Approximately 75% of the items had high loadings on both factors, and there was substantial inter-factor covariance. An oblique solution appeared more consistent with the data, hence Promax was used to

re-position the factors. This yielded eigenvalues of 19.188 and 17.240, and an inter-factor correlation of $r = .56$ ($p < .0001$). With the new solution, 55 of the 80 items demonstrated high ($>.5$) loadings on one factor and low ($<.2$) loadings on the other, providing a clear sense of each factor's meaning: Factor I reflected feelings of powerlessness, inadequacy, frailty and debility, collectively labeled "Personal Vulnerability"; Factor II included perceptions of being burdened by outside demands, responsibilities and pressures, and was labeled "Event Load".

Confirmatory Analyses

Comments from the second sample helped identify more problematic items; five were criticized twice or more, leaving 50 for the following analyses.

Five models were generated for confirmatory analyses. Beyond dictating the general structure (e.g., One-Factor, Three-Factor-Oblique, etc.), the models designated which items were to load on each factor. Owing to multivariate non-normality (normalized Mardia's coefficient = 101.87), all models were tested using polychoric correlations with ROBUST statistics. Four indices were employed to evaluate overall model fit (Sattorra-Bentler χ^2 ; CFI; RMSEA; SRMR); for relative fit, $\Delta S-B\chi^2$ was used for nested models and the AIC (Tanaka 1993) for non-nested ones.

Results showed the One-Factor-Model to fit poorly according to all indices. The Two-Factor-Orthogonal model was plausible according only to two indices (CFI = .90, RMSEA = .03). The Two-Factor-Oblique model fit well according to three indices (CFI = .93, RMSEA = .03, SRMR = .05), and significantly better than the Two-Factor-Orthogonal model [$\Delta S-B\chi^2(1) = 188.40$, $p < .001$]. The Three-Factor-Orthogonal model was plausible according only to the one index (RMSEA = .04) and it fit significantly worse than the Two-Factor-Oblique model (AIC = 518.76 vs. -510.66). The Three-Factor-Oblique model fit well by most indices (CFI = .90, RMSEA = .03, SRMR = .07), but not as well as the Two-Factor-Oblique model (AIC = -289.96 vs. -510.66).

Therefore, based on both fit indices and model comparisons, the Two-Factor-Oblique was deemed the best-fitting model. All standardized factor loadings were found to be large (from .48 to .84) and statistically significant, as was the inter-factor correlation ($r = .60$, $p < .0001$).

Cultural Equivalence Testing

Multi-group analyses were performed to test the equivalence of the Two-Factor-Oblique model across Asians, Hispanics, and Caucasians (the ethnic groups large enough to permit analysis). The metric invariance model was found

to fit reasonably well in comparing Asians to Caucasians and Hispanics to Caucasians and (CFI = 1.00, RMSEA = .01, SRMR = .08, in both comparisons). Making the model more stringent by constraining factor variances and covariance to equality across groups did not alter the results. This restrictive model fit well (CFI = 1.00, RMSEA = .01, SRMR = .08, in both comparisons), and did not significantly differ from the metric invariance model. In sum, the factor loadings, variances, and covariances of the Two-Factor-Oblique model were equivalent across the three ethnic groups.

Discussion

Structural analyses of items judged to reflect overload revealed two underlying factors: Event Load, a sense that life's demands are burgeoning, and Personal Vulnerability, a sense of susceptibility to those demands. And, although distinct, these factors were found to correlate. This model clearly corresponds, in both content and structure, to a conceptualization shared by stress theories—that stress is the overload experienced at the juncture of too many stressors and too few resources. Moreover, the model's universality was demonstrated across diverse population groups, the very ones most likely to differ owing to distinct linguistic and cultural heritages.

Phase 2: Psychometric Tests

Having handpicked overload items that best reflected the underlying constructs of Personal Vulnerability and Event Load, the next step was to determine their psychometric properties.

Construct validity tests examined the relationship between SOS items and three types of validation indices: (1) Personal resistance to stress (Hardiness and Mastery); (2) environmental demands (Life Events and Daily Hassles); and (3) standards in the validation of self-report measures in general (Social Desirability), and stress measures in particular (symptomology). It was predicted that Personal Vulnerability items would relate better to the first set of measures, Event Load items to the second, and that all items would relate to symptom but not social desirability scores.

A separate set of tests determined if the subjective format and concrete wording of SOS items had been successful in achieving both internal and external reliability. Because existing measures vary widely in the test-retest interval used, not one but three different time periods were used to estimate the external reliability of the SOS.

Results from the two psychometric studies were used to pare the SOS item pool to its final size.

Methods

Participants

The quota-sampling method previously described was again employed. For the construct validity study, 500 people were recruited in public settings, of whom 310 (62%) volunteered and completed all protocols. For the reliability study, 500 community residents were approached, 403 of whom agreed to participate and 344 (69%) completed both the test and retest protocols. These were lower than previous compliance rates, possibly due to the greater respondent burden in these studies. Nevertheless, both samples were demographically variegated, in proportions reasonably close to those reported by the Census (see Table 2).

Measures

An SOS consisting of the 50 items surviving from Phase 1 was used in both studies.

For the validity study, a number of other measures were also utilized. As indices of Personal Vulnerability, measures of antithetical constructs were used: (1) Hardiness was assessed by the 45-item Dispositional Resilience Scale (Bartone et al. 1989), which has good reliability (mean Cronbach's $\alpha = .93$; Kosaka 1996) and validity (e.g., negative associations with distress and illness; Kosaka 1996; Ouellette 1993; Roth et al. 1989); (2) Control was measured by means of the Mastery Scale (Pearlin et al. 1981) which, although brief (seven items), has good reliability (demonstrated in LISREL analyses) and validity (negative correlations with strain and depression).

For Event Load, life-events checklists were used: (1) Major events were indexed by the Social Readjustment Rating Scale (Holmes and Rahe 1967), which lists 43 stressors weighted for intensity, and has been shown to be generally reliable and valid (see Table 1; Maddi et al. 1987); (2) minor events were counted by means of the Daily Hassles Scale (Kanner et al. 1981; Lazarus 1985), which lists 117 small but chronic stressors, and yields frequency scores that have shown good test–retest reliability and some validity in predicting ill health (see Table 1).

For the SOS as a whole, commonly used validation indices were employed: (1) Depression was assessed with the Center for Epidemiological Studies Depression scale (CES-D; Radloff 1977), which contains 20 non-clinical symptoms and has shown internal consistency ($\alpha > .80$) and validity; (2) General Illness was evaluated by the General Health Questionnaire (GHQ; Goldberg 1972), which lists a broader spectrum of 30 psychiatric and somatic symptoms, and has demonstrated split-half ($r = .94$) and test–retest ($r = .75$) reliability, as well as validity in predicting clinical

status (Cleary et al. 1982); (3) Social Desirability was measured by the Marlow–Crowne scale (Crowne and Marlowe 1964), whose 33 items focus on the bias most likely to affect stress measures (“impression management”, Paulhaus 1988), and have shown good internal consistency ($K-R 20 = .88$), test–retest reliability ($r = .88$), and validity (Crowne and Marlowe 1964).

Procedure

For the validity study, participants completed the SOS on site, and then received a pre-paid mailer containing all of the validation indices and instructions to complete these 24 h later, mark them with the date, and mail them back within a week. For the reliability study, participants filled out one SOS on site, and then were told to complete a second either one day, four days, or one week later (intervals randomly assigned), dating and returning it via pre-paid mailers.

Results

Construct Validity Tests

Zero-order correlations showed all SOS scores to correlate significantly with all of the validation measures. However, further analysis did reveal differential effects for the Personal Vulnerability and Event Load scales, even though they themselves inter-correlated ($r = .53$). First, *t* tests for the difference between dependent correlations (McNemar 1975) showed significant disparity in the magnitude of most validity coefficients obtained for each scale, in the expected directions. Second, partial correlations (controlling for the influence of the other scale) also produced close to the expected pattern of relationships. In both analyses, Personal Vulnerability proved to be more strongly related to Hardiness and Control, and Event Load more strongly related to major—although not minor—life events (see Table 3).

SOS total scores related as predicted to the standard validation indices, showing significant relationships with Depression and General Illness symptoms, and not with Social Desirability.

Relative Validity Tests

Because some of the most popular of extant stress measures were among the validation indices, it was possible to compare their strength relative to the SOS in terms of predicting pathology. Table 4 shows that the SOS yielded higher correlations with depressive and general symptoms than did the Hardiness, Life Events or Daily Hassles measures—despite that it was the only questionnaire

Table 3 Convergent and discriminant construct validity results

Validity construct	SOS			Difference in PV versus EL correlations <i>t</i> (<i>df</i>)
	Full SOS	PV scale	EL scale	
<i>For the PV scale:</i>				
Hardiness	-.31****	-.40****	-.17**	4.62 (307)**
95% CI	[-.41, -.21]	[-.49, -.30]	[-.28, -.06]	
Partial <i>r</i>		-.36****	-.11	
95% CI		[-.45, -.26]	[-.22, .00]	
Mastery	-.42****	-.56****	-.21**	8.14 (307)***
95% CI	[-.50, -.32]	[-.63, -.48]	[-.31, -.10]	
Partial <i>r</i>		-.46****	.08	
95% CI		[-.54, -.37]	[-.03, .19]	
<i>For the EL scale:</i>				
Life events	.21**	.17**	.29****	2.47 (310)*
95% CI	[.10, .31]	[.06, .28]	[.18, .39]	
Partial <i>r</i>		.09	.27****	
95% CI		[-.02, .20]	[.16, .37]	
Daily hassles	.44****	.38****	.39****	.41 (308)
95% CI	[.35, .53]	[.28, .47]	[.29, .48]	
Partial <i>r</i>		.15*	.19**	
95% CI		[.04, .26]	[.08, .30]	
<i>For the full SOS:</i>				
Social desirability	-.12	-.10	-.13	.44 (307)
95% CI	[-.23, -.01]	[-.21, .01]	[-.24, -.02]	
Partial <i>r</i>		-.08	-.10	
95% CI		[-.19, .03]	[-.21, .01]	
Depression	.53****	.52****	.46****	1.54 (307)
95% CI	[.45, .61]	[.43, .60]	[.37, .54]	
Partial <i>r</i>		.31****	.16**	
95% CI		[.21, .41]	[.05, .27]	
General illness	.75****	.74****	.69****	1.73 (307)
95% CI	[.70, .79]	[.69, .79]	[.63, .74]	
Partial <i>r</i>		.49****	.16**	
95% CI		[.40, .57]	[.05, .27]	

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

administered at a separate time and place (and therefore less likely to benefit from the response bleeding that can inflate correlations). In addition, of the stress measures, only the SOS was unrelated to Social Desirability bias.

To determine if the SOS' advantage rested solely in its structure (in that it assessed two stress-related constructs while the comparison measures assessed only one), a bi-dimensional "Ersatz SOS" was constructed from the Hardiness (as a proxy for Personal Vulnerability) and Life Events (as a proxy for Event Load) scales. Despite being longer (88 items), this ersatz measure did not demonstrate commensurate validity: Differences in correlation coefficients showed it to be inferior to the SOS in predicting Depression ($t = 6.95$, $df = 307$, $p < .001$) and General

Illness ($t = 9.82$, $df = 307$, $p < .001$). Moreover, it was not free of social desirability biases (see Table 4).

Reliability Tests

For these analyses, participants were divided into three groups according to the assigned test–retest interval (One Day, Four Days, One Week). Rates of return were comparable across the groups ($ns = 114$, 113, and 115, respectively), and chi-square tests showed the groups to be demographically equivalent in terms of gender, age and ethnicity.

The SOS was pared to its final length during these tests. The un-attenuated factor loadings from Phase 1, and the convergent and discriminant correlations from the

Table 4 Comparison of construct validity for the SOS versus other measures

Predictor	Social desirability	Depression (CES-D)	General illness (GHQ)
Hardiness	.13*	-.28***	-.36****
95% CI	[.02, .24]	[-.38, -.17]	[-.45, -.26]
Mastery	-.09	-.37****	-.47****
95% CI	[-.20, .02]	[-.46, -.27]	[-.55, -.38]
Life Events	-.15*	.30****	.18**
95% CI	[-.26, -.04]	[.20, .40]	[.07, .29]
Daily Hassles	-.21**	.43****	.32****
95% CI	[-.31, -.10]	[.33, .52]	[.21, .42]
Ersatz SOS	-.19**	.38****	.33****
95% CI	[-.30, -.08]	[.28, .47]	[.23, .43]
SOS (Total Score)	-.12	.53****	.75****
95% CI	[-.23, -.01]	[.45, .61]	[.70, .79]

The “Ersatz SOS” is a combination of the Hardiness and Life Events measures

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

construct validity study, were examined for each of the 50 items. In addition, preliminary analyses of the reliability data yielded item-total and test-retest coefficients (averaged across the three groups) for all 50 items. With an eye towards selecting the best across these five criteria, and an equal number (≥ 10) for each subscale, 24 items were chosen (12 representing each Personal Vulnerability and Event Load). All results presented below derive from this final, 24-item pool.

The internal consistency of these final items was determined using data from the retest SOS, which was completed under conditions more typical of self-report assessments (no researcher present, no distractions of a public venue). Item-total correlations showed the 12 Personal Vulnerability items to relate well to SOS total scores (mean $r = .63$, $SD = .11$, $R = .40-.76$), as did the 12 Event Load items (mean $r = .66$, $SD = .11$, $R = .39-.79$). They also showed more consistency within than across scales: Personal Vulnerability items had stronger correlations with their own adjusted total than with the Event Load total, and vice versa. Cronbach alpha coefficients, averaged across groups, also revealed good internal consistency: .95 for the SOS as a whole, .93 for Personal Vulnerability and .92 for Event Load scales.

In terms of external reliability, test-retest correlations showed shorter time intervals to produce greater score stability. For the One Day group, the correlations were .80 for the whole SOS, .82 for Personal Vulnerability and .79 for Event Load, with 95% CIs of [.72, .86], [.75, .87], and [.71, .85], respectively. For the Four Day group, the corresponding figures were .77, .76 and .78, with 95% CIs of

[.68, .84], [.67, .83], and [.70, .84]. And for the One Week group, the figures were .72, .74 and .70 with 95% CIs of [.62, .80], [.64, .81], and [.60, .78]. Averaged across time intervals, the test-retest correlation of .76 for the full SOS indicated adequate reliability.

Discussion

Validity tests showed the two SOS factor scales, by virtue of covariation with measures of similar constructs, to be aptly named: Personal Vulnerability items were inversely related to indices of stress resistance, and Event Load items were related to tallies of major and minor life events. One discriminant failure was that both scales correlated with Daily Hassles, but this might be explained by the presence of Vulnerability-like (“Concerns about meeting high standards”) as well as Event-like (“Traffic”) items on that checklist. Importantly, the SOS as a whole exhibited strong links to pathology, even though symptoms had been avoided in its item pool. In fact, it surpassed existing stress measures in this regard, both in current comparisons and vis-à-vis published coefficients.

One cautionary note is that construct validity tests evaluate a new measure by comparing it to peers, which likely have their own measurement problems. The Hardiness scale, for example, has been criticized for its psychometric failings (Funk and Houston 1987), so convergence with this measure may constitute more of a condemnation than an endorsement of the SOS. Also, because validity coefficients are correlations, the third-variable problem lurks. Unmeasured person factors (with negative affectivity a particular concern; Costa and McCrae 1987; Watson and Pennebaker 1989) might have biased responses to both the SOS and symptom scales, inflating their inter-correlation. Precautions taken here may not have been wholly effective in curtailing such factors.

Reliability tests provided additional criteria for paring the item pool to its final size, and showed that a shorter SOS could be reliable. The choice of a subjective format was successful in producing good internal consistency. In fact, coefficients for the SOS exceed those of most other subjective measures (see Table 1). But the strategy of “concretizing” the subjective items to maximize score stability was only partially successful. Test-retest reliability for the SOS exceeded the average for the subjective scales and sub-scales, but was less than that for the objective ones, shown in Table 1.

Phase 3: Tests of the Finalized SOS

Put into its final format, the SOS was tested for criterion validity—its ability to identify those persons most

overloaded by a common stressor, and most likely to experience symptoms of ill health in its wake.

To minimize aforementioned problems in validity tests, a journal technique rather than a standardized criterion measure was used. Participants recorded physical complaints (to avoid overlap between stress items and psychiatric symptoms), for an extended period (to minimize response bleeding), in their own words (to minimize negative checklist-response sets), and on a daily basis (to minimize recall problems). In addition, a longitudinal design and baseline controls were employed to further reduce possibilities of third-variable influences.

Several popular stress measures were subjected to the same test, allowing the validity of the finalized SOS to be evaluated on a relative basis again.

Methods

Participants

For the criterion validity test, a convenience sample was drawn from people filing last-minute tax returns at local post offices. To compensate for an onerous respondent load, state lottery tickets were used as incentives. Of 409 who enrolled, 285 (70%) completed the study. Even though quota sampling was not used, the sample proved as diverse as previous ones (see Table 2).

Measures

The SOS' format was finalized for this study (see Appendix). The 24 surviving items were arranged in a non-random order that balanced Personal Vulnerability and Event Load items. The format, instructions, and ambiguous title previously described were retained. Because only one reverse-keyed item had survived, six positive filler items (e.g., "generous") were added to offset the generally negative tone (and help disrupt negative response sets).

For tests of relative validity, three popular stress measures—one objective, one hybrid, and one subjective—were administered. The Social Readjustment Rating Scale (Holmes and Rahe 1967) was the objective measure, providing a weighted sum of Life Events. Daily Hassles (Kanner et al. 1981) was the hybrid measure, giving an objective event tally weighted by subjective severity ratings. The purely subjective measure was the Perceived Stress Scale (Cohen et al. 1983), chosen because it is fairly consistent with theory (70% "overload" items), practical (14 items), and as reliable and valid as its peers (see Table 1).

For baseline measures, the 30-item GHQ was used to assess General Illness, and the Strain-Free Negative

Affectivity-Revised scale (Fortunato and Goldblatt 2002) to assess negativity free of strain or distress. Its 20 items have good internal consistency (mean $\alpha = .87$) and construct validity.

For the criterion measure, participants were supplied with a "Health Log" (40 blank sheets with column headings of "Date", "Symptoms", "Health-Related Visits", and "Missed Work/School"). Before analyses, judges eliminated irrelevant and ambiguous entries and then summed each column, with an average pair-wise agreement of .98.

Procedure

Beginning at noon on the April 15 deadline, research assistants manned tables in front of 10 post offices that remained open until midnight for last-minute tax returns. Above the tables, banners read "STRESSED OUT BY TAXES? Participate in a study for a chance to win your money back". Those approaching the table who met selection criteria (over 18 years old, English-literate, and filing a tax return), were enrolled. Enrollment stopped on April 16, 1:00 am.

Wave 1 assessments took place on site: Participants filled out the four stress measures (in counterbalanced orders), and received one state lottery ticket. Those whose responses were complete ($n = 387$) were mailed the Wave 2 protocol and a second lottery ticket within 24 h. These baseline measures were to be returned within 48 h via pre-paid mailers. Participants who actually returned them within 1 week ($n = 352$) were mailed Wave 3 materials: The Health Log and three more lottery tickets. Instructions were to maintain the Log for 1 month, recording daily physical symptoms (examples provided), visits to health providers (traditional or alternative), or sick time from work or school. Respondents were reminded weekly by phone to keep Logs current; those who mailed back completed Logs within 5 weeks constituted the final sample ($n = 285$).

In an embedded study, the first 100 who returned Wave 2 protocols were asked to volunteer to complete a second SOS exactly 1 week following the first. This request, an additional lottery ticket, and an SOS were included in their Wave 3 materials. Of the 100, 77 complied.

Results

Scale Characteristics

Confirmatory factor analysis verified the Two-Oblique-Factor structure of the final SOS (CFI = .99; RMSEA = .01; SRMR = .06). Standardized factor loadings were all statistically significant (ranging from .59 to

.81). The inter-factor correlation was similar in magnitude to that previously found ($r = .48$).

The embedded study showed the final SOS' reliability to approximate previous figures. Test–retest coefficients over 1 week were .75 for the SOS as a whole, and .77 for Event Load and .73 for Personal Vulnerability scales, with corresponding 95% CIs of [.69, .80], [.72, .81], and [.67, .78]. Alpha coefficients were .96 for the SOS, and .94 for each scale.

Zero-order correlations showed all stress measures to co-vary significantly with the SOS, confirming its construct validity. Life Events tallies related to SOS total ($r = .17$, $p < .01$) and Event Load ($r = .16$, $p < .01$), but not Personal Vulnerability scores ($r = .10$, n.s.). Hassles related to SOS total ($r = .42$, $p < .0001$) and both subscale scores (PV: $r = .38$; EL: $r = .36$; $ps < .0001$), and PSS scores showed the same pattern (SOS total: $r = .53$; PV: $r = .44$; EL: $r = .53$; all $ps < .0001$).

Criterion Validity Tests

Significant zero-order correlations were found between all four stress measures and baseline illness (GHQ). Hybrid and subjective scales also correlated with negative affectivity (SFNA-R), indicating that third-variable concerns might be warranted (see Table 5).

Partial correlations were then used to control for both baseline GHQ and SFNA-R scores in determining the relationships between the stress measures and the Health Log criteria. These showed the SOS to be the only stress

scale that predicted all three illness indices. The SOS and PSS both predicted Symptoms, but the SOS' coefficient was significantly larger ($t = 5.92$, $df = 282$, $p < .001$). All four measures predicted Sick Days, but the SOS' coefficient was again greater than that obtained by the PSS ($t = 3.76$, $df = 282$, $p < .001$), Daily Hassles ($t = 3.75$, $df = 282$, $p < .001$), or Life Events ($t = 1.96$, $df = 282$, $p = .05$) scales. And the SOS alone predicted Practitioner Visits.

Owing to its two-scale structure, the SOS offers the possibility of categorical as well as continuous scoring. That is, by splitting Personal Vulnerability and Event Load scales into high versus low, four categories can be formed to separate those most at risk (in the high-high group) from others. Here, the validity of this scoring option was tested, using scale means to divide the sample into a 2×2 factorial. General Linear Model ANOVAs conducted on this factorial revealed significant main effects for Personal Vulnerability, with the high group exhibiting more Symptoms [$F(1, 284) = 68.14$, $p < .0001$], Missed Days [$F(1, 281) = 51.81$, $p < .0001$] and Practitioner Visits [$F(1, 281) = 10.48$, $p < .001$]. There were also significant main effects for Event Load, with the high group reporting more Symptoms [$F(1, 284) = 5.11$, $p < .03$] and Missed Days [$F(1, 281) = 11.48$, $p < .001$] but not Visits. There were no significant interactions. Most importantly, simple effects tests showed that those in the high-high cell suffered worse health than those in any of the other three (see Table 6).

Table 5 Comparison of predictive validity for the finalized SOS versus other measures

Predictor	Baseline		From health log		
	General illness	Negative affectivity	Physical symptoms	Missed days	Practitioner visits
Life Events	.15*	.06	.00	.19***	-.07
95% CI	[.03, .26]	[-.05, .17]	[-.11, .11]	[.08, .30]	[-.18, .05]
Daily Hassles	.31****	.28****	.07	.27****	-.06
95% CI	[.20, .41]	[.17, .38]	[-.05, .18]	[.16, .37]	[-.16, .06]
Perceived Stress	.38****	.40****	.17**	.29****	.06
95% CI	[.28, .47]	[.30, .49]	[.05, .28]	[.18, .39]	[-.05, .17]
SOS (Total Score)	.70****	.31****	.47****	.48****	.20***
95% CI	[.64, .75]	[.20, .41]	[.37, .56]	[.39, .56]	[.09, .31]
SOS PV scale	.66****	.32****	.53****	.50****	.16**
95% CI	[.59, .72]	[.21, .42]	[.44, .61]	[.41, .58]	[.04, .27]
SOS EL scale	.48****	.25***	.31****	.35****	.21***
95% CI	[.39, .56]	[.14, .36]	[.20, .41]	[.24, .45]	[.10, .32]

Stress-to-baseline (first two columns) correlations are zero-order; all others are partial correlations controlling for baseline illness and negative affectivity

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

Table 6 Means, (Standard Deviations), and [95% CIs] of health indices by SOS categorical group

Health index	SOS EL scale	SOS PV scale		
		Low	High	
Physical symptoms	Low	9.28 ^a (5.02) [8.35, 10.22]	11.28 ^b (7.60) [8.49, 14.07]	
		High	10.84 ^b (5.44) [9.29, 12.39]	16.62 ^c (7.97) [14.92, 18.28]
	Missed days		Low	1.84 ^a (.42) [1.76, 1.92]
		High		1.96 ^a (.35) [1.86, 2.06]
Practitioner visits	Low		0.58 ^a (.56) [.48, .68]	0.75 ^b (.51) [.56, .94]
		High	0.64 ^a (.53) [.49, .79]	0.82 ^c (.47) [.72, .92]

Within each health variable, means with different superscripts differ at least at $p < .05$; *ns* for PV-EL cells are 113 (Low–Low), 50 (Low–High), 31 (High–Low), and 91 (High–High)

General Discussion

“In the future, it will be important to pay careful attention to measurement problems that have made the results from many previous studies on stress and pathology difficult to interpret” (Skodol et al. 1990, p. 17). To this end, a new stress measure was constructed, different in content, scale structure and construction from current stress scales. Of 180 items judged to reflect “overload”, a concept shared by stress theories, only 24 emerged at the end of a sequence of large community studies. These were items that had proven themselves in terms of conformity with a theoretical two-factor structure, comprehensibility across a wide demographic spectrum, construct validity and reliability. Arranged into a finalized SOS, these items demonstrated their collective ability to predict who would become sick in the aftermath of a shared stressful experience. People with higher SOS scores, by either a continuous or categorical scoring method, were more likely to develop physical symptoms, visit health professionals, and miss work or school. The SOS proved better at identifying at-risk individuals than current stress measures of various formats,

even after controlling for a pernicious third-variable confound in stress-illness calculations.

This predictive ability, as well as its practicality and applicability across a broad demographic spectrum, make the SOS well suited to community health work. Identification of population groups at risk for stress-related pathology can be done quickly by comparing group means on total scores. More detailed comparisons can be achieved by utilizing the SOS’ categorical scoring. Splitting and crossing Personal Vulnerability (PV) and Event Load (EL) scales forms a four-cell diagnostic matrix, which can be used to determine the relative proportions of group members in the high-risk versus other cells.

Specific potential applications of the SOS include epidemiological research. In one large population study (Nielsen et al. 2008), an improvised two-item stress measure of unknown reliability and validity was found to predict cause-of-death differentially across gender and age groups. The SOS might have afforded a more accurate and detailed picture of the stress-mortality link, because of its psychometric strength and because it yields a wider range of total and sub-scale scores upon which to plot fatalities. Moreover, given the SOS’ broad demographic applicability, analyses could have been expanded to other population groups with confidence.

In etiological studies, the SOS would be helpful in discerning stress from other psychosocial causes of disease. For example, some have related “minority stress” to an increased prevalence of physical and mental disorders in gay men (Hatzenbuehler et al. 2008; Meyer 1995). Stress levels were not measured directly in these studies, but rather inferred from experiences with rejection, discrimination, and homophobia. To identify the true cause of health disparities between this and the majority population, stress would have to be differentiated from other negative emotions (such as hostility, a factor in coronary disease; or helplessness, a precursor of depression). The SOS, which was able to predict illness after partialling out negative affectivity, could isolate the effects of stress after controlling for competing emotions.

Research exploring individual differences in susceptibility to stress would also benefit from use of the SOS. First, it is unknown whether early exposure to trauma makes a person more or less vulnerable to later stressors (Rutter 1985; Turner and Lloyd 1995). Longitudinal analyses using SOS subscales could resolve this issue, by showing whether peaks in EL scores were followed by subsequent rises or dips in PV scores. Second, although myriad factors (genetic, dispositional, biographical) have been said to predispose people towards stress resilience, few have been empirically verified (Masten 1994). Categorical SOS scores would identify resilient (low-low cell) and vulnerable (high-high cell) individuals, and then

discriminant analysis could be used to find which of the proffered factors actually differentiated the cells. Third, it has been argued that resilience is not a stable personality trait, but varies with time and context (Rutter 1985). The PV scale, which captures momentary perceptions of vulnerability, would be more able to detect such fluctuations than dispositional measures such as the hardiness scale.

As an individual differences measure, the SOS would also aid community prevention and intervention efforts. Even within targeted groups, like the unemployed, such programs are not universally effective, but tend to help high-risk and not low-risk persons (Vinokur et al. 1995). The SOS could be used to triage those in more dire need of, and more likely to benefit from, such efforts. For groups exposed to environmental insults, such as community violence (Wilson et al. 2005), differences in PV scores might be particularly useful in identifying those most susceptible to the pathogens. In already vulnerable groups, such as persons recovering from addiction (Sinha 2008), variations in EL scores would be helpful in pinpointing those in danger of relapse.

In evaluating community programs, repeated stress assessments are often used to gauge effectiveness and the longevity of benefits (e.g., Raeburn et al. 1993). The SOS would be more sensitive to short-term changes in stress than objective measures (which only reflect shifts in external circumstances) and more specific to such changes than other subjective measures (which, being more prone to response biases, obfuscate true change with measurement error). In addition, SOS subscales would be useful in determining if programs had achieved their intended goals. Some, like anti-bullying campaigns, seek to reduce the “piling-up of multiple stressors” (Masten 1994); others,

like that prescribed for individuals with an addiction, seek to enhance self-efficacy (Sinha 2008). Significant drops in EL and PV scores, respectively, would verify that these programs had met their targets.

But there are provisos to the use of the SOS: First, while current samples were matched to regional demographics, these proportions do not reflect national averages. The comprehensibility and validity of SOS items should be verified in the groups under-represented here, such as African-Americans. Second, of the multiple studies conducted, only the last tested the measure in its final format. Further tests are being conducted, both to confirm psychometric strength and to determine optimal cut-off points for categorical scores in the general population. Third, the usefulness of the SOS in predicting health outcomes following single major traumas (tsunamis, etc.) has yet to be demonstrated. And fourth, while subjective wording was deliberately chosen for SOS items, such scales are prone to third-variable confounds. Whether the SOS is vulnerable to nuisance variables other than negative affectivity, and the extent of these influences, must still be determined.

Despite these uncertainties, the SOS emerged from the present studies as a theory-consistent, psychometrically viable, practical and diversity-friendly measure of stress. It will hopefully prove a useful tool for community health researchers and practitioners alike.

Acknowledgments The author wishes to thank California State University Long Beach and the Psychology Department for making possible this five-year endeavor. Scott Roesch is gratefully acknowledged for his statistical assistance, Chi-Ah Chun and Bianca Wilson for their helpful critiques. Requests for the formatted SOS and its scoring rubric (available at no charge for research purposes) should be directed to the author at jamirkha@csulb.edu.

Appendix Final SOS items

SOS item	Item type	Factor loading ^a	
		EL	PV
<i>IN THE PAST WEEK, have you felt:</i>			
1...calm?	Filler		
2...strained?	Event Load	.393	.621
3...inadequate?	Personal Vulnerability	.770	.326
4...overextended?	Event Load	.318	.662
5...confident? (Reverse Key)	Personal Vulnerability	-.578	.040
6...bored?	Filler		
7...no sense of getting ahead?	Personal Vulnerability	.756	.181
8...swamped by your responsibilities?	Event Load	.279	.740
9...that the odds were against you?	Personal Vulnerability	.796	.181
10...that there wasn't enough time to get to everything?	Event Load	.218	.718
11...generous?	Filler		
12...like you were rushed?	Event Load	.021	.784

Appendix continued

SOS item	Item type	Factor loading ^a	
		EL	PV
13...like you couldn't cope?	Personal Vulnerability	.786	.274
14...like you had a lot on your mind?	Event Load	.286	.740
15...like nothing was going right?	Personal Debility	.737	.213
16...carefree?	Filler		
17...powerless?	Personal Vulnerability	.708	.228
18...overcommitted?	Event Load	.049	.539
19...like your life was "out of control"?	Personal Vulnerability	.662	.188
20...like things kept piling up?	Event Load	.105	.759
21...like you could focus on the important things?	Filler		
22...like you had to make quick decisions?	Event Load	.314	.731
23...like asking "what else can go wrong?"	Personal Vulnerability	.768	.150
24...like you didn't have time to breathe?	Event Load	.385	.678
25...like things couldn't get worse?	Personal Vulnerability	.791	.255
26...peaceful?	Filler		
27...like there was no escape?	Personal Vulnerability	.813	.256
28...like you were carrying a heavy load?	Event Load	.086	.809
29...like just giving up?	Personal Vulnerability	.773	.227
30...like there was "too much to do, too little time"?	Event Load	.177	.748

Each page begins with the prompt, "IN THE PAST WEEK, have you felt:". Each item is paired with a 5-point response scale anchored at "Not At All" and "A Lot"

^a Following oblique rotation, using finalized SOS from last study ($n = 285$)

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.
- Auerbach, S. M., & Stolberg, A. L. (1986). *Crisis intervention with children*. Washington, DC: Hemisphere.
- Bartone, P., Ursano, R., Wright, K., & Ingraham, I. (1989). The impact of a military air disaster on the health of assistance workers: A prospective study. *Journal of Nervous and Mental Disease*, 117, 317–328.
- Bernard, L., & Krupat, E. (1994). *Health psychology: Biopsychosocial factors in health and illness*. Fort Worth, TX: Harcourt Brace.
- Brantley, P., Catz, S., & Boudreaux, E. (1998). Daily Stress Inventory. In C. Zalaquett & R. Wood (Eds.), *Evaluating stress: A book of resources* (Vol. 2, pp. 97–112). Lanham, MD: Scarecrow.
- 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: A book of resources* (pp. 405–420). Lanham, MD: Scarecrow.
- Brantley, P., Waggoner, C., Jones, G., & Rappaport, N. (1987). A Daily Stress Inventory: Development, reliability, and validity. *Journal of Behavioral Medicine*, 10, 61–74.
- Brown, G., & Harris, T. (1978). *Social origins of depression: A study of depressive disorder in women*. New York: Free Press.
- Brown, G., & Harris, T. (1989). *Life events and illness*. New York: Guilford.
- Cleary, P., Goldberg, D., Kessler, L., & Nyez, G. (1982). Screening for mental disorder among primary care patients: Usefulness of the General Health Questionnaire. *Archives of General Psychiatry*, 39, 837–840.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385–396.
- 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: Oxford University Press.
- Costa, P., & McCrae, R. (1987). Hypochondriasis, neuroticism, and aging: When are somatic complaints unfounded? *American Psychologist*, 40, 19–28.
- Crowne, D., & Marlowe, D. (1964). *The approval motive*. New York: Wiley.
- DeLongis, A., Coyne, J., Dakof, G., Folkman, S., & Lazarus, R. (1982). Relationship of daily hassles, uplifts, and major life events to health status. *Health Psychology*, 1, 119–136.
- Derogatis, L. (1984). *The Derogatis Stress Profile (DSP): Preliminary administration and scoring manual*. Baltimore, MD: Clinical Psychometric Research.
- Derogatis, L., & Fleming, M. (1997). The Derogatis Stress Profile (DSP). In C. Zalaquett & R. Wood (Eds.), *Evaluating stress: A book of resources* (pp. 113–140). Lanham, MD: Scarecrow.
- Dohrenwend, B. S. (2002). Social stress and community psychology. In T. A. Revenson (Ed.), *A quarter century of community*

- psychology: *Readings from the American Journal of Community Psychology* (pp. 103–117). New York: Plenum.
- Dohrenwend, B. P. (2006). Inventorying stressful life events as risk factors for psychopathology: Towards resolution of the problem of intracategory variability. *Psychological Bulletin*, *132*, 477–495.
- Dohrenwend, B. S., & Dohrenwend, B. P. (1981). *Stressful life events and their contexts*. New York: Prodist.
- Dohrenwend, B. S., & Dohrenwend, B. P. (1984). Symptoms, hassles, social supports, and life events: The problem of confounded measures. *Journal of Abnormal Psychology*, *93*, 222–230.
- Fortunato, V., & Goldblatt, A. (2002). Construct validation of a revised Strain-Free Negative Affectivity scale. *Educational and Psychological Measurement*, *62*, 45–63.
- Funk, S., & Houston, K. (1987). A critical analysis of the Hardiness Scale's validity and utility. *Journal of Personality and Social Psychology*, *53*, 572–578.
- Gentry, W., & Kobasa, S. (1984). Social and psychological resources mediating stress–illness relationships in humans. In W. Gentry (Ed.), *Handbook of behavioral medicine* (pp. 87–116). New York: Guilford.
- Goldberg, D. (1972). *The detection of psychiatric illness by questionnaire*. London: Oxford University Press.
- Goldman, S., Kraemer, D., & Salovey, P. (1996). Beliefs about mood moderate the relationship of stress to illness and symptom reporting. *Journal of Psychosomatic Research*, *41*, 115–128.
- Hatzenbuehler, M., Nolen-Hoeksema, S., & Erickson, S. (2008). Minority stress predictors of HIV risk behavior, substance use, and depressive symptoms: Results from a prospective study of bereaved gay men. *Health Psychology*, *27*, 455–462.
- Hobfoll, S. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, *44*, 513–524.
- 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 and Coping*, *11*, 181–212.
- Holmes, T., & Rahe, R. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research*, *11*, 213–218.
- Hough, R., Fairbanks, D., & Garcia, A. (1976). Problems in the ratio measurement of life stress. *Journal of Health and Social Behavior*, *17*, 70–82.
- Kanner, A., Coyne, J., Schaefer, C., & Lazarus, R. (1981). Comparisons of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of Behavioral Medicine*, *4*, 1–39.
- Kobasa, S., Maddi, S., & Kahn, S. (1982). Hardiness and health: A prospective study. *Journal of Personality and Social Psychology*, *42*, 168–177.
- Kosaka, M. (1996). Relationship between hardiness and psychological stress response. *Journal of Performance Studies*, *3*, 35–40.
- Lazarus, R. (1985). Puzzles in the study of daily hassles. *Journal of Behavioral Medicine*, *7*, 375–389.
- Lazarus, R. (1990). Theory-based stress measurement. *Psychological Inquiry*, *1*, 3–13.
- Lazarus, R., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
- Lazarus, R., & Folkman, S. (1989). *Manual: Hassles and Uplifts scales*. Redwood City, CA: Mind Garden.
- Lei, H., & Skinner, H. (1980). A psychometric study of life events and social readjustment. *Journal of Psychosomatic Research*, *24*, 57–65.
- Levenstein, S., Prantera, V., Varvo, M., Scribano, L., Berto, E., Luzi, C., et al. (1993). Development of the Perceived Stress Questionnaire: A new tool for psychometric research. *Journal of Psychometric Research*, *37*, 19–32.
- Lin, N., & Ensel, W. M. (1989). Life stress and health: Stressors and resources. *American Sociological Review*, *54*, 382–399.
- Littman, A., White, E., Satia, J., Bowen, D., & Kristal, A. (2006). Reliability and validity of 2 single item measures of psychosocial stress. *Epidemiology*, *17*, 398–403.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635–694.
- Maddi, S., Bartone, P., & Puccetti, M. (1987). Stressful events are indeed a factor in physical illness. *Journal of Personality and Social Psychology*, *52*, 833–843.
- Masten, A. (1994). Resilience in individual development: Successful adaptation despite risk and adversity. In M. Wang & E. Gordon (Eds.), *Educational resilience in inner-city America: Challenges and prospects* (pp. 3–25). Hillsdale, NJ: Erlbaum.
- McEwen, B. (2000). The neurobiology of stress: From serendipity to clinical relevance. *Brain Research*, *886*, 172–189.
- McEwen, B. (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: Cambridge University Press.
- McNemar, Q. (1975). *Psychological statistics* (5th ed.). New York: Wiley.
- Meyer, I. (1995). Minority stress and mental health in gay men. *Journal of Health and Social Behavior*, *36*, 38–56.
- Nielsen, R., Kristensen, T., Schnohr, P., & Grenbaek, 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.
- Ouelette, S. (1993). Inquiries into hardiness. In *Handbook of stress: Theoretical and clinical aspects*. New York: Free Press.
- Paulhaus, D. (1988). *Assessing self-deception and impression management in self-reports: The Balanced Inventory of Desirable Responding*. Vancouver, Canada: U. British Columbia.
- Pearlin, L., Lieberman, M., Menaghan, E., & Mullen, J. (1981). The stress process. *Journal of Health and Social Behavior*, *22*, 337–356.
- Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.
- Radmacher, S., & Sheridan, C. (1989). The Global Inventory of Stress: A comprehensive approach to stress assessment. *Medical Psychotherapy: An International Journal*, *2*, 75–80.
- Raeburn, J., Atkinson, J., Dubignon, J., McPherson, M., & Elkind, G. (1993). “Unstress”: A low-cost community psychology approach to stress-management. *Journal of Community Psychology*, *21*, 113–123.
- Rahe, R., Foistad, R., & Beergan, C. (1974). A model for life changes and illness research: Cross-cultural data from the Norwegian navy. *Archives of General Psychiatry*, *31*, 172–178.
- Roth, D., Wiebe, D., Fillingim, R., & Shay, K. (1989). Life events, fitness, hardiness, and health: A simultaneous analysis of proposed stress-resistance effects. *Journal of Personality and Social Psychology*, *57*, 136–142.
- Rutter, M. (1985). Resilience in the face of adversity: Protective factors and resistance to psychiatric disorder. *British Journal of Psychiatry*, *147*, 598–611.
- Sagrestano, L., Rodriguez, A., Caroll, D., Bieniarz, B., Greenberg, A., Castro, L., et al. (2006). A comparison of standardized measures of psychosocial variables with single-item screening measures used in an urban obstetric clinic. *Journal of Obstetric, Gynecological, and Neonatal Nursing*, *31*, 147–155.
- Sarason, M., & Sarason, B. (1984). Life changes, moderators of stress, and health. In A. Baum, S. Taylor, & J. Singer (Eds.), *Handbook of psychology and health: Social psychological aspects of health* (Vol. 4, pp. 279–299). Hillsdale, NJ: Erlbaum.
- Selye, H. (1956). *The stress of life*. New York: McGraw-Hill.

- Sheridan, C., & Radmacher, L. (1998). The Global Inventory of Stress (GIS). In C. Zalaquett & R. Wood (Eds.), *Evaluating stress: A book of resources* (Vol. 2, pp. 119–139). Lanham, MD: Scarecrow.
- Sheridan, C., & Smith, S. (1987). Towards a comprehensive scale of stress assessment: Norms, reliability, and validity. *International Journal of Psychosomatics*, 34, 48–54.
- Sinha, R. (2008). Chronic stress, drug use, and vulnerability to addiction. *Annals of the New York Academy of Sciences*, 1141, 105–130.
- Skodol, A., Dohrenwend, B., Link, B., & Shrout, P. (1990). The nature of stress: Problems of measurement. In J. Noshpitz & R. Coddington (Eds.), *Stressors and the adjustment disorders* (pp. 3–20). Oxford: Wiley.
- Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629.
- Stone, A. (1995). Measurement of affective response. In S. Cohen, R. Kessler, & L. Gordon (Eds.), *Measuring stress* (pp. 148–171). New York: Oxford University Press.
- Tanaka, J. (1993). Multifaceted conceptions of fit in structural equation models. In K. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 10–39). Newbury Park, CA: Sage.
- Turner, R., & Lloyd, D. (1995). Lifetime traumas and mental health: The significance of cumulative adversity. *Journal of Health and Social Behavior*, 36, 360–376.
- Vinokur, A., Price, R., & Schul, Y. (1995). Impact of the JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology*, 23, 39–74.
- Watson, D., & Pennebaker, J. (1989). Health complaints, stress, distress: Exploring the central role of negative affectivity. *Psychological Review*, 96, 234–254.
- Wilson, W., Rosenthal, B., & Austin, S. (2005). Exposure to community violence and upper respiratory illness in older adolescents. *Journal of Adolescent Health*, 36, 313–319.