The Safety Organizing Scale
Development and Validation of a Behavioral Measure of Safety Culture in Hospital Nursing Units

Timothy J. Vogus, PhD,* and Kathleen M. Sutcliffe, PhD†

Background: Evidence that medical error is a systemic problem requiring systematic solutions continues to expand. Developing a “safety culture” is one potential strategy toward improving patient safety. A reliable and valid self-report measure of safety culture is needed that is both grounded in concrete behaviors and is positively related to patient safety.

Objective: We sought to develop and test a self-report measure of safety organizing that captures the behaviors theorized to underlie a safety culture and demonstrates use for potentially improving patient safety as evidenced by fewer reported medication errors and patient falls.

Subjects: A total of 1685 registered nurses from 125 nursing units in 13 hospitals in California, Indiana, Iowa, Maryland, Michigan, and Ohio completed questionnaires between December 2003 and June 2004.

Research Design: The authors conducted a cross-sectional assessment of factor structure, dimensionality, and construct validity.

Results: The Safety Organizing Scale (SOS), a 9-item unidimensional measure of self-reported behaviors enabling a safety culture, was found to have high internal reliability and reflect theoretically derived and empirically observed content domains. The measure was shown to discriminate between related concepts like organizational commitment and trust, vary significantly within hospitals, and was negatively associated with reported medication errors and patient falls in the subsequent 6-month period.

Conclusions: The SOS not only provides meaningful, behavioral insight into the enactment of a safety culture, but because of the association between SOS scores and reported medication errors and patient falls, it also provides information that may be useful to registered nurses, nurse managers, hospital administrators, and governmental agencies.

Key Words: high reliability, medical error, patient safety, safety culture, safety organizing

On the heels of several troubling reports by the Institute of Medicine regarding the incidence and consequences of medical error1,2—the failure of a medical or nursing action to be completed as intended (eg, patients falling from their beds because nurses do not pull up the bedside railing) or the use of a wrong plan to achieve an aim (eg, administering a wrong drug or dosage to a patient) that may result in harm—patient safety has begun to receive renewed attention by researchers and practitioners alike. Although the actual number of medical errors3,4 and their preventability5 have been intensely debated, the fact remains that such errors are frequent6,7 and costly both in human lives and expense to the healthcare system.8 In the search for ways to reduce the incidence of such errors and improve safety, recent research has concluded that the most fruitful solutions focus on defects in the system of care delivery that give rise to errors.1,9 One such systemic solution frequently recommended to healthcare organizations is developing a “culture of safety.” A safety culture is the product of the shared values, attitudes, and patterns of behavior that determine the observable degree of effort with which all organizational members direct their attention and actions toward minimizing patient harm that may result from the process of care delivery.10,11 An effective safety culture relies on ongoing collecting, analyzing, and disseminating of information from errors as well as proactive checks on the organization’s vital signs.10

Recent studies provide evidence that safety cultures and patient safety are enabled by supportive leaders.12,13 An emerging body of literature has also begun to empirically link safety culture to medical errors14,15 and mortality rates.16 Although these studies are instructive, they offer only indirect evidence regarding how a safety culture can reduce the incidence of medical errors. Without a way to directly assess the behaviors underlying a safety culture poses problems because it impedes ascertaining the exact mechanisms through which leader and other interventions reduce errors. As a result, interventions cannot be designed to ensure that they trigger behaviors that reduce the incidence of error.
To better understand the behaviors underlying an effective safety culture, we have developed the Safety Organizing Scale (SOS). Our primary objective was to develop a measure that is both substantively and methodologically consistent with the concept of safety culture. This means that it must be behavioral, shared at the unit level (indicative of collective and shared processes rather than individual behaviors), and associated with indicators of patient safety. Our measure needs to be shared and grounded in concrete behaviors because if a safety culture is present and consequential, it must be manifest in both shared attitudes and observable behaviors.\(^{19,20}\)

We focus on safety culture at the level of the caregiving unit because a safety culture is enacted collectively with one’s most proximate colleagues and because variations in safety cultures, attitudes, and outcomes are evident at the unit level even within the same hospital.\(^{14,19,20}\)

Theoretically, we ground our development of the SOS in the detailed case studies of “high-reliability organizations” (HROs). HROs such as aircraft carrier flight decks\(^21\) and nuclear power plants\(^22\) are organizations that operate hazardous technologies in a nearly error-free manner under trying conditions rife with complexity, interdependence, and time pressure. Case studies of HROs as well as healthcare organizations following the principles of HROs\(^23\) suggest that safety cultures enable their reliable performance. Safety culture is seen as coming to life in HROs through behavioral processes of “collective mindfulness” enacted by front-line employees.\(^23\)–\(^26\)

Collective mindfulness consists of 5 interrelated behavioral processes: preoccupation with failure, reluctance to simplify interpretations, sensitivity to operations, commitment to resilience, and deference to expertise.\(^24\)–\(^26\)

We validate the SOS using a sample of registered nurses (RNs) on hospital nursing units. We use RNs because nurses (RNs) on hospital nursing units. We use RNs because RNs: reported medication errors and patient falls.

### METHODS

#### Setting

The units and RNs for this study were drawn from private, nonprofit Catholic hospitals that are members of a large Catholic Health System in the United States with member hospitals in California, Idaho, Indiana, Iowa, Maryland, Michigan, and Ohio. We conducted our multisite cross-sectional study between December 2003 and June 2004 using a convenience sample of 13 hospitals from this system. The participating hospitals included 5 urban hospitals (in the Columbus, Ohio, Detroit, Michigan, and Washington DC, Metropolitan Statistical Areas), 5 midsized metropolitan hospitals (in the Fresno, California, South Bend, Indiana, Battle Creek/Kalamazoo, Michigan, and Grand Rapids/Muskegon, Michigan, Metropolitan Statistical Areas), and 3 rural hospitals (in the rural Michigan and rural Iowa Metropolitan Statistical Areas). In addition to varying in geographic location, the participating hospitals also varied in size from 89 beds to 478 acute care beds. To increase generalizability of the SOS within hospitals, we sampled and received responses from a wide array of inpatient units (n = 125). Additional details regarding hospital, unit, and RN characteristics are presented in Table 1.

#### Response Rate

We surveyed all RNs in the participating hospitals and received responses to 1685 of 3298 questionnaires sent for a...
response rate of 51.1%. Although this rate is consistent with published studies in the medical literature using a mailed questionnaire,28 such a response rate may be problematic if it biases or meaningfully affects the conclusions. Therefore, we investigated whether the response rate for a given unit was correlated with the SOS as well as other variables of interest, including reported medication errors, patient falls, RN staffing levels, or unit size. None of the correlations between response rate and these variables were statistically significant.

To further assess whether our sample was in any way biased, we conducted a series of $t$ tests comparing the unit-level means reported by RN respondents with unit-level means provided by nurse managers derived from the hospital’s human resource information systems to determine if the RNs that responded to our survey were significantly different from those who did not respond in terms of age, 3 types of tenure (within the nursing profession, with the hospital, and on the unit), and level of education. None of these $t$ tests revealed significant differences between respondents and nonrespondents, which suggests response bias is less likely to threaten the validity of our analyses.

**Measure Validation Strategy**

An effective measure should be theory-based, reliable, valid, relevant to the unit of analysis, and relatively easy to administer. Therefore, our first step was to determine the content validity of the SOS—the degree to which an instrument captures its intended construct in a manner consistent with relevant theory. To do so, we developed survey items that were closely linked to the theoretical literature on HROs that identifies safe performance as being a function of “collective mindfulness” as reflected by the 5 processes of preoccupation with failure, reluctance to simplify interpretations, sensitivity to operations, commitment to resilience, and deference to expertise.24–26 Each of these processes is precisely defined and linked to corresponding survey items in Table 2. Additional descriptive data on the SOS and its items are provided in the Appendix and mean

![FIGURE 1. Mean levels of the Safety Organizing Scale by hospital nursing unit.](image)
levels of the SOS by unit are displayed in Figure 1. We tailored our items to hospital nursing based on case studies of hospitals attempting to become HROs and our own qualitative fieldwork in 2 hospital nursing units in a large Midwestern teaching hospital and in 2 geographically separate academic emergency departments. To further ensure that our survey items were consistent with existing theory and clear to RNs, we also had 7 experts in HROs and nursing assess the measure’s content validity and pretested the SOS questionnaire with a sample of 45 RNs on a coronary care unit in a large Midwestern teaching hospital. Based on these pretests, we made one minor wording modification to a single item.

A measure is reliable when a set of items form an internally consistent scale. Reliability is a necessary but not sufficient condition for validity. The construct validity of survey-based measures like the SOS is empirically determined and evidence of validity exists when a scale actually measures what it intends to measure (convergent validity), differentiates from other constructs (discriminant validity), and is related to its theoretical causes, correlates, and effects (criterion validity). It is important to note that we assess criterion validity using responses to the SOS aggregated to the unit level. We do so because the SOS is only theoretically meaningful at the unit level, and we want to assess the relationship between the SOS of a given nursing unit and its safety performance. We justify the use of the aggregate measure and conduct all other analyses related to convergent, discriminant, and criterion validity next.

RESULTS

Reliability and Convergent Validity

Convergent validity is the degree to which multiple items agree and is tested using confirmatory factor analysis (CFA). In this case, a CFA model estimated using AMOS 5.0 tests the hypothesis that the 9 items of the SOS are reflective of a single underlying factor. A CFA measurement model displays convergent validity if items load significantly onto the intended factor and model fit indices suggest adequate fit. The results of the measurement model are shown in Table 3, column 1. Our model demonstrated excellent fit across all fit indices (CFI = 0.964, incremental fit index = 0.964, root mean square error of approximation = 0.055, standardized root mean square residual = 0.033) and all factor loadings were highly significant (P < 0.001). Taken in total, these results provide strong evidence for convergent validity. The Cronbach’s alpha for the SOS was 0.88, strongly supporting scale reliability.

Discriminant Validity

We assessed discriminant validity by conducting “pairwise tests” of theoretically related constructs. If items from other constructs are distinct from the SOS (ie, load on different factors), it is evidence of discriminant validity. Two constructs theorized to be essential to the development of safety cultures—employee commitment and trust in manager—served as comparison constructs. Commitment was measured using 3 items adapted from Meyer and Allen (eg, “I feel emotionally attached to my current work unit”) and 2 items for trust in manager (“My manager has a reputation for fairness in dealing with RNs” and “My manager demonstrates absolute integrity”).

We evaluated discriminant validity by comparing the fit of 4 CFA models. First, we specified a 3-factor measurement model, including the SOS, commitment, and trust in manager. Second, we specified a 2-factor model in which one factor consisted of the items for the SOS and commitment and the second factor consisted of the items comprising trust in manager. Third, we specified a 2-factor model in which one factor consisted of the items for the SOS and trust in manager and the second factor consisted of the items comprising commitment. Lastly, we specified a one-factor model, including all the items for the SOS, commitment, and trust in manager scales. If the fit of the 3-factor model provides a significantly better fit than the alternative 2-factor (in which the SOS and commitment or the SOS and trust in manager load onto the same factor) or one-factor models using a χ² statistic, one can conclude that commitment and trust in manager are distinct from the SOS. Table 3 columns 3, 4, and 5 reports the results of these analyses and suggest that commitment and trust in manager are indeed distinct from the SOS (changes in χ² of 1193.17, 1670.99, and 2779.90, respectively, all differences significant at P < 0.001).

Statistical Justification for Aggregation

To ensure that the SOS was meaningful at the unit level, we constructed survey items to direct respondents’ attention to common experiences of all RNs on the unit (eg, “we” or “RNs on this unit”). The survey instructions also encouraged respondents to assume the shared perspective of the group. Having items and directions that focus respondent attention on shared experience do, in fact, engender less within-group variability and more between-group variability than comparable survey items that reference individual experiences and perceptions.

For aggregation to be statistically appropriate, it is necessary to demonstrate that 1) the members of each unit reported similar scores for the unit on a given measure and 2) the units have significant between unit variance for a given measure. Four complementary measures of within-group agreement were used to determine the degree of congruence between individual RNs’ survey responses and the appropriateness of aggregating these measures to the unit level: the median r_wg(i), the F-statistic from a one-way analysis of variance (ANOVA), intraclass correlation coefficient (ICC)(1) and ICC(2). r_wg(i) measures the degree to which individual responses within a group are interchangeable; median r_wg(i) values of 0.70 or greater provide evidence of acceptable agreement among individual responses on a scale. Every unit in the sample had an r_wg(i) Value greater than 0.87 with a median value of 0.98. A significant F-statistic resulting from a one-way ANOVA with unit membership as the independent variable and the SOS as the dependent variable indicates that responses differ between RNs in different nursing units. The one-way ANOVA of
The Safety Organizing Scale

TABLE 3. Confirmatory Factor Analyses Results for Convergent and Discriminant Validity of the SOS

<table>
<thead>
<tr>
<th>Construct and Items</th>
<th>3-Factor Model (n = 1685)</th>
<th>2-Factor Model SOS/Commitment and Trust in Manager (n = 1685)</th>
<th>2-Factor Model SOS/Trust in Manager and Commitment (n = 1685)</th>
<th>1-Factor Model (n = 1685)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1: Good “map” of skills.</td>
<td>0.606</td>
<td>0.616</td>
<td>0.599</td>
<td>0.607</td>
</tr>
<tr>
<td>Q2: Discuss mistakes to learn</td>
<td>0.723</td>
<td>0.705</td>
<td>0.711</td>
<td>0.694</td>
</tr>
<tr>
<td>Q3: Know others’ skills</td>
<td>0.697</td>
<td>0.686</td>
<td>0.685</td>
<td>0.674</td>
</tr>
<tr>
<td>Q4: Discuss alternatives to normal work process</td>
<td>0.676</td>
<td>0.668</td>
<td>0.675</td>
<td>0.666</td>
</tr>
<tr>
<td>Q5: During report, discuss risks</td>
<td>0.559</td>
<td>0.540</td>
<td>0.550</td>
<td>0.532</td>
</tr>
<tr>
<td>Q6: Take advantage unique skills</td>
<td>0.670</td>
<td>0.663</td>
<td>0.665</td>
<td>0.657</td>
</tr>
<tr>
<td>Q7: Identify activities do not want to go wrong</td>
<td>0.683</td>
<td>0.666</td>
<td>0.680</td>
<td>0.662</td>
</tr>
<tr>
<td>Q8: Discuss error prevention</td>
<td>0.723</td>
<td>0.706</td>
<td>0.722</td>
<td>0.704</td>
</tr>
<tr>
<td>Q9: Pool expertise in crisis</td>
<td>0.613</td>
<td>0.607</td>
<td>0.613</td>
<td>0.604</td>
</tr>
<tr>
<td>Commitment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1: Personal meaning in work unit</td>
<td>0.815</td>
<td>0.515</td>
<td>0.814</td>
<td>0.525</td>
</tr>
<tr>
<td>Q2: Emotionally attached to unit</td>
<td>0.802</td>
<td>0.505</td>
<td>0.805</td>
<td>0.513</td>
</tr>
<tr>
<td>Q3: Belonging in work unit</td>
<td>0.698</td>
<td>0.475</td>
<td>0.697</td>
<td>0.485</td>
</tr>
<tr>
<td>Trust in manager</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1: Reputation of fairness</td>
<td>0.889</td>
<td>0.890</td>
<td>0.464</td>
<td>0.481</td>
</tr>
<tr>
<td>Q2: Absolute integrity</td>
<td>0.931</td>
<td>0.930</td>
<td>0.480</td>
<td>0.497</td>
</tr>
<tr>
<td>Model fit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2$ baseline model</td>
<td>10362.39 (df = 91)</td>
<td>10362.39 (df = 91)</td>
<td>10362.39 (df = 91)</td>
<td>10362.39 (df = 91)</td>
</tr>
<tr>
<td>$\chi^2$ default model</td>
<td>442.24 (df = 72)</td>
<td>1635.41 (df = 74)</td>
<td>2113.23 (df = 74)</td>
<td>3222.14 (df = 75)</td>
</tr>
<tr>
<td>$\Delta \chi^2$ (from 3 factor)</td>
<td>1193.17 (df = 2)</td>
<td>1670.99 (df = 2)</td>
<td>2779.90 (df = 3)</td>
<td></td>
</tr>
<tr>
<td>CFI*</td>
<td>0.964</td>
<td>0.848</td>
<td>0.801</td>
<td>0.694</td>
</tr>
<tr>
<td>IFI†</td>
<td>0.964</td>
<td>0.848</td>
<td>0.802</td>
<td>0.694</td>
</tr>
<tr>
<td>RMSEA‡</td>
<td>0.055</td>
<td>0.112</td>
<td>0.128</td>
<td>0.158</td>
</tr>
<tr>
<td>SRMR§</td>
<td>0.033</td>
<td>0.075</td>
<td>0.073</td>
<td>0.095</td>
</tr>
</tbody>
</table>

Q1 through Q9 indicate the question number associated with the corresponding scale.

The comparative fit index represents the fit of the model relative to the baseline model of independence among the observed variables. Values range from 0 to 1 with a good model fit being 0.95. The incremental fit index represents the fit of the model relative to the baseline model of independence among the observed variables. Values range from 0 to 1 with a good model fit being 0.95. The root mean square error of approximation is a population discrepancy function of the acceptability of the model that takes into account the complexity of the model. RMSEA assesses how well the model would fit the population covariance matrix were it known. A value less than 0.06 indicates good model fit. The standardized root mean square residual is a summary measure of average standardized residuals when the model variance–covariance matrix is compared with the sample data variance–covariance matrix. Values range from 0 to 1. Values less than 0.08 represent good model fit. The Safety Organizing Scale indicates Safety Organizing Scale; n, the number of registered nurse respondents.

unit on SOS has a highly significant F-statistic ($F = 2.73$, $P < 0.001$).

The 2 forms of the intraclass correlation, referred to as ICC(1) and ICC(2), provide omnibus indices of homogeneity and are calculated from a one-way ANOVA in which the SOS is the dependent variable and unit membership is the independent variable. ICC(1) can be interpreted as the proportion of total variance that is explained by unit membership with values ranging from −1 to +1 and values between 0.05 and 0.30 being most typical. ICC(1) is computed as

$$
\text{ICC}(1) = \frac{\text{MSB} - \text{MSW}}{\text{MSB} + [(k - 1) \times \text{MSW}]}
$$

where MSB is between-group mean square, MSW is within-group mean square, and k is average unit size. The ICC(1) value for the SOS is 0.20. Whereas ICC(1) provides an estimate of the reliability of a single RN’s assessment of the unit mean, ICC(2) provides an overall estimate of the reliability of unit means. The ICC(2) is computed as (MSB − MSW)/MSB. The closer ICC(2) is to 1.00, the more reliably nursing units can be distinguished based on individual RNs’ perceptions of the SOS with values equal to or above 0.70 being acceptable. The ICC(2) value for the SOS is 0.74. In sum, the results of these 4 analyses strongly support the idea that the SOS reflects a unit-level construct and aggregation of individual responses is justified. Given this strong support, we created the unit-level SOS as the average of all the individual responses.
Criterion Validity

We assessed criterion validity based on hypothesized relationships between the unit-level SOS and its proposed causes and effects. First, safety organizing emerges when leaders create a supportive context that makes it safe for RNs to engage in the behaviors that comprise the SOS (eg, discussing errors).12,13,22-24 Specifically, these behaviors should be more prevalent when RNs trust managers and administrators to treat them fairly and otherwise act with integrity.10,22 Therefore, we expected that trust in manager should be more prevalent when RNs trust managers and should be sufficiently motivated to engage in safety organizing. Third, studies have consistently linked lower patient-to-RN ratios and performance outcomes such as mortality39,40 failure to rescue,40,41 complications,41 and length of stay.41 Fewer patients per nurse enable RNs collectively to be more vigilant in their ongoing surveillance such that they can detect and correct emerging and manifest errors in a timely manner that minimizes negative outcomes for patients.27 As such, patients per RN should be negatively related to the SOS. Lastly, past case studies of HROs23-25 have shown that the behaviors measured by the SOS are responsible for their nearly error-free performance. Thus, we expected that the SOS would be negatively related to reported medication errors and patient falls.

As previously mentioned, trust in manager and commitment were measured using survey items. Patient-to-RN ratio was measured using data provided by nurse managers. We also conducted our analyses using alternate measures of this ratio from survey data and the results remained substantively identical. Reported medication errors were defined as occurring whenever the right medication was not given to either the right patient, at the right time, in the right dose, or through the right route (eg, intravenous). Reported patient falls were defined as occurring any time a previously standing, sitting, or lying patient is found on the floor. Both measures were derived from incident reports filed at the time of a medication error or patient fall and were measured as the number of errors or falls reported on a unit for the 6 months after the collection of the survey data. In all our regression models, we also control for other factors that might be associated with the SOS, reported medication errors, and reported patient falls, including average RN tenure on the unit, the percentage of RNs on a unit with at least a bachelor’s degree in nursing, the type of unit (a series of dummy variables for emergency department, intensive care, labor and pediatrics, surgery, and internal medicine), and hospital location (dummy variables for rural, suburban, and urban).

Criterion validity was examined through the regression analyses reported in Table 4.32,43 The first model shows the effects of theorized antecedents on the SOS and the results confirm the expected relationships as trust in manager (B = 0.164, P < 0.001) and commitment (B = 0.295, P < 0.001)

### TABLE 4. Multilevel Regression Analysis Results for Criterion Validity of the SOS

<table>
<thead>
<tr>
<th>Variables</th>
<th>SOS</th>
<th>Reported Medication Errors</th>
<th>Reported Patient Falls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (P)</td>
<td>95% CI</td>
<td>Coefficient (P)</td>
</tr>
<tr>
<td>Trust in manager</td>
<td>0.164 (&lt;0.001)</td>
<td>0.09 to 0.238</td>
<td>-0.678 (&lt;0.001)</td>
</tr>
<tr>
<td>Commitment</td>
<td>0.295 (&lt;0.001)</td>
<td>0.182 to 0.408</td>
<td>0.198 (&lt;0.001)</td>
</tr>
<tr>
<td>Patient-to-RN ratio</td>
<td>-0.056 (&lt;0.001)</td>
<td>-0.084 to -0.027</td>
<td>0.101 (&lt;0.001)</td>
</tr>
<tr>
<td>Percent RNs with BSN</td>
<td>0.214 (0.108)</td>
<td>0.047 to 0.474</td>
<td>-0.568 (&lt;0.001)</td>
</tr>
<tr>
<td>Average tenure on unit</td>
<td>-0.001 (0.985)</td>
<td>-0.020 to 0.019</td>
<td>0.044 (&lt;0.001)</td>
</tr>
<tr>
<td>Emergency department</td>
<td>-0.361 (0.002)</td>
<td>-0.589 to -0.132</td>
<td>-0.643 (&lt;0.001)</td>
</tr>
<tr>
<td>Intensive care unit</td>
<td>-0.051 (0.582)</td>
<td>-0.233 to 0.113</td>
<td>0.399 (0.003)</td>
</tr>
<tr>
<td>Obstetrics and pediatrics</td>
<td>-0.058 (0.507)</td>
<td>-0.229 to 0.113</td>
<td>-0.212 (0.109)</td>
</tr>
<tr>
<td>Surgery</td>
<td>-0.006 (0.955)</td>
<td>-0.208 to 0.196</td>
<td>-0.035 (0.813)</td>
</tr>
<tr>
<td>Constant</td>
<td>3.044 (&lt;0.001)</td>
<td>2.512 to 3.576</td>
<td>3.453 (&lt;0.001)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level-2 Model</th>
<th>Variance (SE)</th>
<th>Variance (SE)</th>
<th>Variance (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.021 (0.033)</td>
<td>0.093 (0.051)</td>
<td>0.004 (0.027)</td>
</tr>
<tr>
<td>Urban</td>
<td>0.021 (0.047)</td>
<td>0.887 (0.413)</td>
<td>0.575 (0.280)</td>
</tr>
<tr>
<td>Suburban</td>
<td>0.021 (0.043)</td>
<td>0.318 (0.226)</td>
<td>0.589 (0.281)</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-17.794</td>
<td>-508.239</td>
<td>-369.038</td>
</tr>
<tr>
<td>n</td>
<td>118*</td>
<td>94</td>
<td>94</td>
</tr>
</tbody>
</table>

All analyses were conducted in GLLAMM with hospital-level random effects. Including random effects accounts for the multilevel structure of the data (ie, units nested within hospitals).42

*The sample size is 118 resulting from missing patient-to-RN ratio data for 7 units. The units with missing data are not systematically different from the included units in terms of size (number of beds), scores on the SOS, or RN characteristics (age, education, or tenure).

1Multilevel Poisson regression models were estimated using GLLAMM to model reported medication errors and patient falls. We used Poisson rather than negative binomial models because hospital-level random effects sufficiently account for extra-Poisson variation (ie, overdispersion such that the variance is greater than the expectation).43

2The sample size is 94 for these analyses because these units did not provide their reported medication error or patient fall data. The units with missing data are not systematically different from the included units in terms of size (number of beds), scores on the SOS, or RN characteristics (age, education, or tenure).

SOS indicates Safety Organizing Scale; CI, confidence interval, RN, registered nurse, SE, standard error; n, the number of nursing units included in the analysis.
are positively related to the SOS and patient-to-RN ratio is negatively related to the SOS (ie, more patients per nurse is associated with a lower score on the SOS, B = −0.06, P < 0.001). Models 2 and 3 show the SOS is negatively related to reported medication errors (B = −0.678, P < 0.001) and reported patient falls (B = −0.629, P < 0.001).

**DISCUSSION**

The 9-item SOS appears to be a precise, unidimensional measure of the self-reported behaviors underlying a safety culture. The factor structure closely resembles the initial content domains identified in the extensive case study literature on HROs and we substantively differentiated the SOS from 2 related concepts. The data also supported the SOS as a unit-level construct because aggregation of individual RN responses to the unit level was well justified. The strong relationship between the SOS and reported medication errors and patient falls provides additional support for its construct validity as well as its use. Consequently, the SOS may help differentiate between safe and unsafe units by examining the ongoing actions and interactions on the unit.

Although some studies suggest that supportive leadership and a well-developed safety culture should be associated with more reporting of errors, others have not. We find that high levels of the SOS are associated with fewer reported medication errors and patient falls. We argue that these results coupled with additional data analyses suggest that our measure did not only reflect the tendency to report, but also provided a partial picture of patient safety on the units. If reporting more errors is associated with being a high-functioning (ie, safe) unit, we would expect reported medication errors and patient falls to be positively associated with other indicators of safety performance. We find that nurse managers’ assessments of their unit’s safety performance (a 2-item survey measure) are actually negatively associated with reported medication errors (P < 0.01) and patient falls (P = 0.012). That is, high numbers of reported medication errors and patient falls were associated with low ratings of quality of care by nurse managers.

To date, there have been few self-report measures developed that focus on the behaviors underlying a safety culture. To the extent that measures related to safety culture do exist, they have tended to focus on factors that create a context that promotes safety, including leaders, processes, and procedures that prioritize safety; encourage open communication and reporting of errors; and otherwise ensure that errors are handled appropriately and handled appropriately rather than how those delivering health care (eg, RNs) act to promote safety. Although some of the more comprehensive questionnaires do possess items similar to the SOS, they only capture a portion of the behaviors documented in case studies of HROs. Given that a supportive context is essential for safety organizing, it would be fruitful to explore the relationship among measures of safety culture, the SOS, and safety outcomes.

**Limitations**

The findings of our study should be considered in light of its limitations. First, the research was conducted in a convenience sample of 13 Catholic hospitals that were all members of the same health system. Although this may limit the generalizability of our findings to all Catholic hospitals, other types of hospitals (eg, for-profit hospitals, academic medical centers) or health systems, the hospitals in our sample range in size from small (89 beds) to large (478 beds) and in location, including rural, suburban, and urban sites, both of which should increase confidence in the potential generalizability of the findings. In addition, the characteristics of the sample of RNs surveyed are nearly identical to the characteristics identified in a comprehensive national sample of RNs in terms of age, education, and gender. However, further research is warranted to confirm that the SOS applies to a broader sample of Catholic hospitals as well as other hospital types and health systems.

Second, although our response rate was reasonable for a mailed questionnaire, it varied significantly across units and across hospitals. Subsequent analyses revealed that RN response rates at the unit and hospital levels were not correlated with reported medication errors, reported patient falls, or the SOS. Thus, when these analyses are coupled with the fact that the RNs responding to our questionnaire did not differ in any significant manner from nonrespondents, we are confident that our results are not systematically biased.

Third, our measure was validated using a sample composed exclusively of RNs. Although RNs play an important role in determining unit and hospital safety and their reports of errors and falls are not an adequate measure of error rates. Therefore, before the SOS can conclusively be determined to be associated with reductions in medical error, subsequent work should build on our strong initial findings and test criterion validity with other methods (eg, direct observation, prospective clinical surveillance) assessing explicitly defined errors and adverse events.

Despite these limitations, the SOS fills an important gap in assessing the behavioral underpinnings of a unit’s safety culture. The development of the SOS follows calls for investigating the behaviors facilitating error prevention and the content of this measure is closely aligned with case studies documenting the association of such behaviors with high levels of safety in medical settings and HROs. The SOS items are also “actionable” in the sense that they entail concrete behaviors that can readily be changed by RNs and influenced by nurse managers through how they manage their employees on a daily basis.

In summary, the SOS makes 2 key contributions. First, it provides a self-report measure of the behaviors that lead to the emergence of a safety culture and is strongly associated with fewer reported medication errors and patient falls. Second, the
SOS is also a useful tool for understanding both how and under what conditions interventions designed to improve safety culture and safety outcomes such as improving patient-to-RN ratios\(^{19-41}\) and safety-conscious leadership\(^{3,14,17}\) are effective. That is, such interventions are likely to generate improvements by fostering safety organizing or, at least, their benefits should be enhanced in the presence of high levels of safety organizing. Future research could explicitly test the sensitivity of the SOS to safety-oriented design, leadership, or even technologic interventions and its subsequent effects on safety outcomes.

ACKNOWLEDGMENTS

The authors thank Bruce Cooli, Jody Hoffer Gittell, Ryan Quinn, Sara Singer, Anita Tucker, and Klaus Weber for insightful comments on earlier drafts of the manuscript. The authors also thank Deputy Editor Lisa Meredith and the journal’s 3 anonymous reviewers for thoughtful and constructive comments that substantially strengthened the quality of the manuscript.

REFERENCES


APPENDIX

Means, Standard Deviations, and Missing Values for the 9 SOS Items

<table>
<thead>
<tr>
<th>Item*</th>
<th>Mean</th>
<th>SD</th>
<th>Percent Missing</th>
<th>Number Missing</th>
<th>Alpha if Item Deleted†</th>
<th>Total n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>5.05</td>
<td>1.20</td>
<td>1.10</td>
<td>18</td>
<td>0.87</td>
<td>1685</td>
</tr>
<tr>
<td>Q2</td>
<td>4.95</td>
<td>1.28</td>
<td>1.18</td>
<td>20</td>
<td>0.86</td>
<td>1685</td>
</tr>
<tr>
<td>Q3</td>
<td>4.76</td>
<td>1.41</td>
<td>0.65</td>
<td>11</td>
<td>0.86</td>
<td>1685</td>
</tr>
<tr>
<td>Q4</td>
<td>4.29</td>
<td>1.41</td>
<td>0.53</td>
<td>9</td>
<td>0.86</td>
<td>1685</td>
</tr>
<tr>
<td>Q5</td>
<td>5.55</td>
<td>1.20</td>
<td>1.37</td>
<td>23</td>
<td>0.87</td>
<td>1685</td>
</tr>
<tr>
<td>Q6</td>
<td>5.46</td>
<td>1.17</td>
<td>0.42</td>
<td>7</td>
<td>0.86</td>
<td>1685</td>
</tr>
<tr>
<td>Q7</td>
<td>4.87</td>
<td>1.23</td>
<td>1.01</td>
<td>17</td>
<td>0.86</td>
<td>1685</td>
</tr>
<tr>
<td>Q8</td>
<td>5.16</td>
<td>1.15</td>
<td>0.65</td>
<td>11</td>
<td>0.86</td>
<td>1685</td>
</tr>
<tr>
<td>Q9</td>
<td>5.86</td>
<td>1.10</td>
<td>1.01</td>
<td>17</td>
<td>0.87</td>
<td>1685</td>
</tr>
</tbody>
</table>

Q1 through Q9 indicates the question number associated with the scale.

*The instructions read “The following questions ask you to assess the degree to which you and the other RNs with which you currently and primarily work engage in certain behaviors and practices. By work unit, we mean your current hospital unit (eg, Cardiac Intensive Care Unit).” The stem question asks “To what extent do the following characterize your current work unit?”

†Scale: 1 = not at all, 2 = to a very limited extent, 3 = to a limited extent, 4 = to a moderate extent, 5 = to a considerable extent, 6 = to a great extent, 7 = to a very great extent.

These values are offered to those who might be interested in developing a short-form version of the SOS. We encourage use of the full-scale to capture the richness of all behaviors described.

SOS indicates Safety Organizing Scale; SD, standard deviation; n, the number of registered nurse respondents.
AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1—AUTHOR: In Table 1, does N/A indicate not applicable or not available?