Workforce Perceptions of Hospital Safety Culture: Development and Validation of the Patient Safety Climate in Healthcare Organizations Survey

Sara Singer, Mark Meterko, Laurence Baker, David Gaba, Alyson Falwell, and Amy Rosen

Objective. To describe the development of an instrument for assessing workforce perceptions of hospital safety culture and to assess its reliability and validity.

Data Sources/Study Setting. Primary data collected between March 2004 and May 2005. Personnel from 105 U.S. hospitals completed a 38-item paper and pencil survey. We received 21,496 completed questionnaires, representing a 51 percent response rate.

Study Design. Based on review of existing safety climate surveys, we developed a list of key topics pertinent to maintaining a culture of safety in high-reliability organizations. We developed a draft questionnaire to address these topics and pilot tested it in four preliminary studies of hospital personnel. We modified the questionnaire based on experience and respondent feedback, and distributed the revised version to 42,249 hospital workers.

Data Collection. We randomly divided respondents into derivation and validation samples. We applied exploratory factor analysis to responses in the derivation sample. We used those results to create scales in the validation sample, which we subjected to multitrait analysis (MTA).

Principal Findings. We identified nine constructs, three organizational factors, two unit factors, three individual factors, and one additional factor. Constructs demonstrated substantial convergent and discriminant validity in the MTA. Cronbach's α coefficients ranged from 0.50 to 0.89.

Conclusions. It is possible to measure key salient features of hospital safety climate using a valid and reliable 38-item survey and appropriate hospital sample sizes. This instrument may be used in further studies to better understand the impact of safety climate on patient safety outcomes.

Key Words. Safety culture, safety climate, survey, psychometric evaluation

Since the Institute of Medicine's identification of safety culture as a key determinant of the ability of health care organizations to address and reduce risks to patients due to medical care (Institute of Medicine 2001), initiatives to improve and measure safety culture have proliferated (McCarthy and Blumenthal 2006). Recognition of the critical need to assess safety culture and the impact of innovative interventions aimed at improving it has led to the development of surveys designed to measure hospital worker perceptions of safety culture or "safety climate." Instruments vary according to length, dimensions covered, intended sample population (hospital-wide or unit-level personnel), and extent of psychometric evaluation. Assessments of these early attempts to measure safety climate agree that more consideration of the psychometric factors in the design and selection of health care safety climate instruments is appropriate (Nieva and Sorra 2003; Colla et al. 2005; Flin et al. 2006; Singla et al. 2006).

This paper describes the development and psychometric evaluation of the Patient Safety Climate in Healthcare Organizations (PSCHO) survey (appendix A), which has been used for assessing safety climate in hospitals in the United States and abroad (Singer et al. 2003; Ginsburg et al. 2005; Sharek 2005; Cooper et al. forthcoming). The PSCHO survey was developed as part of a Stanford-based patient safety research program sponsored by the Agency for Healthcare Research and Quality (RO1 HS013920).

DEVELOPMENT AND EVOLUTION OF THE PSCHO SURVEY

Theoretical Development

Theory guiding development of the PSCHO survey was based on research regarding high-reliability organizations (HROs) such as nuclear aircraft

Address correspondence to Sara Singer, M.B.A., Center for Health Policy, Stanford University, 117 Encina Commons, Stanford, CA 94305-6019. Ms. Singer is also with the Ph.D. Program in Health Policy Management, Harvard University, Center for Public Leadership, Kennedy School of Government, Cambridge, MA. Mark Meterko, Ph.D., is with the Center for Organization, Leadership, and Management Research, VA Medical Center (152-M), Boston, MA. Laurence Baker, Ph.D., is with the Department of Health Research and Policy, Stanford University, Stanford, CA. David Gaba, M.D., is with the Department of Anesthesia, VA Palo Alto Health Care System, Anesthesia Service, Stanford University, Palo Alto, CA. Alyson Falwell, M.P.H., is with the Center for Health Policy/Primary Care and Outcomes Research, Stanford University, Stanford, CA. Amy Rosen, Ph.D., is with Risk Assessment and Patient Safety, Center for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center, Bedford, MA.

carriers and commercial aviation (Rochlin, La Porte, and Roberts 1987). HROs engage in extremely complex and often fast-paced activities yet avoid catastrophic error. Observational studies of HROs suggest that successful operations require a culture of reliability centering on safety. A safety culture is necessary to encourage uniformly appropriate responses by frontline personnel (Weick 1987; Roberts 1990).

HRO theory was first applied to health care in 1994 (Gaba, Howard, and Jump 1994), and has been used increasingly in this context over the last 5 years (Knox, Simpson, and Garite 1999; Gaba 2000; Sutcliffe 2000; Roberts and Tadmor 2002; Gaba et al. 2003; Singer et al. 2003; Weick and Sutcliffe 2003; Roberts, Madsen, and Van Stralen 2005; Wilson et al. 2005; Yates et al. 2005). While there is reasonable consensus across these applications about many of the key features needed to ensure high reliability in hazardous industries, the importance of a strong safety culture is a primary element of all formulations. However, to date no theoretical model has codified discrete and measurable dimensions of safety culture. Discrepancies among existing safety climate questionnaires are due in part to the absence of a strong theoretical model with which to guide hypotheses and explain relationships among variables.

At the time of the development of the initial version of the PSCHO survey instrument, several surveys had been developed to measure safety climate in high-hazard industries other than health care (Ciavarella, Figlock, and Sengupta 1999; Libuser and Roberts 1998). In hospitals, investigators had sought to measure specific elements of safety climate, such as teamwork and production pressure in specific units or among clinicians in a single discipline (Gaba, Howard, and Jump 1994; Helmreich and Schaefer 1994). No available survey, however, had been designed to measure safety climate among all hospital personnel and across multiple hospitals of different types, and no organization-wide survey had been systematically administered and subjected to rigorous psychometric assessment.

A review of existing safety climate survey instruments led to the identification of 16 topics with theoretical support from the HRO literature (see appendix B). These topics represented characteristics expected to comprise safety climate in a given organization.

Questionnaire Development

The initial version of the survey instrument was constructed by the Patient Safety Center of Inquiry at the VA Palo Alto Health Care System under the direction of Dr. David Gaba. It was adapted with permission from five existing

survey instruments (Augustine et al. 1998; Ciavarelli, Figlock, and Sengupta 1999; Libuser and Roberts 1998; Gaba, Howard, and Jump 1994; Helmreich and Schaefer 1994). Items from each were reviewed and modified for application to hospitals. Additional questions were generated where gaps became apparent upon comparison with the 16 safety climate themes. Throughout this process, we applied several guiding assumptions: (a) that patient outcomes are the product of care delivered by many health professionals operating across multiple work areas; (b) that individual values, beliefs, and behaviors become acculturated over time through daily interactions within work units; and (c) that these norms are in turn influenced by institutional policies, procedures, and decisions. Consequently, individual survey items asked subjects to consider safety-related issues at three levels: individual, unit, and the overall organization. Items forming the core of the pilot questionnaire were consolidated during team meetings.

The initial instrument consisted of 122 items, most of which used a five-point Likert response scale ranging from strongly disagree to strongly agree with a midpoint labeled "neither agree nor disagree." A small number of items utilized a yes/no/uncertain response set, and others a frequency scale (always, frequently, sometimes, rarely, never). Three open-ended response questions were included, as were 12 questions about respondents' demographic characteristics.

Pilot Testing and Revision

This long-form survey instrument was tested in two pilot studies in 1999 and 2000. In addition, abridged versions of the survey instrument were used in two other pilot studies in 2001 and 2002 (Gaba and Park 2000; Singer 2003; Singer et al. 2003). The major results of these pilot efforts with respect to the PSCHO survey can be summarized as follows. First, we reduced the length of the instrument to 38 items over several administrations in response to very low initial response rates. This was accomplished primarily by dropping one or more items in those cases where high intercorrelations were observed. Other items were deleted or modified to clarify their meaning in response to comments indicating potential ambiguity in item interpretation or to increase variance in response. Third, we added two items to address additional gaps, including training in teamwork and emphasis on patient safety with new hires. Fourth, we added a "not applicable" response option given relatively high percentages of missing data on some items regarding practices relevant primarily in clinical domains about which nonclinical personnel were generally

not aware. Finally, we revised all remaining items to be compatible with a five-point Likert response scale. In creating the short-form survey, we purposefully retained more items derived from the Naval Command Assessment Tool to facilitate longitudinal comparison of safety climate between hospital personnel and navy aviators (Gaba et al. 2003). This process resulted in a 38-item survey (Appendix A) containing questions related to all of the original 16 HRO topics. The survey also contains six demographic and background questions.

Pilot efforts focused on the feasibility of implementing safety climate surveys and on obtaining comparative information across hospitals and did not include a thorough examination of instrument reliability and validity. Building on this prior work, the goal of the present study was to evaluate the instrument in a systematic way in order to finalize the content of the survey and to validate it as a measure that captures the key elements of safety climate as described by HRO theory. A secondary goal was to structure that measure so that the specific dimensions represented would provide hospital managers with practical feedback that they could use to facilitate their patient safety improvement efforts.

PSYCHOMETRIC EVALUATION

Sample

We tested the psychometric properties of the 38-item PSCHO instrument using data obtained through surveying personnel from 105 hospitals representing three size categories and four regions of the United States. Participating hospitals became members of a national Patient Safety Consortium (PSC) formed for this research project and administered by the Center for Health Policy & Center for Primary Care and Outcomes Research at Stanford University (CHP/PCOR). The sample consisted of 42,249 individuals, representing all disciplines and levels of hierarchy in participating hospitals.

Survey Administration

Survey implementation followed a standard protocol similar to that in our previous studies. We distributed surveys up to three times in waves spaced approximately 6 weeks apart to 100 percent of each hospital's active, hospital-based physicians, 100 percent of senior executives (defined as department head or above), and a 10 percent random sample of all other hospital employees. Surveys were sent to a hospital liaison at each facility who distributed

the individually addressed envelopes through interoffice mail. Packets included a cover letter explaining the project and requesting participation, a survey instrument, a U.S. mail postage prepaid reply envelope, and a prepaid Questionnaire Completion Notification postcard, which recipients were instructed to complete and return separately from their completed questionnaire. The postcard avoided repeat mailings to respondents while maintaining their anonymity. Implementation commenced in March 2004 and concluded in May 2005. The Institutional Review Boards of all participating institutions granted approval to conduct this survey.

We obtained a total of 21,496 completed questionnaires from hospital personnel. After accounting for individuals who no longer worked at the hospital, this represented a 51 percent overall response rate. Hospital response rates ranged from 17 to 100 percent. Responses among physicians (28 percent) were lower than among senior managers (74 percent) and other personnel (66 percent). The percent of missing responses for individual items was very low.

Analysis Strategy

To begin, we randomly split the pool of respondents in half to create a derivation sample ($n = 10,748$) and a validation sample ($n = 10,748$). This strategy provided adequate power for all analyses. After splitting the sample, analyses proceeded in two phases. First, we applied exploratory factor analysis (EFA) to the derivation sample in order to identify the underlying constructs in the data. We then used the data in the remaining half of the sample to validate that set of constructs by applying multitrait analysis (MTA) to scales based on the item-to-factor loadings from phase one.

The EFA in phase one involved initial extraction of factors using principal components analysis, followed by varimax rotation, to identify a simple structure that included coherent and relatively independent groups of items (Tabachnick and Fidell 1983). We experimented with four exploratory analyses in our effort to identify the most logical set of dimensions. These differed in the items that were included in the EFA. The first analysis included all survey items; subsequent analyses excluded sets of items that appeared qualitatively different from the other items and therefore might warrant separate consideration. In particular, (a) the second EFA excluded six "personal" items that focused on the individual respondents' personal reactions to patient safety-related situations (e.g., "Telling others about my mistakes is embarrassing"); (b) the third analysis excluded three items in which respondents were asked to report on the occurrence of actual safety-related events in which they were

involved or that they witnessed (e.g., “In the last year, I have done something that was not safe for the patient”); and (c) the fourth analysis excluded both the personal and report-type items. In each analysis, we used the standard eigenvalues-greater-than-one decision rule (Kaiser 1960) and Cattell’s (1966) criteria for identifying distinct breaks in the slope of plotted factors against their eigenvalues to determine the number of factors to extract. Investigators convened by teleconference to discuss the results of each EFA and collectively to select a factor structure on which to perform the validation analysis. This choice was driven by both statistical and practical merit, i.e., we sought a factor structure that could easily be understood and acted upon.

To test the replicability of the proposed dimensions identified in the derivation sample, we applied MTA to the data in the validation sample. MTA is based on the multitrait/multimethod strategy (Campbell and Fiske 1959) and assesses both the reliability and validity of proposed multi-item scales. Reliability is established by application of internal reliability consistency criteria, e.g., Cronbach’s $\alpha \geq 0.70$. Validity is established by comparing the strength of the correlation of each item with its assigned scale (convergent validity) versus the correlation of the item with all other proposed scales (discriminant validity) (Hays and Hayashi 1990). Proposed scale scores were computed by averaging with equal weight those items with factor loading coefficients > 0.40 for a given factor. Items with loadings > 0.40 on two different factors were provisionally assigned to the factor with the highest loading.

RESULTS

Table 1 presents self-reported background characteristics of respondents. The success of the randomization of respondents into derivation and validation samples was supported by finding no significant differences between the two groups on the background characteristics presented.

EFA Results

Applied to the full set of 38 items, Kaiser’s eigenvalues-greater-than-one criteria suggested retention of seven factors. Cattell’s scree test revealed noteworthy inflection points in the plot of eigenvalues after four and seven factors. Based on the convergence and subjective evaluation of the practical value of these results, we elected to retain seven factors. The resulting rotated simple structure is reported in Table 2.

Table 1: Demographic Characteristics of Survey Respondents (Percentages)

<i>Variable/Category</i>	<i>Total Sample (n = 21,496)</i>	<i>Derivation Sample (n = 10,748)</i>	<i>Validation Sample (n = 10,748)</i>	<i>Difference: p-Value (Two-Tailed)*</i>
Clinician: yes	60.4	60.2	60.6	0.63
Gender: male	35.9	36.4	35.5	0.17
Respondent age				0.6
18–25	3.5	3.6	3.5	
26–30	5.7	5.5	5.9	
31–40	21.9	21.7	22.1	
41–50	33.6	33.7	33.5	
51–60	27.3	27.3	27.3	
> 60	7.9	8.1	7.6	
Facility Size				0.57
Small (0–99 beds)	17.9	18.2	17.7	
Medium (100–249 beds)	24.4	24.5	24.4	
Large (250+ beds)	57.6	57.3	57.9	

*Probabilities are for χ^2 or, in the case of gender and clinician (yes/no), Fisher's exact test.

Considering factor loadings of 0.40 or greater, the first factor consisted of 13 items related to hospital-wide safety-related issues, such as senior managers' awareness of and engagement with safety hazards at the front lines, availability of organizational resources supportive of safety, and the quality of communication related to safety issues up and down the chain of command. The second factor consisted of seven items that focused on unit-level norms related to safety, such as the intensity of peer pressure for safety, and the prevalence of violations of standard procedures and safety rules. The third factor (four items) concerned the alignment of rewards, recognition, and training with the goal of patient safety. The five items that constituted the fourth factor concerned personal feelings (embarrassment, shame, concerns about appearing incompetent) that might deter behaviors such as revealing mistakes or asking for help that would promote patient safety. The fifth factor consisted of three report-type items asking about the frequency of the respondent's own unsafe practices or those the respondent may have witnessed among coworkers. Factor 6 consisted of two items related to respondents' self-awareness of factors (fatigue, personal problems) that could affect their performance. The two items that constituted the seventh factor asked about the likelihood of punishment related to mistakes and patient care errors, and thus appeared to assess respondents' level of fear of negative consequences for being associated with patient safety incidents.

Two items emerged from the EFA as requiring additional consideration. One item (Q16) exhibited no loadings above 0.30 on any factor, and its

Table 2: Exploratory Factor Analysis Results

Questionnaire Item (R) indicates item was reverse-coded before being included in factor analysis	Factor Loadings ≥ 0.40						
	1	2	3	4	5	6	7
Factor 1							
Q5	0.76						
Q8	0.76						
Q1	0.72						
Q19	0.66						
Q38	0.65						
Q11	0.64						
Q17	0.63						
Q4	0.61	0.48					
Q30	0.61						
Q28	0.57						
Q29	0.52						
Q22	0.49						
Q6	0.44						

Continued

Table 2: Continued

<i>Questionnaire Item</i> <i>(R) indicates item was reverse-coded before being included in factor analysis</i>		<i>Factor Loadings ≥ 0.40</i>						
		1	2	3	4	5	6	7
Factor 2								
Q3	In my unit, disregarding policy and procedures is rare		0.66					
Q7	In my unit, anyone who intentionally violates standard procedures or safety rules is swiftly corrected		0.61					
Q12	My unit does a good job managing risks to ensure patient safety	0.48	0.58					
Q9	My unit takes the time to identify and assess risks to ensure patient safety	0.47	0.57					
Q2	My unit emphasizes patient safety procedures and goals to new hires in their first 6 months of work	0.45	0.57					
Q32	In my unit, there is significant peer pressure to discourage unsafe patient care		0.48					
Q27	Individuals in my unit are willing to report behavior that is unsafe for patient care		0.45					
Factor 3								
Q35	I am rewarded for taking quick action to identify a serious mistake			0.7				
Q14	My unit recognizes individual safety achievement through rewards and incentives			0.65				
Q37	My unit provides training on teamwork in order to improve patient care performance and safety			0.55				
Q24	My unit follows a specific process to review performance against defined training goals			0.43				
Factor 4								
Q13	If I make a mistake that has significant consequences and nobody notices, I do not tell anyone about it (R)				0.73			
Q10	Asking for help is a sign of incompetence (R)				0.62			
Q15	Telling others about my mistakes is embarrassing (R)				0.51			

Q21	I will suffer negative consequences if I report a patient safety problem (<i>R</i>)	0.51			
Q36	I have made significant errors in my work that I attribute to my own fatigue (<i>R</i>)	0.5			
Factor 5					
Q25	In the last year, I have witnessed a coworker do something that appeared to me to be unsafe for the patient (<i>R</i>)	0.78			
Q33	I have never witnessed a coworker do something that appeared to me to be unsafe patient care	0.77			
Q34	In the last year, I have done something that was not safe for the patient (<i>R</i>)	0.49			
Factor 6					
Q18	I am less effective at work when I am fatigued	0.78			
Q20	Personal problems can adversely affect my performance	0.71			
Factor 7					
Q26	If people find out that I made a mistake, I will be disciplined (<i>R</i>)				0.81
Q31	Clinicians who make serious mistakes are usually punished (<i>R</i>)				0.76
Orphan items					
Q16	It is hard for doctors or nurses to hide serious mistakes	0.07	0.21	0.3	-0.03
Q23	I have learned how to do my own job better by learning about mistakes made by my coworkers	0.07	0.16	0.26	0.07
				0.2	-0.29
				0.13	0.35
					-0.02

content appeared already to be adequately represented by the other items in that item's highest-loading factor. We therefore elected to drop Q16 from our further analyses. The other notable item (Q23) did have one relatively high loading (0.35) and a reasonable conceptual fit with the other items in that scale, which also focused on factors that affect an individual's performance. However, this item alone addressed individuals' ability to learn from the mistakes of others, an important facet of safety culture. We therefore elected to relax slightly the 0.40 factor loading coefficient criteria and retain Q23 as we moved forward with our analyses.

When we applied EFA to the three alternate subsets of items described above (omitting the "personal" items in one analysis, the "report" items in another, and both in the third), the resulting factor structures did not improve the overall interpretability of the measure. We thus saw no empirical justification for limiting the content of the instrument at this point and a possible liability in the premature foreclosure of content. We therefore elected to retain 37 items in the second validation phase of the analysis.

We modified the seven-factor EFA model in one way before conducting the MTA. The first factor was long (13 items) and had a very high level of internal consistency reliability (0.90). These characteristics can be indicative of redundancy within a scale. On the other hand, item homogeneity (the average of all item covariances within the scale) was relatively modest (0.42), suggesting instead the possibility of latent subscales. Upon careful review of factor 1, we identified three potential subscales related to safety culture at the facility overall: (1) senior managers' engagement (seven items); (2) organizational resources for safety (three items); and (3) overall emphasis on safety (three items). We tested empirically this more differentiated version of factor 1 along with the six other scales identified by the EFA in the next phase of the analysis.

MTA Results

We created nine summated rating scales using data from the validation sample, weighing each item equally in the computation of the scale score. In order to be included in the analysis, a respondent was required to have answered at least half of the items in each proposed scale; a total of 8,535 (79.4 percent) of the respondents in the validation sample met this criterion.

Key MTA results are summarized in Table 3. The first four columns report the question number, text, mean, and standard deviations for all items. The remainder of the table consists of item-to-scale correlations. The highlighted coefficients are the corrected correlations between each item and the

Table 3: Multi-Trait Analysis Results: Item Descriptive Statistics and Item-to-Scale Correlations

Scale/Items		Mean	SD	Factor 1.1	Factor 1.2	Factor 1.3	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
<i>(R) indicates that the item was reverse-coded before being included in scale score</i>												
Factor 1.1 (scale 1)												
Q5	Senior management provides a climate that promotes patient safety	4	1.01	0.77	0.56	0.6	0.62	0.5	0.25	0.3	0.02	0
Q8	Senior management has a clear picture of the risk associated with patient care	3.82	1.11	0.75	0.52	0.54	0.56	0.47	0.24	0.29	0	-0.01
Q19	Senior management considers patient safety when program changes are discussed	3.7	0.96	0.66	0.44	0.52	0.51	0.5	0.19	0.23	0.07	-0.02
Q11	Senior management has a good idea of the kinds of mistakes that actually occur in this facility	3.57	1.07	0.67	0.45	0.48	0.53	0.48	0.21	0.24	0.03	-0.01
Q17	Good communication flow exists up the chain of command regarding patient safety issues	3.65	1.07	0.71	0.5	0.57	0.6	0.55	0.24	0.3	0.04	-0.01
Q4	Patient safety decisions are made at the proper level by the most qualified people	4.01	0.96	0.68	0.49	0.52	0.63	0.46	0.23	0.3	0.01	-0.03
Q6	Reporting a patient safety problem will not result in negative repercussions for the persons reporting it	4.18	0.93	0.53	0.35	0.41	0.5	0.34	0.29	0.2	0.05	0.09
Factor 1.2 (Scale 2)												
Q1	I am provided with adequate resources (personnel, budget, and equipment) to provide safe patient care	3.83	1.11	0.61	0.51	0.53	0.52	0.41	0.2	0.27	-0.02	0
Q30	I have enough time to complete patient care tasks safely	3.57	1.05	0.48	0.53	0.46	0.41	0.37	0.22	0.28	-0.07	0
Q29	Loss of experienced personnel has negatively affected my ability to provide high-quality patient care (R)	3.04	1.27	0.36	0.43	0.4	0.29	0.25	0.27	0.29	-0.19	0.07
Factor 1.3 (scale 3)												
Q38	Overall, the level of patient safety at this facility is improving	3.86	0.92	0.64	0.52	0.55	0.56	0.54	0.24	0.29	0.02	-0.02
Q22	Compared with other facilities in the area, this facility cares more about the quality of patient care it provides	3.78	1.01	0.49	0.36	0.46	0.44	0.4	0.18	0.2	0.06	-0.01
Q28	I am asked to cut corners to get the job done (R)	3.83	1.15	0.45	0.49	0.39	0.35	0.3	0.36	0.29	-0.1	0.07

Continued

Table 3: Continued

Scale/Items	Mean	SD	Factor 1.1	Factor 1.2	Factor 1.3	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
<i>(R) indicates that the item was reverse-coded before being included in scale score</i>											
Factor 2 (scale 4)											
Q3 In my unit, disregarding policy and procedures is rare	4.02	0.99	0.48	0.34	0.37	0.56	0.34	0.2	0.27	0.02	-0.02
Q7 In my unit, anyone who intentionally violates standard procedures or safety rules is swiftly corrected	3.87	0.97	0.56	0.37	0.42	0.61	0.46	0.23	0.3	0.02	-0.06
Q12 My unit does a good job managing risks to ensure patient safety	4.1	0.82	0.63	0.47	0.52	0.68	0.46	0.28	0.32	0.01	-0.02
Q9 My unit takes the time to identify and assess risks to ensure patient safety	4.04	0.88	0.63	0.44	0.49	0.66	0.49	0.27	0.27	0.03	-0.02
Q2 My unit emphasizes patient safety procedures and goals to new hires in their first 6 months of work	4.07	0.93	0.59	0.45	0.48	0.61	0.48	0.24	0.25	0.02	-0.01
Q32 In my unit, there is significant peer pressure to discourage unsafe patient care	3.56	1.07	0.26	0.15	0.2	0.35	0.24	0.07	0.07	0.14	0.01
Q27 Individuals in my unit are willing to report behavior that is unsafe for patient care	3.94	0.86	0.43	0.32	0.37	0.49	0.36	0.21	0.22	0.06	-0.04
Factor 3 (scale 5)											
Q35 I am rewarded for taking quick action to identify a serious mistake	3.05	1.07	0.41	0.29	0.35	0.36	0.51	0.11	0.14	0.08	-0.02
Q14 My unit recognizes individual safety achievement through rewards and incentives	2.74	1.16	0.44	0.31	0.35	0.38	0.57	0.07	0.2	0	-0.08
Q37 My unit provides training on teamwork in order to improve patient care performance and safety	3.44	1.16	0.5	0.4	0.48	0.49	0.57	0.2	0.25	-0.03	-0.05
Q24 My unit follows a specific process to review performance against defined training goals	3.58	0.96	0.48	0.32	0.39	0.49	0.5	0.18	0.2	0.07	-0.05
Factor 4 (scale 6)											
Q13 If I make a mistake that has significant consequences and nobody notices, I do not tell anyone about it (R)	4.39	0.83	0.16	0.15	0.2	0.2	0.09	0.43	0.14	-0.04	0.1
Q10 Asking for help is a sign of incompetence (R)	4.48	0.91	0.19	0.16	0.22	0.22	0.1	0.38	0.11	-0.02	0.15
Q15 Telling others about my mistakes is embarrassing (R)	3.23	1.18	0.08	0.1	0.12	0.06	0.09	0.26	0.15	-0.18	0.05
Q21 I will suffer negative consequences if I report a patient safety problem (R)	4.07	0.93	0.35	0.27	0.35	0.31	0.19	0.38	0.18	-0.04	0.2
Q36 I have made significant errors in my work that I attribute to my own fatigue (R)	3.94	0.98	0.17	0.26	0.22	0.18	0.1	0.3	0.33	-0.25	0.02

Factor 5 (scale 7)												
Q25	In the last year, I have witnessed a coworker do something that appeared to me to be unsafe for the patient (<i>R</i>)	3.27	1.22	0.3	0.32	0.31	0.31	0.23	0.26	0.55	-0.15	0.04
Q33	I have never witnessed a coworker do something that appeared to me to be unsafe patient care	2.93	1.22	0.29	0.27	0.24	0.28	0.26	0.11	0.51	-0.11	-0.08
Q34	In the last year, I have done something that was not safe for the patient (<i>R</i>)	3.94	1.08	0.2	0.23	0.23	0.2	0.12	0.34	0.37	-0.16	0
Factor 6 (scale 8)												
Q18	I am less effective at work when I am fatigued	4.06	0.93	-0.01	-0.14	-0.06	0.02	-0.04	-0.12	-0.11	0.39	0.06
Q20	Personal problems can adversely affect my performance	3.42	1.13	-0.01	-0.12	-0.06	-0.01	0	-0.2	-0.15	0.37	0.01
Q23	I have learned how to do my own job better by learning about mistakes made by my coworkers	3.8	0.89	0.13	0	0.11	0.14	0.12	-0.05	-0.11	0.2	0.03
Factor 7 (scale 9)												
Q26	If people find out that I made a mistake, I will be disciplined (<i>R</i>)	3.18	1.09	0.09	0.09	0.09	0.05	0.01	0.19	0.05	0.04	0.44
Q31	Clinicians who make serious mistakes are usually punished (<i>R</i>)	3.13	1.05	-0.09	-0.04	-0.05	-0.11	-0.13	0.08	-0.08	0.04	0.44

Table 4: Reliability Estimates and Inter-Scale Correlations

<i>Scale</i>	1	2	3	4	5	6	7	8	9
1 Senior Managers' engagement	(0.89)								
2 Organizational resources for safety	0.61	(0.67)							
3 Overall emphasis on safety	0.67	0.59	(0.65)						
4 Unit safety norms	0.73	0.51	0.58	(0.82)					
5 Unit recognition and support for safety efforts	0.61	0.44	0.53	0.57	(0.74)				
6 Fear of shame	0.30	0.30	0.35	0.30	0.19	(0.58)			
7 Provision of safe care	0.35	0.36	0.34	0.34	0.27	0.30	(0.66)		
8 Learning	0.04	(0.13)	(0.01)	0.06	0.04	(0.18)	(0.18)	(0.50)	
9 Fear of blame	0.00	0.03	0.02	(0.03)	(0.07)	0.16	(0.02)	0.05	(0.61)

remaining items in its hypothesized scale, a measure of convergent validity. Comparing these correlations with others in the same row indicates the discriminant validity of each item, that is, the extent to which the item measures the hypothesized dimension of patient safety rather than other dimensions.

The pattern of convergent and divergent correlations presented constitutes strong evidence of the reliability and validity of the hypothesized scales. Convergent item–scale correlations were substantial in magnitude, ranging from 0.20 to 0.77 across the nine proposed dimensions (median 0.51). Each item in a Likert scale should carry a substantial and roughly equal amount of information about the construct in question (Kerlinger 1973; Ware et al. 1997), as indicated by correlations of 0.40 or higher between an item and its overall scale score (adjusted for overlap). In our study, this “item internal consistency” criterion was met by 73.0 percent of the items.

Examination of the correlations between each item and its hypothesized scale in contrast to other scales revealed good item discriminant validity. For example, the first row of Table 3 (Q5) shows a significantly higher correlation between the item and its hypothesized scale (0.77) in contrast to other scales (0.00–0.62). Correlations between items and their hypothesized scales were significantly higher than correlations with any other scale in 283 of 296 comparisons (95.6 percent), and were higher although not significantly so in four additional comparisons. Thus, the overall discriminant validity quotient for the proposed set of items and scales was 97 percent.

Additional data related to the reliability and validity of the scales can be found in Table 4, which presents the correlations among the scales (off-diagonal entries) and the scale internal reliability consistency estimates (diagonal entries). The pattern of relationships observed is generally consistent with our

expectations for valid measures of related yet distinct aspects of safety climate. Specifically, the correlations range from 0.00 to 0.73 (absolute value) with a mean of 0.29. The α coefficients (diagonal entries) are considerably higher than the interscale correlations (off-diagonal entries), with only one exception out of 72 comparisons. If this were not the case, the scales could be said to be interchangeable and as such not measures of distinguishable aspects of hospital personnel's perceptions.

DISCUSSION

Our empirical results support the integrity of a nine-dimension model of hospital safety culture (see Figure 1). Three of these dimensions are organizational factors, two are work unit factors, three are individual factors, and one factor relates to report-type questions about the actual incidence of unsafe care. The organizational factors include *senior managers' engagement* in patient safety, which consists of seven items whose endorsement indicates that senior managers accurately understand current safety issues in their facility and take supportive action when necessary, but also appreciate that those best qualified to solve safety issues would often be those on the frontlines of patient care. We label the second organizational factor *organizational resources for patient safety*. Its three items elicit perceptions regarding the adequacy of personnel, time, equipment, and other resources necessary to provide safe patient care. The third organizational factor consists of three items related to the *overall level of emphasis on patient safety* at a facility and whether or not the respondent feels that safety is improving there.

Figure 1: Theoretical Model.

Among work unit dimensions, one consists of seven items regarding *unit norms for patient safety*. Endorsement of these items would characterize the immediate work environment as one in which safety issues are proactively assessed and addressed, patient safety is a genuine and pervasive value among staff, and concern for safety defines the norms of socially acceptable behavior. The second unit factor is *unit recognition and support for safety efforts*. It consists of four items that ask about the extent to which (a) actions that promoted safe patient care are explicitly acknowledged, and (b) patient safety standards are formally used in training and evaluation of performance. These two work unit factors are complementary in that the former is concerned with informal, less conscious prosafety social pressures, whereas the latter focuses on the formal and more explicit ways in which concern for patient safety guides staff behavior.

We identify three individual factors. The first, which we label *fear of shame*, consists of five items that are concerned with the respondents' level of comfort admitting to mistakes and gaps in knowledge and seeking help. Our hypothesis is that individuals who feel embarrassed or concerned in this regard would be less likely to ask for assistance in situations where they are not entirely sure about what to do, and that this in turn could have a negative impact on patient safety. The second individual factor, *fear of blame*, consists of two items that focus on respondents' perception that revealing mistakes would result in discipline and punishment. The final individual factor we label *learning* and self-awareness of safety risks. It consists of three items related to respondents' knowledge of the potential to learn from others to reduce future errors and the relationship between certain personal factors (fatigue, personal problems) and their potential impact on patient safety. Finally, three items ask respondents whether they had witnessed or been directly involved in the *provision of unsafe care*.

Overall, empirical support for this model is strong. Evidence for discriminant validity is excellent, and that for convergent validity and reliability is good, with six of nine scales demonstrating Cronbach's α levels at or near the criterion for group comparisons. In addition, the scales are highly consistent with concepts that appear in the HRO literature (Appendix C compares survey dimensions with theoretical topics initially identified). The hierarchical segmentation of factors into tiers related to (a) the organization as a whole, (b) the immediate work unit, and (c) the individual may be particularly informative, suggesting the importance of approaches to patient safety that simultaneously address all three levels.

Items and dimensions covered in the PSCHO survey make it one of the most comprehensive yet parsimonious patient safety climate surveys available

(Singla et al. 2006). Additionally, project investigators have reported results to hospital managers organized by the dimensions outlined above. They reacted very favorably to this approach, citing improved ability to focus their interventions on the basis of these factor scales.

Our testing of the instrument has limitations. Although the number of individuals who responded to the questionnaire was large ($>20,000$), the overall response rate of 51 percent leaves room for questions about selection bias among the respondents. Response was particularly low among physicians. Rates are consistent, however, with published studies of similar length in the medical literature (Jepson et al. 2005). Still, we do not know whether the respondents differed systematically from the nonrespondents in their attitudes toward safety culture. In addition, response rates varied among hospitals. Specifically, higher rates of response were positively associated with stronger hospital safety culture scores in multivariate regressions, albeit only among the organizational dimensions. The more uniformly favorable responses one would expect in this situation may have contributed to the lack of differentiation we observed within the first factor of items related to facility-wide safety climate characteristics. Elsewhere we have sought to mitigate this potential bias by recruiting hospitals from predetermined strata of safety performance (Gaba et al. 2004). We encourage future researchers to attempt replication of the scale structure reported here, particularly regarding the reliability of the proposed differentiation among organization-wide characteristics.

Other important opportunities exist to improve the instrument itself. For example, it may be possible to reduce the size of the largest scales by eliminating items less correlated with the overall scale score. In addition, the three individual dimensions—learning, fear of shame, and fear of blame—currently demonstrate relatively low internal consistency. Wider variance among individuals relative to units and organizations likely contributes to this finding. It might be possible to achieve greater reliability among these scales by replacing poorly correlated items and adding one or two items to each.

However, adequate hospital-level reliability can be achieved with current versions of these scales by increasing sample size. Using the Spearman-Brown prophecy formula to solve for the number of respondents needed to obtain hospital-level reliability of 0.70 (Hays et al. 1999) yields sample-size requirements of 123 (learning), 255 (fear of shame), and 105 (fear of blame). These are relatively high but achievable sample sizes. They contrast those required by more reliable PSCHO subscales (e.g., 51 for senior managers' engagement).

The three-item learning dimension had the lowest reliability. Currently, two items in this scale deal with the impact of specific performance shaping factors (enough sleep; personal problems), while a third item addresses learning. To achieve satisfactory reliability with smaller sample sizes will require separating items into two constructs or eliminating items related to one of them and adding items concerning the remaining one. For this instrument, which is used both for research and for operational improvement, hospital leaders suggested retaining the collective learning construct as it represents an area that they are more likely to target for improvement.

Additional survey modifications might improve the instrument. However, changes should be made judiciously to facilitate longitudinal comparison of survey results in hospitals that have previously used it and cross-industry comparisons relying on analogous instruments used in other hazardous industries. Any new items should be added to the end of the instrument to minimize differences between survey versions.

ACKNOWLEDGMENTS

The authors thank Tobias Rathgeb, Shou-tzu Lin, and Priti Shokeen for assistance with data analysis and manuscript preparation. Development and validation of this instrument was financially supported by grants from the U.S. Agency for Healthcare Research and Quality and the Veteran Administration's Health Services Research and Development Service. The first author also acknowledges fellowships provided by the Harvard Business School and the Center for Public Leadership at the Kennedy School of Government.

Disclosures: None reported.

Disclaimers: None reported.

REFERENCES

- Augustine, C., K. Weick, J. Bagian, and C. Lee. 1998. "Predispositions towards a Culture of Safety in a Large Multi-facility Health System." *Enhancing Patient Safety and Reducing Errors in Health Care Conference*, Rancho Mirage, CA, pp. 138–41. Chicago: National Patient Safety Foundation.
- Campbell, D., and D. Fiske. 1959. "Convergent and Discriminant Validity by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56: 81–105.
- Cattell, R. B. 1966. "The Meaning and Strategic Use of Factor-Analysis." In *Handbook of Multivariate Experimental Psychology*, edited by R. Cattell, pp. 174–243. Chicago: Rand McNally.

- Ciavarelli, A., R. Figlock, and K. Sengupta. 1999. "Organizational Factors in Aviation Accidents." *Naval Postgraduate School Res* 9: 4–5, 40–3.
- Colla, J., A. Bracken, L. Kinney, and W. Weeks. 2005. "Measuring Patient Safety Climate: A Review of Surveys." *Quality and Safety in Health Care* 14 (5): 364–6.
- Cooper, J., R. Blum, J. Carroll, M. Dershwitz, D. Feinstein, D. Goba, J. Morey, and A. Singla. Forthcoming. "Differences in Safety Climate between Anesthesia Groups and the Effect of a Realistic Simulation Based Training Program." Submitted for review.
- Flin, R., C. Burns, K. Mearns, S. Yule, and E. Robertson. 2006. "Measuring Safety Climate in Health Care." *Quality and Safety in Health Care* 15 (2): 109–15.
- Gaba, D. 2000. "Structural and Organizational Issues in Patient Safety: A Comparison of Health Care to Other High-Hazard Industries." *California Management Review* 43 (1): 83–102.
- Gaba, D., S. Howard, and B. Jump. 1994. "Production Pressure in the Work Environment: California Anesthesiologists' Attitudes and Experiences." *Anesthesiology* 81: 488–500.
- Gaba, D., and K. Park. 2000. "Preliminary Analysis of Hospital Safety Culture Surveys." Patient Safety Center of Inquiry, VA Palo Alto Health Care System.
- Gaba, D., S. Singer, A. Sinaiko, and J. Bowen. 2003. "Differences in Safety Climate between Hospital Personnel and Navy Aviators." *Human Factors* 45 (2): 173–85.
- Gaba, D. M., A. Rosen, M. Meterko, L. Baker, S. J. Singer, J. J. Geppert, and A. Ciavarelli. 2004. "Improving Safety Culture and Outcomes of Veteran Administration Hospitals." VA Health Services Research and Development.
- Ginsburg, L., P. G. Norton, A. Casebeer, and S. Lewis. 2005. "An Educational Intervention to Enhance Nurse Leaders' Perceptions of Patient Safety Culture." *Health Services Research* 40 (4): 997–1020.
- Hays, R., and T. Hayashi. 1990. "Beyond Internal Consistency Reliability: Rationale and User's Guide for Multitrait Analysis Program on the Microcomputer." *Behavior Research Methods, Instruments and Computers* 22 (2): 167–75.
- Hays, R., J. Shaul, V. Williams, J. Lubalin, L. Harris-Kojetin, S. Sweeny, and P. Cleary. 1999. "Psychometric Properties of the CAHPS 1.0 Survey Measures. Consumer Assessment of Health Plans Study." *Medical Care* 37 (3 suppl): MS22–31.
- Helmreich, R., and H. Schaefer. 1994. "Team Performance in the Operating Room." In *Human Error in Medicine*, edited by M. Bogner, pp. 225–53. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hofmann, D. A., and F. P. Morgeson. 1999. "Safety-Related Behavior as a Social Exchange: The Role of Perceived Organizational Support and Leader-Member Exchange." *Journal of Applied Psychology* 84 (2): 286.
- Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.
- Jepson, C., D. Asch, J. Hershey, and P. Ubel. 2005. "In a Mailed Physician Survey, Questionnaire Length Had a Threshold Effect on Response Rate." *Journal of Clinical Epidemiology* 58 (1): 103–5.

- Kaiser, H. 1960. "The Application of Electronic Computers to Factor Analysis." *Educational and Psychological Measurement* 20: 141–51.
- Kerlinger, F. 1973. *Foundations of Behavioral Research*, 2d Edition. New York: Holt, Rinehart and Winston Inc.
- Knox, G., K. Simpson, and T. Garite. 1999. "High Reliability Perinatal Units: An Approach to the Prevention of Patient Injury and Medical Malpractice Claims." *Journal of Healthcare Risk Management* 19 (2): 24–32.
- Libuser, C., and K. Roberts. 1998. "Risk Mitigation through Organizational Structure." *Annual Meeting of the Academy of Management*, San Diego, CA.
- McCarthy, D., and D. Blumenthal. 2006. "Stories from the Sharp End: Case Studies in Safety Improvement." *Milbank Quarterly* 84 (1): 165–200.
- Nieva, V., and J. Sorra. 2003. "Safety Culture Assessment: A Tool for Improving Patient Safety in Healthcare Organizations." *Journal of Quality and Safety in Healthcare* 12 (suppl II): ii17–23.
- Roberts, K. H. 1990. "Some Characteristics of One Type of High Reliability Organization." *Organization Science* 1 (2): 160–77.
- Roberts, K. H., P. D. V. Madsen, and D. Van Stralen. 2005. "A Case of the Birth and Death of a High Reliability Healthcare Organisation." *Quality and Safety in Health Care* 14 (3): 216–20.
- Roberts, K. H., and C. Tadmor. 2002. "Lessons Learned from Non-Medical Industries: The Tragedy of the USS Greenville." *Quality and Safety in Health Care* 11 (4): 355–7.
- Rochlin, G., T. La Porte, and K. Roberts. 1987. "The Self-Designing High Reliability Organization: Aircraft Carrier Flight Operations at Sea." *Naval War College Review* 42: 76–90.
- Sharek, P. 2005. "Culture of Safety at CHCA: How We Stack Up against the 'Big Boys'" Child Health Corporation of America Annual Meeting. Chicago, IL.
- Singer, S. J. 2003. "Safety Culture Survey Results from 12 Patient Safety Consortium Hospitals: 2001–2002." *Health Care Quality and Outcomes Conference*. San Francisco, CA: University of California.
- Singer, S. J., D. M. Gaba, J. J. Geppert, A. D. Sinaiko, S. K. Howard, and K. C. Park. 2003. "The Culture of Safety in California Hospitals." *Quality and Safety in Health Care* 12 (2): 112–8.
- Singla, A. K., B. T. Kitch, J. S. Weissman, and E. G. Campbell. 2006. "Assessing Patient Safety Culture: A Review and Synthesis of the Measurement Tools." *Journal of Patient Safety* 2 (3): 105–15.
- Sutcliffe, K. 2000. "A Mindful Infrastructure for Increasing Reliability." *Ambul Outreach* Spring:30–4.
- Tabachnick, B., and L. Fidell. 1983. *Using Multivariate Statistics*. New York: Harper & Row.
- Ware, J., W. Harris, B. Gandek, B. Rogers, and P. Reese. 1997. *Multitrait/Multimethod Analysis Program—Revised*. Boston: Health Assessment Lab.
- Weick, K. 1987. "Organizational Culture as a Source of High Reliability." *California Management Review* 29 (2): 112–27.

- Weick, K. E., and K. M. Sutcliffe. 2003. "Hospitals as Cultures of Entrapment: A Re-Analysis of the Bristol Royal Infirmary." *California Management Review* 45 (2): 73.
- Wilson, K., C. Burke, H. Priest, and E. Salas. 2005. "Promoting Health Care Safety through Training High Reliability Teams." *Quality and Safety in Health Care* 14 (4): 303–9.
- Yates, G., D. Bernd, S. Sayles, C. Stockmeier, G. Burke, and G. Merti. 2005. "Building and Sustaining a Systemwide Culture of Safety." *Joint Commission Journal on Quality and Patient Safety* 31 (12): 684–9.

SUPPLEMENTARY MATERIAL

The following supplementary material for this article is available:

Appendix A: Patient Safety Climate in Healthcare Organizations Survey.

Appendix B: Theoretical Topics in the Original PSCHO Survey.

Appendix C: Comparison of Survey Dimensions with Theoretical Topics.

This material is available as part of the online article from <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1475-6773.2007.00706.x> (this link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.