Safety Science 69 (2014) 37–47

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Exploring the relationship between safety culture and safety performance in U.S. nuclear power operations

Stephanie L. Morrow [a,*], G. Kenneth Koves [b], Valerie E. Barnes [a]

[a] United States Nuclear Regulatory Commission, Washington, DC, USA
[b] Institute of Nuclear Power Operations, Atlanta, GA, USA

ARTICLE INFO

Article history:
Available online 19 March 2014

Keywords:
Safety culture
Nuclear power
Safety performance
Human and organizational factors
Organizational culture
Safety climate

ABSTRACT

How do nuclear power plant workers, within a single national culture, perceive safety culture within their organizations? What is the relationship between safety culture and other indicators of safety? Is the construct of safety culture useful for predicting future plant performance? These questions were addressed in the current study using a survey administered to a sample of personnel at 97% of the nuclear power plants in the United States, resulting in 2876 responses from 63 nuclear power plant sites. Exploratory and confirmatory factor analysis revealed a multi-factor structure to the safety culture survey. For each nuclear power plant, the mean score for the total survey results and the factor means were correlated with organization-level performance indicators both concurrently and one year following the survey administration. Correlations suggested meaningful, statistically significant relationships between safety culture, as measured by the survey, and multiple nuclear power plant performance indicators. This study presents a unique look at safety culture across the United States nuclear power industry and takes a critical step toward establishing that safety culture is empirically related to safety performance.

Published by Elsevier Ltd.

1. Introduction

The term "safety culture" was first introduced to the nuclear industry as part of the International Atomic Energy Agency (IAEA) assessment of the causes of the 1986 Chernobyl accident. The International Nuclear Safety Advisory Group (INSAG), an advisory group reporting to the Director General of the IAEA, concluded that "Formal procedures, properly reviewed and approved, must be supplemented by the creation and maintenance of a nuclear safety culture" (INSAG, 1986). Although not labeled "safety culture" at the time, the U.S. Nuclear Regulatory Commission (NRC) also recognized the contribution of organizational factors to accidents in their investigation of the Three Mile Island (TMI) accident in 1979. The NRC's investigation report stated that "The one theme that runs through the conclusions we have reached is that the principal deficiencies in commercial reactor safety today are not hardware problems, they are management problems" (Rogovin, 1980). These accident investigations helped spur research in the area of safety culture to understand how shared, underlying beliefs and values in an organization may help or hinder safe performance.

In 2002, Sorensen published a critical review of the state-of-the-art of safety culture research as applied to the nuclear industry. Sorensen asserted that safety culture research cannot progress until safety culture has been defined, the characteristics or attributes of safety culture have been delineated, and a link between safety culture and safe operations has been established. It is only in the last 10 years that researchers have begun to publish more rigorous studies explicitly testing for relationships between safety culture and safety performance and reviews of the safety culture literature have begun to reach agreement around common themes in safety culture definitions and dimensions.

The primary purposes of the current study were to investigate the factors that comprise the concept of safety culture in the nuclear power industry and evaluate the relationships between these safety culture factors and other measures of organizational and safety performance. The nuclear industry collects and trends vast amounts of data gauging equipment reliability and operating performance. However, to our knowledge this study is the first comprehensive look at potential linkages between a measure of safety culture in nuclear power organizations and these other types of data. The administration of this survey provided a unique opportunity to explore how perceptions of safety culture are related to organizational-level performance measures across nearly all operating plants in the U.S. nuclear power industry.

* Corresponding author. Address: Nuclear Regulatory Commission, 11555 Rockville Pike, MailStop O-4A15A, Rockville, MD 20852-2738, USA. Tel.: +1 1(301)415 1986.
E-mail address: Stephanie.Morrow@nrc.gov (S.L. Morrow).

http://dx.doi.org/10.1016/j.ssci.2014.02.022
0925-7535/Published by Elsevier Ltd.

1.1. The underlying theory of organizational safety culture

Schein's (1992, 2010) model of organizational culture is perhaps the most widely-adopted model in both nuclear and non-nuclear domains. Schein defines organizational culture as "a pattern of shared basic assumptions that the group learned as it solved its problems of external adaptation and internal integration, that has worked well enough to be considered valid, and, therefore to be taught to new members as the correct way to perceive, think, and feel in relation to those problems" (1992, p. 12). Safety culture is generally considered to be a specific aspect of organizational culture regarding the organization's shared beliefs, values, and attitudes that contribute to ensuring safe operations.

The practical utility of assessing an organization's safety culture is that the assessment may be used as a performance indicator, in addition to more established indicators like safety management audits or analyses of events and near-misses (Guldenmund, 2000). Further, it is possible that safety culture assessments may serve as leading indicators of performance, and provide opportunities for intervention before significant events occur. Post-event investigations, like TMI and Chernobyl, have repeatedly shown that weaknesses in an organization's safety culture can create opportunities for significant adverse events.

At an individual level, the relationship between safety culture and safety performance may be best described using the Theory of Planned Behavior (Ajzen and Fishbein, 1977). Employees' beliefs about the importance of safety are shaped by the safety culture of the organization, which then influences their attitudes toward safety, perceived norms for working safely, and perceptions of control over safe working behaviors. For example, an employee working in a strong, positive safety culture is more likely to have positive attitudes toward the importance of safety, perceive that performing a job safely (e.g., by following procedures or wearing personal protective equipment) is the accepted norm, and perceive that they have more control over safety because the organization promotes and prioritizes doing the job safely over competing demands. This logic is also consistent with Neal and Griffin's (2006) exploration of safety motivation and compliance as mediators of the relationship between safety climate and safety behaviors. The organizational safety culture provides the contextual cues that the employee uses to determine whether to behave in a safe or unsafe manner while performing work. The more employees engage in unsafe behaviors the more likely an adverse event will occur because each unsafe behavior creates holes in the organization's defenses and barriers, as described by Reason's (1997) Swiss Cheese Model. The aggregated behaviors then influence or directly determine the overall performance of the organization. A weak or negative safety culture shapes employee beliefs that it is acceptable to take shortcuts or become complacent, which may degrade the safety performance of the organization over time and lead to a significant adverse event.

1.2. Relationships between safety culture and safety performance

A series of meta-analytic studies published between 2006 and 2010 significantly advanced the state of safety culture research by providing comprehensive analyses of past safety culture studies (Christian et al., 2009; Clarke, 2006; Beus et al., 2010). The studies included in these meta-analyses measured safety culture using surveys where employees were asked various questions regarding their perceptions of the extent to which their organization valued safety.

Safety performance is used as an umbrella term to refer to the various types of safety outcomes that have been used as dependent variables in safety culture studies, ranging from observed or self-reported employee safety behaviors (e.g., following procedures, wearing personal protective equipment, participating in safety meetings) to organization-level safety outcomes like accident and injury rates. The studies hypothesized that there should be a relationship between measures of safety culture and other indicators of safety performance.

The results from the meta-analytic studies found consistent evidence of a statistically significant linear relationship between safety culture and accidents/injuries, ranging from a correlation of −.22 to −.39 ($p < .05$), and even larger statistically significant correlations between safety culture and employees' self-reported safety behaviors, ranging from .43 to .61 ($p < .05$). Using Cohen's (1988) labels, the relationship between safety culture and safety performance appears to be a medium effect, and the relationship between safety culture and safety behaviors appears to be a large effect. Effect sizes can also be interpreted in terms of the percent of variance shared by two variables. In the case of correlation analyses, the square of the correlation coefficient represents the percent of shared variance. The results of the meta-analyses suggest that, overall, safety culture may account for 5–15% of the variance in an organization's accident and injury rates, and 18–37% of the variance in employees' safety behaviors.

Longitudinal studies have reached mixed conclusions about the relationship between safety culture and safety performance. Mearns et al. (2003) found some support for a relationship between safety climate at the organizational level and safety performance in offshore oil and gas installations. However, the study suffered from a lack of statistical power when the data were analyzed at the organization level because only 13 installations were included in the study. Correlations between the organizations' safety climate survey results and measures of accident and incident rates were in the expected directions, but were not statistically significant, and the effects were much stronger in time one as compared to time two.

Neal and Griffin (2006) found support for group-level safety climate as a predictor of safety motivation, which subsequently influenced safety behaviors. The study tested these relationships over a five-year time period, focusing on the causal chain linking safety climate to safety performance. Safety behaviors were also a significant predictor of accident rates in the following year, but safety climate did not predict accident rates in the following year. Neal and Griffin argue that this is because safety climate is a distal predictor of safety performance, whereas safety behavior is a more proximal predictor. Zohar (2000) tested a group-level model of safety climate within a single manufacturing organization and found support for safety climate as a predictor of microaccidents (i.e., minor on-the-job injuries requiring medical attention) over a five-month period. Zohar's use of microaccidents as an objective measure of safety was an important contribution to the literature, but the examination of group-level climate makes it difficult to generalize these findings to organizational safety culture. The current study expands on previous longitudinal research by examining safety culture at the organizational level across a large number of organizations in a single industry, and using a diverse set of objective safety performance measures.

1.3. Safety performance in the nuclear industry

In high reliability industries, like nuclear power, accidents are extremely rare occurrences. As a result, less significant events that occur more frequently are relied upon as indicators of potentially degrading performance. These more frequent events are also more conducive to quantitative data analysis because there are more data points and more variability across organizations. Although there are many reasons a plant could see declines in their safety performance measures, some of those reasons could theoretically relate to safety culture. For instance, it is possible that declines in safety performance could be attributable, in part, to an emerging pattern of management decisions that emphasize productivity over safety, growing complacency among employees with regard to following procedures or using human performance tools (e.g., STAR: Stop, Think, Act, Review), or other systemic weaknesses in the organization's safety culture.

The NRC collects and monitors many different types of data related to the performance of operating nuclear power plants. The data are evaluated regularly to identify indications of degrading performance and determine appropriate regulatory responses. The safety performance measures chosen for use in this study were based on performance indicators from the NRC's industry trends program (ITP), performance indicators and inspection reports associated with the NRC's reactor oversight process (ROP), and reports made to the NRC regarding nuclear safety concerns and other matters at nuclear power plants, referred to as allegations. In the following paragraphs we provide an overview of some of the safety performance data used in the nuclear industry. The methods section of the paper includes a detailed description of the specific variables included in this study as indicators of safety performance.

The ITP offers insights about plant performance using information provided to the NRC from each nuclear power plant. One of the objectives of the ITP is to "collect and monitor industry-wide data that can be used to assess whether the nuclear industry is maintaining the safety performance of operating plants" (IMC-0313, NRC, 2008). The data are consolidated annually to create indicators representing aspects of a site's performance.

The ROP is the NRC's regulatory framework for overseeing the safe operation of commercial nuclear power plants. The ROP is designed to focus on those plant activities that are most important to safety, and uses inspection findings and performance indicators for on-going monitoring of each plant's performance (IMC-0308, Attachment 1; NRC, 2006). During inspections, NRC inspectors document findings that may indicate a deficiency in performance. These findings can range in severity and significance to safety and can be issued for incidents ranging from not following plant procedures to directly violating a regulation. Each quarter, every plant is assigned a status in the ROP action matrix based on their most recent performance indicators and inspection findings. A plant's status in the action matrix determines the level of NRC oversight of the plant, including supplemental inspections and pertinent regulatory actions ranging from management meetings up to and including orders for plant shutdown. The columns in the ROP action matrix, ordered by increasing regulatory response and degraded performance, are: (1) licensee response, (2) regulatory response, (3) degraded cornerstone, (4) multiple/repetitive degraded cornerstone, and (5) unacceptable performance.

The ROP also includes three "cross-cutting areas" that are categorizations used to identify performance deficiencies that cut across many different cornerstones of safe plant operation. These areas are human performance (HP), problem identification and resolution (PI&R), and safety conscious work environment (SCWE). Each cross-cutting area is further divided into various components, and the components are sub-divided into aspects. Inspectors can assign an aspect from one of the cross-cutting areas to inspection findings if they determine that the aspect characterizes the most significant contributor to the performance deficiency cited in the finding (IMC-0310; NRC, 2011a). For example, if an inspector cites a nuclear power plant with a finding because a worker did not have an approved work order while completing a maintenance activity, then the inspector may also attribute this finding to a more widespread concern about the quality of the plant's work control processes. As a result, the inspector may assign to the finding a cross-cutting aspect under the human performance cross-cutting area in the ROP. This allows the NRC to track issues that may indicate potential safety deficiencies with common underlying themes. The cross-cutting areas, components, and aspects are considered to be conceptually related to safety culture because they intend to identify potential widespread issues at the plant rather than just isolated incidents.

During mid-cycle and end-of-cycle assessment meetings, the NRC reviews the aspects tagged to inspection findings within each cross-cutting area for substantive cross-cutting issues (SCCIs). According to IMC-0305 (NRC, 2011b), an SCCI is a cross-cutting theme that has been identified in one of the cross-cutting areas (HP, PI&R, or SCWE), about which the NRC staff has a concern with the licensee's scope of efforts or progress in addressing the cross-cutting theme. Cross-cutting themes in the HP and PI&R areas are assigned when multiple inspection findings (i.e., four or more) are assigned the same cross-cutting aspect within a 12-month period. In all cases, the NRC determines that an SCCI exists only if there is a concern with the plant's scope of efforts or progress in addressing the cross-cutting theme.

The NRC encourages employees working in regulated activities to raise safety concerns and maintains an allegations program to evaluate and respond to such concerns. An allegation is "a declaration, statement, or assertion of impropriety or inadequacy associated with NRC-regulated activities, the validity of which has not been established." Safety concerns may include areas like operations, maintenance, radiation protection, security, harassment, discrimination, wrongdoing, or a work environment that discourages workers from raising safety concerns. Any member of the public or any individual who is performing work at a site licensed by the NRC may report an allegation. In particular, allegations from plant personnel may be related to the plant's safety culture. For instance, a sharp increase in allegations could suggest that personnel are losing confidence in the plant's internal corrective action program or no longer feel safe raising safety concerns within the organization.

Through the Institute for Nuclear Power Operations (INPO), the power reactor industry also collects and monitors data related to their own safety and operating performance. These key performance indicators (KPIs) are evaluated on an ongoing basis to look for performance trends at the plant, fleet, and industry levels. A number of KPIs were selected for examination in this study because they were distinct from the NRC performance measures and also representative of aspects of plant performance that may be related to the safety culture of the organization, including chemistry performance, human error rate, forced loss rate, and industrial safety accident rate. Chemical impurities in power plant systems may indicate a lack of care for housekeeping and ensuring the plant is operating at peak condition. Human errors may result from not following procedures or using human performance tools to check work and ensure it is done correctly. Loss of power production due to unplanned events suggests potential issues with the plant's management systems and directly impacts profitability. Injury rates relate to concerns with personnel safety which may also relate to the overall priority of safety in the plant.

1.4. Hypotheses

The current study provides a unique contribution to the safety culture literature by examining the relationship between safety culture and a diverse set of performance measures that focus on the overall operational safety of a nuclear power plant. This is in contrast to previous studies that have only measured outcomes related to personal safety (e.g., injury rates) or individual safety behaviors (e.g., wearing personal protective equipment); and studies that have lacked statistical power to detect effects at the organizational level of analysis. Although the impact of safety culture on global measures of operational performance may be more distal, this study aims to establish a much-needed link between safety culture and safety performance at the organizational level of analysis.

Given the results of previous studies, we expected to find small to medium effect sizes between safety culture and safety performance, with correlation coefficients of .20 to .30. We also expected the safety culture survey to be negatively related to the NRC and INPO performance measures, such that higher scores on the safety culture survey are associated with lower values on the NRC and INPO safety performance measures (e.g., fewer scrams, inspection findings, and allegations). The following specific hypotheses are proposed.

H1. Organizational safety culture will be negatively related to the following concurrent measures of safety performance:

H1a. Unplanned scrams.

H1b. NRC allegations.

H1c. Reactor oversight process (ROP) cross-cutting aspects.

H1d. Human performance cross-cutting area.

H1e. Problem identification and resolution cross-cutting area.

H1f. Substantive cross-cutting issues (SCCIs).

H1g. ROP action matrix oversight.

H1h. Chemistry performance index.

H1i. Human performance error rate.

H1j. Forced loss rate.

H1k. Industrial safety accident rate.

H2. Organizational safety culture will be negatively related to the following future measures of safety performance:

H2a. Unplanned scrams.

H2b. NRC allegations.

H2c. Reactor oversight process (ROP) cross-cutting aspects.

H2d. Human performance cross-cutting area.

H2e. Problem identification and resolution cross-cutting area.

H2f. Substantive cross-cutting issues (SCCIs).

H2g. ROP action matrix oversight.

H2h. Chemistry performance index.

H2i. Human performance error rate.

H2j. Forced loss rate.

H2k. Industrial safety accident rate.

2. Methods

2.1. Initial survey development

The initial survey consisted of 110 items drafted from multiple sources, including INPO's Principles for a Strong Nuclear Safety Culture (INPO, 2004), IAEA's safety culture characteristics and attributes (IAEA, 2006), inputs from the NRC's inspection program, relevant items from published surveys in the safety culture research literature, and behaviors characterizing a positive safety culture developed by subject matter experts in the nuclear industry. The survey also included demographic questions asking respondents to indicate their plant, work group (e.g., operations, maintenance, engineering), and work status (permanent or contractor). Survey participants were asked to rate their degree of agreement with each statement using a 7-point Likert scale ranging from strongly disagree to strongly agree. Each item also included a "do not know/no opportunity to observe" response option.

2.2. Survey administration

Lists of personnel, including long-term contractors, were obtained from each operating nuclear power plant in the U.S., and approximately 100 individuals from each site were randomly selected to participate in the survey. The survey was administered through the internet, and e-mail invitations with a link to the survey were sent to the selected participants at each site. Senior management at each site was contacted to request that the survey be announced and participation encouraged. The survey administrator sent an average of two reminder e-mails to people who did not respond to the first invitation, until a minimum of 30 respondents from each site was obtained. Data collection began on June 14 and ended on August 11, 2010.

2.3. Participants

The total number of people invited to complete the survey over the administration period was 6333. Of those, 3031 individuals responded to the invitation for a 48% response rate. 2876 respondents provided valid answers to the majority of the survey items and their responses were retained for subsequent data analysis. The average number of respondents per site was 46. Sixty-three sites were in the sample, or 97% of the operating nuclear power plants in the U.S. Two sites were not included in the study. One site was participating in another survey in preparation for an organizational effectiveness review and requested not to participate in this survey. Another site was inadvertently omitted from the survey because of an administrative error.

In addition to the overall response rate of the survey, we were interested in obtaining a sample that was representative of the population under study (Krosnick, 1999). At least 30 employees from all but 2 nuclear power plants in the U.S. provided usable data from the survey. The demographic information collected from respondents indicates a good distribution of respondents from different workgroups within a nuclear power plant. For example, approximately 16% of respondents worked in operations, 17% in maintenance, 10% in security, 5% in systems engineering, 6% in training, 6% in radiation protection, 3% in chemistry, and 7% were contractors. These distributions are generally representative of workforce distributions in a nuclear power plant.

2.4. Factor analysis

One of the objectives of any survey development effort is to maximize content validity while maintaining parsimony and simple structure (Hinkin, 1998). Because the survey items covered diverse aspects of the construct of safety culture we conducted a multi-phased approach to establish the factor structure of the survey.

First, we randomly split the sample into two halves. We used the first half to conduct an exploratory factor analysis using principal components analysis (PCA). The PCA resulted in 9 interpretable factors. Labels were developed for each of the 9 factors based on the survey items that demonstrated high factor loadings for each factor. The items with factor loadings of .40 or greater and no major cross-loadings on other factors were judged as meaningful and representative of the construct under examination (Ford et al., 1986). Four items did not have high factor loadings on any of the 9 factors and were therefore removed from subsequent analyses. Overall, the 9 factors accounted for 58% of the variance in the survey data.

Following the exploratory factor analysis, the 110 survey items were reviewed within the context of the 9 factors identified by the PCA to assess whether items could be removed from the survey. For instance, cases where many respondents did not respond to an item (i.e., missing data), or responded by choosing the "do not know" option may indicate that respondents were confused by the item or did not find it applicable to their work. Items that have low correlations with all of the other items in the survey, with the rule of thumb being inter-item correlations less than .40 (Kim and Mueller, 1978), may not be good representations of the construct of interest (i.e., safety culture). Extremely similar items that group together under the same factor may also be unnecessarily repetitive. This review resulted in the elimination of 50 items from the survey, bringing the total number of items used in subsequent analyses to 60.

We then conducted a confirmatory factor analysis (CFA) on the second half of the sample using only the 60 items retained in the final survey. The CFA confirmed that all items had statistically significant factor loadings on their respective latent variables and the measurement model demonstrated adequate fit to the data (χ^2 (1674) = 8977.14, $p < .01$; RMSEA = .05; CFI = .88).

2.5. Within-group reliability analysis

A key underlying premise of safety culture is that it is shared among members of an organization. The concept of "relatedness" can be determined by assessing within-group reliability, or the degree to which respondents at the same site had similar responses to items on the safety culture survey. Within-group reliability is necessary to justify aggregating survey data from individuals to the site level and provides support for generalizing data from a sample of employees at a site to the entire site. One method for determining within-group reliability is by using intraclass correlations (ICCs; McGraw and Wong, 1996). In addition, we conducted an analysis of variance (ANOVA) to determine whether there was a significant amount of variability between organizations.

Two types of ICCs are relevant for determining within-group reliability: ICC(1) measures reliability among individuals in a group, and ICC(2) measures the reliability of the group mean (James, 1982). The first, referred to as ICC(1), is similar in interpretation to a traditional correlation and describes the extent to which members within a group had the same responses, in terms of agreement/disagreement, to the items in the safety culture survey. The measurement of ICC(2) is very similar in concept to measuring Cronbach's alpha for internal consistency, and uses a cutoff of .70 for acceptability (McGraw and Wong, 1996). However, instead of measuring consistency among items in a factor, the ICC(2) statistic indicates consistency in responses among members in a group and thus the stability of the mean score for the group.

The ICC(1) values were statistically significant for safety culture overall and the 9 factors, indicating that at least some portion of the variation in individual responses is shared at the site level. The ICC(2) values for safety culture overall and four of the factors exceeded .70 at all sites, indicating that the group mean scores on each of those factors are relatively stable. The results of the ANOVAs were also statistically significant (safety culture $F(62) = 3.71$, $p < .01$, management commitment $F(62) = 5.09$, $p < .01$; willingness to raise concerns $F(62) = 2.06$, $p < .01$; decision making $F(62) = 3.29$, $p < .01$; supervisor responsibility for safety $F(62) = 2.20$, $p < .01$; questioning attitude $F(62) = 2.65$, $p < .01$; safety communication $F(62) = 2.87$, $p < .01$; personal responsibility for safety $F(62) = 1.42$, $p < .05$; prioritizing safety $F(62) = 3.17$, $p < .01$; training quality $F(62) = 1.71$, $p < .01$), suggesting adequate variance in safety culture between the sites included in the study.

Altogether, these measures provide enough evidence to justify conceptualizing safety culture as an organizational-level construct and aggregating the individual-level survey data to derive a mean score for each site. Aggregating data to the site level is a necessary step to allow for comparisons between the survey data and safety performance metrics, which are traditionally collected at the unit or site level. The analyses discussed in the remainder of the paper use the mean scores on the factors and safety culture survey as a whole at each site to represent the station's safety culture.

2.6. Safety culture measures

Organizational safety culture was measured using 60 items, comprised of the following 9 factors: management commitment to safety, willingness to raise safety concerns, decision making, supervisor responsibility for safety, questioning attitude, safety communication, personal responsibility for safety, prioritizing safety, and training quality. The overall measure of safety culture demonstrated adequate internal consistency with a Cronbach's alpha of .98. Table 1 presents an overview of the 9 safety culture factors. The table includes the factor label, Cronbach's alpha, number of items, and an example item for each factor.

2.7. Safety performance measures

Unplanned scrams is based on one of the inputs to the ROP performance indicator program. It represents the number of unplanned scrams at a site per year for every 7000 h the nuclear reactor is critical. A scram is an immediate shutdown of a nuclear reactor. There can be both planned and unplanned scrams. Unplanned scrams are usually a response to unexpected or emergency conditions. In 2010, the range of values across the U.S. nuclear power industry for this variable was from 0 to 5.10 ($M = 1.05$, SD = 1.27). In 2011, the range of values was from 0 to 3 ($M = .24$, SD = .67).

NRC allegations represents the total number of allegations reported to the NRC by personnel at a site (including contractors) over the course of the year. In 2010, the range of values was from 1 to 66 ($M = 6.03$, SD = 8.48), and the range in 2011 was from 0 to 16 ($M = 4.08$, SD = 3.55).

ROP cross-cutting aspects represents the total number of inspection report findings that were assigned as being attributable to one of the aspects in any one of the ROP cross-cutting areas per year. In 2010 the range of inspection findings attributed to ROP aspects was 1–49 ($M = 11.62$, SD = 8.14), and in 2011 the range was 2–37 ($M = 9.74$, SD = 6.23).

Human performance cross-cutting area is the variable representing the subset of inspection findings where the findings were attributed to an ROP cross-cutting aspect in the area of human performance. The range of ROP aspects in the human performance area in 2010 was 1–36 ($M = 7.77$, SD = 5.38), and in 2011 the range was 0–21 ($M = 6.86$, SD = 4.28).

Problem identification and resolution cross-cutting area is the variable representing the subset of inspection findings where the findings were attributed to an aspect in the area of problem identification and resolution. The range of ROP aspects in the problem identification and resolution area in 2010 was from 0–23 ($M = 3.85$, SD = 4.29), and in 2011 the range was 0–16 ($M = 2.88$, SD = 2.73).

Substantive cross-cutting issues (SCCIs) represents the total number of outstanding SCCIs in the human performance or problem identification and resolution area at each site during the end-of-cycle assessment. No sites had SCCIs in the Safety Conscious Work Environment area during the end-of-cycle 2010 or 2011 time

Table 1
Factor labels, number of items, Cronbach's alpha, and an example item for each factor.

Factor label	# Items	Cronbach's alpha	Example item
1. Management commitment to safety	20	.96	At this station, people are routinely rewarded for identifying and reporting nuclear safety issues
2. Willingness to raise concerns	6	.90	When I make a mistake, I am not afraid to report it to my supervisor
3. Decision-making	5	.88	Decision-making at this site reflects a conservative approach to nuclear safety
4. Supervisor responsibility for safety	6	.88	My supervisor is usually available when I have a question or problem
5. Questioning attitude	6	.85	Personnel promptly identify and report conditions that can affect nuclear safety
6. Safety communication	7	.87	There is good communication about nuclear safety issues that affect my job
7. Personal responsibility for safety	3	.77	It is my responsibility to raise nuclear safety concerns
8. Prioritizing safety	4	.83	At this station, nuclear safety takes priority over production goals
9. Training quality	3	.78	Training at this site provides me with the knowledge I need to perform my job

period. In 2010, the range of values was 0–5 ($M = .25$, SD = .82), and in 2011 the range was 0–3 ($M = .24$, SD = .67).

ROP action matrix oversight is a high-level indicator reflecting overall plant performance. This variable is an indicator of sites that have elevated oversight by the NRC due to degraded performance. Each site was assigned a value from 1 to 5, based on their level of oversight within the ROP action matrix in 2010 ($M = 1.29$, SD = .66), and then also in 2011 ($M = 1.38$, SD = .87), with higher values indicating greater oversight.

Chemistry performance index is an index maintained by INPO that indicates the concentration of selected chemical impurities and corrosion products within power plant systems. The levels of impurities are compared to corresponding limiting values to create an index. The mean was 1.02 in 2010 and 2011 (SD = .05 and .04 respectively).

Human performance error rate is a performance indicator maintained by INPO that measures the rate of negative events per month over the course of the year when an event occurs as a result of a human performance error. The mean was .01 (SD = .01) in both 2010 and 2011.

Forced loss rate is a performance indicator maintained by INPO that measures the ratio of all unplanned forced energy losses during a given period of time to the reference energy generation minus energy generation losses corresponding to planned outages and any unplanned outage extensions of planned outages, during the same period, expressed as a percentage. Essentially it is a measure of the amount of production lost due to unplanned events. In 2010 the mean value was 2.57 (SD = 4.79), and in 2011 the mean value was 2.24 (SD = 2.45).

Industrial safety accident (ISA) rate is a performance indicator maintained by INPO that measures the number of accidents for all utility personnel assigned to the station, which resulted in one or more days away from work, or one or more days of restricted work, or fatalities, per 200,000 man-hours worked. In both 2010 and 2011 the mean was .10 (SD = .08).

3. Results

3.1. Descriptive analysis of safety culture factors

We first conducted a descriptive analysis of the safety culture survey and associated factors to aid in the interpretation of the results. The mean value for safety culture overall and the means for the safety culture factors are higher than the mid-point of the response scale (4), suggesting that respondents generally rated the survey items favorably. Most of the means fall between a value of 5 and 6 on the 7-point scale, which correspond to the response options "somewhat agree" (5) and "agree" (6). Table 2 provides the means, standard deviations, and intercorrelations for the overall measure of safety culture and each of the safety culture factors.

3.2. Hypothesis testing

The concurrent validity analysis sought to assess whether there was an apparent relationship between safety culture and safety performance during the same time period, and corresponds to Hypothesis 1. The predictive validity analysis sought to assess the relationship between safety culture and future safety performance, corresponding to Hypothesis 2. We used site-level mean scores on the safety culture survey as the indicator of safety culture and aggregated site-level values for performance metrics as the indicator(s) of safety performance. Because the safety culture survey was administered in 2010, we used safety performance measures from calendar year 2010 for the concurrent analysis, and safety performance measures from 2011 for the predictive analysis. The sample size was 63; the total number of sites that participated in the safety culture survey. Pearson correlations are reported for the overall measure of safety culture and the 9 safety culture factors. Relationships between individual safety culture factors and the safety performance measures may suggest instances where some aspects of safety culture are more strongly related to safety performance than other aspects.

Table 3 presents the Pearson's correlations between safety culture and the safety performance measures in 2010 and 2011. Correlations that are statistically significant ($p < .05$ or $p < .01$) are highlighted.

3.2.1. Concurrent validity analysis (Hypothesis 1)

The overall measure of safety culture demonstrated statistically significant correlations with the 2010 safety performance measures of unplanned scrams (H1a), total ROP cross-cutting aspects (H1c), human performance cross-cutting area (H1d), problem identification and resolution cross-cutting area (H1e), chemistry performance index (H1h), and human performance error rate (H1i). The correlations were negative, and thus in the expected direction, demonstrating at least partial support for Hypothesis 1.

The following performance measures were not significantly correlated with the overall measure of safety culture: NRC allegations (H1b), total substantive cross-cutting issues (H1f), ROP action matrix oversight (H1g), forced loss rate (H1j), and industrial safety accident rate (H1k). However, three of the five performance measures that were not significantly correlated with the overall measure of safety culture demonstrated significant correlations with some of the safety culture factors. For example, questioning attitude was significantly correlated with NRC allegations and total substantive cross-cutting issues, and willingness to raise concerns was significantly correlated with ROP action matrix oversight. Only the performance measures of forced loss rate and industrial safety accident rate were uncorrelated with the overall measure of safety culture and the safety culture factors.

Sites with more than two unplanned scrams were generally below the overall mean score for safety culture of 5.61. Sites with

Table 2
Means, standard deviations, and intercorrelations among safety culture factors.

		Mean	Std. dev.	Safety culture	1	2	3	4	5	6	7	8	9
	Safety culture	5.61	0.79	–									
1	Management commitment	5.07	1.02	.98**	–								
2	Willingness to raise concerns	5.81	1.02	.87**	.81**	–							
3	Decision-making	5.98	0.83	.96**	.93**	.85**	–						
4	Supervisor responsibility	5.63	1.01	.86**	.81**	.71**	.81**	–					
5	Questioning attitude	5.93	0.7	.87**	.83**	.76**	.80**	.69**	–				
6	Safety communication	5.78	0.78	.95**	.89**	.83**	.89**	.82**	.85**	–			
7	Personal responsibility	6.54	0.52	.37**	.26*	.45**	.37**	.28*	.37**	.46**	–		
8	Prioritizing safety	5.98	0.86	.90**	.86**	.79**	.89**	.75**	.77**	.85**	.28*	–	
9	Training quality	5.75	0.95	.80**	.76**	.64**	.75**	.66**	.70**	.80**	.48**	.72**	–

* $p < .05$.
** $p < .01$.

Table 3
Correlations between safety culture factors and safety performance measures in 2010 and 2011.

	Unplanned scrams		NRC allegations		Total ROP aspects		Human perf. area		Problem ID & Res. area		Total SCCIs		ROP action matrix oversight		Chemistry perf. index		Human perf. error rate		Forced loss rate		ISA rate	
	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011
SAFETY CULTURE	-.37**	-.27*	-.21	-.36**	-.39**	-.20	-.28*	-.12	-.37**	-.27*	-.23	-.26*	-.23	-.30*	-.27*	-.45**	-.36**	-.22	-.11	-.31*	.01	-.25
1. Management commitment	-.34**	-.25	-.24	-.38**	-.44**	-.21	-.30*	-.11	-.40**	-.30*	-.28*	-.27*	-.20	-.29*	-.27*	-.45**	-.38**	-.19	-.11	-.28*	-.01	-.23
2. Willingness to raise concerns	-.29*	-.31*	-.16	-.38**	-.21	-.10	-.16	-.03	-.24	-.18	-.10	-.18	-.31*	-.24	-.22	-.36**	-.37**	-.16	-.08	-.23	-.04	-.21
3. Decision-making	-.33**	-.30*	-.17	-.33**	-.37**	-.21	-.25*	-.13	-.35**	-.29*	-.21	-.32*	-.27*	-.32*	-.28*	-.40**	-.36**	-.20	-.08	-.33**	.05	-.28*
4. Supervisor responsibility	-.26*	-.26*	-.15	-.20	-.28*	-.17	-.27*	-.12	-.24	-.20	-.14	-.18	-.27*	-.30*	-.14	-.38**	-.40**	-.27*	-.06	-.26*	.01	-.23
5. Questioning attitude	-.29*	-.30*	-.41**	-.48**	-.47**	-.19	-.40**	-.15	-.37**	-.19	-.38**	-.23	-.17	-.26*	-.26*	-.45**	-.28*	-.19	-.19	-.29*	-.11	-.33**
6. Safety communication	-.35**	-.24	-.12	-.28*	-.31*	-.18	-.20	-.14	-.32*	-.20	-.14	-.21	-.19	-.29*	-.28*	-.44**	-.39**	-.23	-.14	-.33**	.00	-.23
7. Personal responsibility	-.24	-.07	.17	-.08	.04	-.06	.05	-.05	.00	-.06	.13	-.10	-.28*	-.05	-.27*	-.26*	-.21	-.20	-.06	-.25*	-.11	-.14
8. Prioritizing safety	-.23	-.21	-.09	-.21	-.28*	-.21	-.18	-.14	-.30*	-.25*	-.11	-.20	-.11	-.27*	-.21*	-.31*	-.25*	-.22	-.03	-.22	.10	-.16
9. Training quality	-.46**	-.09	.04	-.07	-.22	-.11	-.14	-.06	-.25*	-.15	-.10	-.20	-.23	-.23	-.30*	-.37**	-.19	-.25*	-.13	-.44**	.07	-.10

* $p < .05$.
** $p < .01$.

higher scores on the safety culture survey were more likely to have fewer unplanned scrams in 2010. Most notably, the correlation between training quality and unplanned scrams was −.46. Sites with higher quality training, as judged by site personnel, had fewer unplanned scrams in 2010.

The overall measure of safety culture evidenced a medium-sized correlation with the total number of ROP cross-cutting aspects (−.41), mainly due to the high correlations with the factors management commitment to safety (−.44) and questioning attitude (−.45). Sites where management was perceived as demonstrating a strong commitment to safety and fostering a questioning attitude among their workforce were less likely to have inspection findings that inspectors attributed to a cross-cutting aspect. The same factors, management commitment to safety and questioning attitude, were significantly correlated with substantive cross-cutting issues. Sites where employees reported better management commitment to safety and more of a questioning attitude were less likely to have a substantive cross-cutting issue at the end of 2010.

Human performance error rate was correlated with the overall measure of safety culture, and seven out of nine of the safety culture factors. Chemistry performance index was correlated with the majority of the safety culture factors, the exceptions being willingness to raise concerns and supervisor responsibility for safety.

The safety culture factors of decision making and questioning attitude had the most frequent statistically significant correlations with the variety of safety performance measures examined in this study. On the other hand, personal responsibility for safety was uncorrelated with the majority of the safety performance measures.

3.2.2. Predictive validity analysis (Hypothesis 2)

The overall measure of safety culture demonstrated statistically significant and negative correlations with the following 2011 performance measures: unplanned scrams (H2a), NRC allegations (H2b), problem identification and resolution cross-cutting area (H2e), substantive cross-cutting issues (H2f), NRC action matrix oversight (H2g), chemistry performance index (H2h), and forced loss rate (H2j). These correlations suggest partial support for Hypothesis 2.

The ROP cross-cutting aspects (H2c), human performance cross-cutting area (H2d) and human performance error rate (H2i), were not significantly correlated with the overall measure of safety culture or any of the safety culture factors. Although there were concurrent relationships between these safety performance measures and safety culture, predictive relationships were not evident using the 2011 data. Industrial safety accident rate (H2k) was the only safety performance measure that did not have a significant correlation with safety culture in 2010 or 2011.

Unplanned scrams, problem identification and resolution cross-cutting area, and chemistry performance index had significant correlations with safety culture in both 2010 and 2011. NRC allegations, substantive cross-cutting issues, NRC action matrix oversight, and forced loss rate had significant correlations with safety culture in 2011, but not in 2010. Although the overall measure of safety culture was not correlated with NRC allegations in 2010, the factor of questioning attitude had a fairly strong correlation with NRC allegations in both 2010 and 2011 (−.41 and −.48, respectively). Forced loss rate was not correlated with overall measure of safety culture nor any of the factors in 2010, but was significantly correlated with all but two factors in 2011.

3.3. Follow-up analyses

The bivariate correlations between safety culture and safety performance demonstrated partial support for both Hypothesis 1 and 2. However, there were many instances where the overall measure of safety culture was not significantly correlated with a safety performance measure, but one or more safety culture factors were correlated with the safety performance measure. Because of the variability in the bivariate correlations, we sought to test our original hypotheses in an alternative way by focusing on the safety culture factors. Specifically, we ran a series of regression analyses with the safety culture factors entered together in the regression equation as independent variables, and each safety performance measure as the dependent variable. The purpose of the follow-up analyses was to determine if the safety culture factors collectively account for a significant portion of the variance in safety performance, and if so, which factors seem to drive that relationship. The results of the follow-up analyses are presented in Tables 4 and 5.

The results of the regression analyses suggest statistically significant relationships between safety culture, as measured by the factors, and the 2010 safety performance measures of NRC allegations, ROP cross-cutting aspects, human performance cross-cutting area, substantive cross-cutting issues, and human performance error rate. In addition, the safety culture factors had marginally significant ($p > .10$) relationships with the 2010 measures of unplanned scrams, problem identification and resolution cross-cutting area, NRC action matrix oversight, and chemistry performance index. The safety culture factors were not significantly related to forced loss rate in 2010. With the exception of forced loss rate, the safety culture factors accounted for 23–52% of the variance in concurrent safety performance measures. In particular, the factor questioning attitude seemed to drive the relationship between safety culture and NRC allegations, total ROP aspects, human performance cross-cutting area, and substantive cross-cutting issues.

The safety culture factors also demonstrated significant predictive relationships with NRC allegations and chemistry performance index (i.e., using the 2011 data). The safety culture factors accounted for 52% of the variance in NRC allegations in 2010, and 48% of the variance in 2011. The factors accounted for 25% of the variance in chemistry performance index in 2010, and 27% of the variance in 2011. There were also marginally significant relationships between the safety culture factors and forced loss rate (24% of variance) and industrial safety accident rate (23% of variance). The other predictive relationships were non-significant, with the safety culture factors only accounting for 0–20% of the variance in safety performance. Because the regression analyses include

Table 4
Regression analysis testing relationship between safety culture factors and NRC safety performance measures.

Safety culture factors	Safety performance measures													
	Unplanned scrams		NRC allegations		Total ROP aspects		Human performance area		Problem id. & res. area		Total SCCIs		ROP action matrix oversight	
	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011
1. Management commitment	−.15	.44	−.22	−.57	−.47	−.21	−.25	.08	−.58	−.60	−.44	−.06	.39	.28
2. Willingness to raise concerns	−.14	−.27	−.01	−.24	.33	.43	.28	.46	.28	.27	.31	.30	−.39	−.10
3. Decision-making	−.34	−.59	−.25	−.24	−.30	−.28	−.25	−.19	−.24	−.33	−.44	−1.00	−.79†	−.61
4. Supervisor responsibility	.10	−.18	−.13	.17	−.07	−.01	−.28	−.04	.22	.05	.03	.14	−.22	−.11
5. Questioning attitude	.05	−.40	−1.11**	−.75**	−.67**	−.11	−.82**	−.21	−.24	.07	−.91**	−.11	−.15	−.26
6. Safety communication	.12	.13	.56†	.31	.35	.01	.58	−.17	−.05	.28	.57†	.21	.52	.25
7. Personal responsibility	−.03	.08	.19	−.04	.07	−.13	.05	−.09	.07	−.15	.15	−.05	.01	.05
8. Prioritizing safety	.40	.27	.48*	.53*	.28	−.19	.28	−.21	.19	−.09	.50†	.28	.61	.11
9. Training quality	−.45*	.26	.38	.49**	.14	.25	.14	.23	.10	.21	.10	.05	−.30	.08
R	.49	.45	.72	.69	.58	.31	.56	.29	.48	.38	.64	.42	.48	.38
R^2	.24	.20	.52	.48	.34	.10	.31	.08	.23	.14	.41	.18	.23	.15
Adjusted R^2	.12	.06	.43	.39	.25	−.06	.19	−.07	.10	−.00	.31	.04	.10	.00
Std. error	1.20	.95	6.55	2.79	7.33	6.5	4.95	4.47	4.16	2.79	.67	.65	.62	.79
F	1.9†	1.46	6.06**	5.36**	3.00**	.64	2.66**	.54	1.78†	.97	4.02**	1.27	1.78†	1.01

All safety culture factors entered in the same step in the regression equation. Standardized coefficients (β) reported.
† $p < .10$.
* $p < .05$.
** $p < .01$.

Table 5
Regression analysis testing relationship between safety culture factors and INPO safety performance measures.

Safety culture factors	Safety performance measures							
	Chemistry performance index		Human performance error rate		Forced loss rate		Industrial safety accident rate	
	2010	2011	2010	2011	2010	2011	2010	2011
1. Management commitment	−.82†	−.51	−.75†	.02	−.21	.22	−.68	.21
2. Willingness to raise concerns	.25	.09	−.15	.10	.17	.13	−.04	.18
3. Decision-making	.32	.10	.24	.17	−.08	−.65	.15	−.83†
4. Supervisor responsibility	.40†	−.05	−.30	−.37	.17	.04	.08	−.10
5. Questioning attitude	−.01	−.26	.14	−.06	−.29	−.10	−.21	−.54*
6. Safety communication	−.34	−.15	.09	.28	−.15	.06	.27	.14
7. Personal responsibility	−.34*	−.10	−.22	−.18	−.09	−.02	−.35	−.09
8. Prioritizing safety	.08	.34	.24	−.21	.31	.40	.23	.44
9. Training quality	.05	.01	.42†	−.17	.01	−.47*	.39	.32
R	.50	.52	.54	.35	.29	.49	.39	.48
R^2	.25	.27	.30	.13	.09	.24	.15	.23
Adjusted R^2	.12	.15	.18	−.02	−.07	.11	.00	.10
Std. error	.05	.04	.01	.01	5.02	2.34	.08	.09
F	1.98†	2.17*	2.46*	.84	.552	1.86†	1.03	1.76†

All safety culture factors entered in the same step in the regression equation. Standardized coefficients (β) reported.
† $p < .10$.
* $p < .05$.

nine independent variables (for the nine safety culture factors) there are fewer degrees of freedom and less power to detect effects as compared to the bivariate correlations.

4. Discussion

This study sought to investigate the factors that comprise the concept of safety culture in the nuclear power industry and evaluate the relationships between safety culture and measures of safety performance. The results confirm that the safety culture survey is multidimensional and demonstrates statistically significant concurrent and predictive relationships with some measures of safety performance. However, there was variability in the results, depending on how safety culture was measured, the specific safety performance measures used, and whether safety culture was compared to concurrent or future safety performance.

The significant correlations between overall safety culture and measures of safety performance ranged from −.26 to −.45, suggesting a medium effect and that safety culture accounts for 7–21% of the variance in most of the measures of safety performance examined in this study. The correlations in this study are generally consistent with the magnitude of correlations between safety culture and safety performance reported in safety climate meta-analysis studies. For example, Christian et al. (2009) found that the mean correlation between organizational safety climate and safety outcomes was −.39 across six studies when corrected for unreliability. Beus et al. (2010) found that the mean corrected correlation between organizational safety climate and injury rates across 11 studies was −.24. Clarke (2006) found a mean corrected correlation of .22 between safety climate and accident involvement across 25 retrospective studies, and a mean corrected correlation of .35 across six prospective studies. Therefore, relationships between safety culture and safety performance outcomes in the nuclear industry seem consistent with results from other industries, and the correlations between safety culture and process-based measures of safety performance are similar to, if not slightly stronger than, correlations in other studies using personal safety outcomes like injuries and accident involvement.

When all of the safety culture factors were examined in the context of a regression equation for our follow-up analyses, some of the results were even more promising, suggesting that the factors collectively accounted for 23–52% of the variance in concurrent safety performance. In addition, the safety culture factors accounted for a significant amount of predictive variance in allegations (48% variance explained), and chemistry performance (27% variance explained).

Overall, we saw that the relationships between safety culture and safety performance were more consistent when safety performance was measured concurrently, as opposed to being measured one year after the survey administration. Concurrent correlations also tended to be stronger than the predictive correlations. This is consistent with the trends identified by Mearns et al. (2003) in their study of safety climate and safety performance in offshore environments. Although lacking in power to detect significant effects, they found that the trends in correlations between safety culture and safety performance were more consistent in year 1 as compared to year 2.

The most consistent relationship across both the correlation and regression analyses seemed to be between the safety culture factor questioning attitude, and the outcome variable NRC allegations. Organizations where employees perceived less of a questioning attitude were more likely to receive higher numbers of allegations in the same year and the following year. When employees are more likely to report safety concerns through the allegation program it may be an indicator that they have less confidence in their internal processes for identifying and resolving safety issues. Questioning attitude was also a significant predictor of concurrent counts of inspection findings associated with ROP cross-cutting aspects, the cross-cutting area of human performance, and total number of SCCIs. Fostering a questioning attitude may be a particularly important component of the overall safety culture of an organization.

Unplanned scrams and the problem identification and resolution cross-cutting area were the only performance measures to demonstrate significant correlations with the overall measure of safety culture concurrently and predictively. Organizations with lower overall scores on the safety culture survey were more likely to have higher counts of unplanned scrams, and have inspection findings related to inadequacies in problem identification and resolution. The concurrent correlation between safety culture and unplanned scrams seemed to be driven by training quality, whereas the factor that had the strongest correlation with problem identification and resolution was management commitment to safety. However, the follow-up regression analyses did not meet the significance threshold of $p < .05$ and none of the standardized coefficients for the factors in the regression analyses were statistically significant.

Some of the strongest predictive correlations between overall scores on the safety culture survey and NRC safety performance measures relate to the broad measures of SCCIs, and ROP action matrix oversight. The overall measure of safety culture was not significantly related to these variables using the 2010 performance data, but demonstrated significant negative correlations with the 2011 data. Sites with lower scores on the safety culture survey were more likely to receive substantive cross-cutting issues, and be in an elevated oversight condition within the ROP action matrix in the following year. Lower scores on the safety culture survey may indicate problems at a site that are starting to add up, and it is only at future points in time that those problems, if not adequately addressed, require NRC response through the issuance of SCCIs and movement in the ROP action matrix.

The safety culture factor personal responsibility for safety seemed to be the poorest performing safety culture factor in terms of its relationship to the performance measures included in this study. Overall, personal responsibility seemed to be different from the other safety culture factors included in the survey. Personal responsibility was comprised of items that included statements that referenced the individual rather than the organization, such as, "it is my responsibility to raise nuclear safety concerns," and demonstrated the lowest intercorrelations with the other safety culture factors. In addition, personal responsibility had a mean very near the high end of the response scale (6.54), and a low standard deviation (.52). Most respondents indicated that they strongly agreed with the items that comprised this factor, likely because people tend to respond more favorably toward items that ask about their personal safety attitudes as compared to coworkers' attitudes or the organization's overall safety values. Lee and Harrison (2000) found a similar result in their survey of safety culture in nuclear power stations. The factor of personal responsibility, as measured in the survey used in this study, may not be as useful or valid as an indicator of safety culture as compared to the other factors in the survey.

It is particularly interesting that the only measure of safety performance that was not significantly correlated with safety culture was industrial safety accident rate. Accident rates are one of the most commonly used independent measures of safety performance. However, many have noted concerns with using accident rates to represent performance, primarily that accidents may be few and far between and do not capture the frequency of micro-events, such as near misses or minor incidents that do not meet accident reporting requirements (Zohar, 2000). In the current study, industrial safety accident rate seems to be a measure of performance that is less relevant to an organization's safety culture as compared to other measures of performance.

4.1. Limitations

Overall, the safety culture survey demonstrated sound psychometric properties (i.e., reliability and validity) and a factor structure supported by other safety culture research. However, these results should be interpreted within the context of the study, which used data from a survey administered to employees at a single point in time to approximate an organization's safety culture. The single administration of the survey, combined with the correlational analyses, does not permit conclusions to be drawn regarding a causal relationship between safety culture and safety performance.

In particular, the findings presented here are exploratory, mainly because the correlational analyses cannot be used to verify causality and the data used represent snapshots of safety culture and safety performance. These individual variables do not, by themselves, adequately capture the full scope and dynamic nature of an organization's safety culture or its safety performance. Additional analyses would be necessary to establish confidence that the results represent a causal phenomenon that is stable over time. For instance, structural equation modeling using additional longitudinal data (i.e., multiple administrations of the safety culture survey and annual safety performance metrics over successive years) would allow for more complex testing of combinations of safety culture factors and performance variables to establish whether safety culture consistently predicts safety performance.

Additional, ongoing research is necessary to determine whether the relationships observed are consistent over time, whether the same factors consistently emerge in subsequent survey administrations within the nuclear power industry, and whether different safety culture factors are uniquely related to different aspects of performance. And, finally, the generalizability of this study's results to other industries is unknown. Comparable data, both in terms of a safety culture survey and robust performance metrics, would be necessary to replicate this study in other domains.

5. Conclusions

Safety culture has received considerable attention in high reliability industries, including nuclear power operations, because of its potential contribution to ensuring safe performance. Post-event analyses and accident investigations in many different industries have frequently cited weaknesses in an organization's safety culture as significant contributors to event occurrence. Some recent examples of accidents where safety culture was identified as a contributing cause include BP's Texas City refinery explosion in 2005 (Chemical Safety Board, 2007), the Washington Metropolitan Area Transit Authority rail collision in 2009 (National Transportation Safety Board, 2010), the Deepwater Horizon oil spill in 2010 (United States Coast Guard, 2011), the Upper Big Branch mine explosion in 2010 (Mine Safety and Health Administration, 2011), and the Fukushima nuclear accident in 2011 (National Diet of Japan, 2012).

This study represents a first step in the empirical exploration of relationships between safety culture and other measures of nuclear power plant performance. The results provide evidence that safety culture is correlated with concurrent measures of safety performance and may be related to future performance for some measures. But even within a single industry in the same country, the relationship between safety culture and safety performance is highly dependent on how and when both safety culture and safety performance are measured.

Moreover, it should be recognized that a survey is only one potential indicator of safety culture and does not constitute a full assessment of an organization's safety culture. There are no established thresholds for determining what constitutes a "healthy" or "unhealthy" safety culture. Nor do the results presented in this paper attempt to draw any conclusions about appropriate thresholds. The mean score on the safety culture survey was 5.61, which indicates that respondents generally agreed with the positive statements concerning their organization's safety cultures. There is considerable debate in the research literature about whether a threshold can be defined or is appropriate for evaluating safety culture. It may be that organizations can have significantly different cultures and perform equally well in terms of safety (Reiman and Oedewald, 2007). Edgar Schein echoes this idea, noting that, "In most organizational change efforts, it is much easier to draw on the strengths of the culture than to overcome the constraints by changing the culture" (2010, p. 327).

Often, because of data limitations, organizational culture research is limited to comparing subcultures within a single company. Because of the breadth, consistency, and warehousing of data in the U.S. nuclear power industry there is a unique opportunity for comparative organizational research that is not available in most other industries. This research has shown promise for not only demonstrating the relationship between safety culture and current measures of plant performance but hints at ways that safety culture may be a predictor of future plant performance.

Acknowledgements

This paper presents information that does not currently represent an agreed-upon staff position. The U.S. Nuclear Regulatory Commission has neither approved nor disapproved its technical content.

The safety culture survey data presented in this paper are the property of the Institute of Nuclear Power Operations. Data were shared under a Memorandum of Agreement between the Nuclear Regulatory Commission and the Institute of Nuclear Power Operations, dated December 10, 2007.

References

Ajzen, I., Fishbein, M., 1977. Attitude-behavior relations: a theoretical analysis and review of empirical research. Psychol. Bull. 84, 888–918.

Beus, J., Payne, S., Berman, M., Arthur, W., 2010. Safety climate and injuries: an examination of theoretical and empirical relationships. J. Appl. Psychol. 95 (4), 713–727.

Chemical Safety and Hazard Investigation Board, 2007. Investigation Report, Refinery explosion and fire, BP, Texas City, TX, March 23, 2005. [2005-04-I-TX]. Washington, DC.

Christian, M.S., Bradley, J.C., Wallace, J.C., Burke, M.J., 2009. Workplace safety: a meta-analysis of the roles of person and situation factors. J. Appl. Psychol. 94 (5), 1103–1127.

Clarke, S., 2006. The relationship between safety climate and safety performance: a meta-analytic review. J. Occup. Health Psychol. 11 (4), 315–327.

Cohen, J., 1988. Statistical power analysis for the behavioral sciences, second ed. Lawrence Earlbaum Associates, Hillsdale, NJ.

Ford, J., MacCallum, R., Tait, M., 1986. The application of exploratory factor analysis in applied psychology: a critical review and analysis. Pers. Psychol. 39, 291–314.

Guldenmund, F., 2000. The nature of safety culture: a review of theory and research. Safety Sci. 34 (1–3), 215–257.

Hinkin, T., 1998. A brief tutorial on the development of measures for use in survey questionnaires. Organ. Res. Methods 1 (1), 104–121.

Institute for Nuclear Power Operations, 2004. Principles for a Strong Nuclear Safety Culture. Institute of Nuclear Power Operations, Atlanta.

International Atomic Energy Agency, 2006. Application of the management system for facilities and activities: safety guide. IAEA Safety Standards Series No. GS-G-3.1. International Atomic Energy Agency, Vienna.

International Nuclear Safety Advisory Group, 1986. Summary Report on the Post-Accident Review Meeting on the Chernobyl Accident, Safety Series No. 75-INSAG-1, International Atomic Energy Agency, Vienna.

James, L.R., 1982. Aggregation bias in estimates of perceptual agreement. J. Appl. Psychol. 67, 219–229.

Kim, J., Mueller, C., 1978. Introduction to Factor Analysis: What it is and how to do it. Sage, Beverly Hills, CA.

Krosnick, J., 1999. Survey research. Annu. Rev. Psychol. 50, 537–567.

Lee, T., Harrison, K., 2000. Assessing safety culture in nuclear power stations. Safety Sci. 34, 61–97.

McGraw, K., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychol. Methods 1 (1), 30–46.

Mearns, K., Whitaker, S., Flin, R., 2003. Safety climate, safety management practice, and safety performance in offshore environments. Safety Sci. 41, 641–680.

Mine Safety and Health Administration, 2011. Report of Investigation: Fatal Underground Mine Explosion, April 5, 2010. Coal Mine Safety and Health.

National Diet of Japan, 2012. Report of the Fukushima Independent Investigation Commission, March 11, 2011. Tokyo, Japan.

National Transportation Safety Board, 2010. Collision of Two Washington Metropolitan Area Transit Authority Metrorail Trains Near Fort Totten Station, Washington, D.C., June 22, 2009. Railroad Accident Report NTSB/RAR-10/02. Washington, DC.

Neal, A., Griffin, M., 2006. A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. J. Appl. Psychol. 91 (4), 946–953.

Nuclear Regulatory Commission, 2006. Technical Basis for Inspection Program. NRC Inspection Manual, Chapter 0308, Appendix 2. United States Nuclear Regulatory Commission, Washington, DC.

Nuclear Regulatory Commission, 2008. Industry Trends Program. NRC Inspection Manual, Chapter 0313. United States Nuclear Regulatory Commission, Washington, DC.

Nuclear Regulatory Commission, 2011a. Components within the Cross-Cutting Areas. NRC Inspection Manual, Chapter 0310. United States Nuclear Regulatory Commission, Washington, DC.

Nuclear Regulatory Commission, 2011b. Operating Reactor Assessment Program. NRC Inspection Manual, Chapter 0305. United States Nuclear Regulatory Commission, Washington, DC.

Reason, J., 1997. Managing the Risks of Organizational Accidents. Ashgate.

Reiman, T., Oedewald, P., 2007. Assessment of complex sociotechnical systems – theoretical issues concerning the use of organizational culture and organizational core task concepts. Safety Sci. 45, 745–768.

Rogovin, M., 1980. Three Mile Island – A Report to the Commissioners and the Public, vol. 1.

Schein, E.H., 1992. Organizational Culture and Leadership, Second ed. Jossey-Bass, San Francisco.

Schein, E.H., 2010. Organizational Culture and Leadership, fourth ed. Jossey-Bass, San Francisco.

Sorensen, J.N., 2002. Safety culture: a survey of the state-of-the-art. Reliab. Eng. Syst. Safety 76, 189–204.

United States Coast Guard, 2011. Report of Investigation into the Circumstances Surrounding the Explosion, Fire, Sinking and Loss of Eleven Crew Members Aboard the Mobile Offshore Drilling Unit Deepwater Horizon in the Gulf of Mexico, April 20–22, 2010. Vol. 1, Washington, DC.

Zohar, D., 2000. A group-level model of safety climate: testing the effect of group climate on microaccidents in manufacturing jobs. J. Appl. Psychol. 85 (4), 587–596.