NPS-OR-11-004

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

AN ASSESSMENT OF THE RELATIONSHIP BETWEEN SAFETY CLIMATE AND MISHAP RISK IN U.S. NAVAL AVIATION

by

Paul O'Connor, Samuel E. Buttrey, Angela O'Dea, and
Quinn Kennedy

October 2011

Approved for public release; distribution is unlimited

Prepared for: Defense OSD Readiness Programming and Assessment
Defense Safety Oversight Council
4000 Defense Pentagon
Washington, D.C. 20301-4000

THIS PAGE INTENTIONALLY LEFT BLANK

NAVAL POSTGRADUATE SCHOOL

Monterey, California 93943-5000

Daniel T. Oliver
President

Leonard A. Ferrari
Executive Vice President and
Provost

This report was prepared for the Defense OSD Readiness Programming and Assessment, Defense Safety Oversight Council, 400 Defense Pentagon, Washington, D.C. 20301-4000 and funded by the Defense Safety Oversight Council, 110 Army Pentagon, Room 3E464, Washington, D.C. 20310-0110.

This report was prepared by:

Paul O'Connor
Senior Research Fellow
National University of Ireland

Samuel E. Buttrey
Associate Professor
Department of Operations Research

Angela O'Dea
Investigator

Quinn Kennedy
Lecturer
Department of Operations Research

Reviewed by:

Ronald D. Fricker
Associate Chairman for Research
Department of Operations Research

Robert F. Dell
Chairman
Department of Operations Research

Released by:

Karl A. van Bibber, Ph.D.
Vice President and Dean of Research

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED**
01-10-2011	Technical Report	Oct 09-Sept 10

4. TITLE AND SUBTITLE
An Assessment of the Relationship Between Safety Climate and Mishap Risk in U.S. Naval Aviation

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Paul O'Connor
Samuel E. Buttrey
Angela O'Dea
Quinn Kennedy

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-OR-11-004

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Safety Oversight Council
110 Army Pentagon Room 3E464
Washington, D.C. 20310-0110

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

14. ABSTRACT
This study used a prospective design to assess whether 12 items from the Command Safety Assessment Survey (CSAS) can be used to differentiate between U.S. Naval aviation squadrons who have had a mishap within a recent period of time, and those that have not. Logistic regression modeling was carried out using the survey responses of U.S. Naval aircrew (n = 23,442) and mishap data. The models that were used to attempt to predict severe and moderately severe mishaps together, performed better than the models that used subsets of the mishaps data. It was found that three of the CSAS items had some limited value in predicting mishap risk. Personnel in squadrons with a low probability of mishap more strongly agree with the need to monitor personnel and integrate safety and operations, than aircrew in squadrons with a higher probability of mishap. However, the aircrew in squadrons with a higher probability of mishap also more strongly agrees that persistent rule violators will jeopardize their career, compared to personnel in squadrons with a low probability of mishaps. This finding suggests that blame and punishment are not constructive in efforts to promote safety at work. This study would seem to support the premise that safety climate and safety performance are weakly related. It is recommended that researchers would be better advised to attempt to establish the discriminate validity of their questionnaire through self-reported safety attitudes and behaviors, rather than mishap data.

15. SUBJECT TERMS Aviation Safety, Survey

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
			UU	44	Samuel E. Buttrey
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified			(831) 656-3035

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This study used a prospective design to assess whether 12 items from the Command Safety Assessment Survey (CSAS) can be used to differentiate between U.S. Naval aviation squadrons who have had a mishap within a recent period of time, and those that have not. Logistic regression modeling was carried out using the survey responses of U.S. Naval aircrew (n = 23,442) and mishap data. The models that were used to attempt to predict severe and moderately severe mishaps together, performed better than the models that used subsets of the mishaps data. It was found that three of the CSAS items had some limited value in predicting mishap risk. Personnel in squadrons with a low probability of mishap more strongly agree with the need to monitor personnel and integrate safety and operations, than aircrew in squadrons with a higher probability of mishap. However, the aircrew in squadrons with a higher probability of mishap also more strongly agrees that persistent rule violators will jeopardize their career, compared to personnel in squadrons with a low probability of mishaps. This finding suggests that blame and punishment are not constructive in efforts to promote safety at work. This study would seem to support the premise that safety climate and safety performance are weakly related. It is recommended that researchers would be better advised to attempt to establish the discriminate validity of their questionnaire through self-reported safety attitudes and behaviors, rather than mishap data.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

This work was funded by the Defense Safety Oversight Council (DSOC). All opinions stated in this paper are those of the authors and do not necessarily represent the opinion or position of the U.S. Navy, the Naval Postgraduate School, or the National University of Ireland, Galway.

THIS PAGE INTENTIONALLY LEFT BLANK

1. INTRODUCTION

Safety climate describes employees' perceptions, attitudes, and beliefs about risk and safety (Mearns & Flin, 1999). The most commonly used method for measuring safety climate is through the use of questionnaires. Questionnaires have been used to assess safety climate for over three decades in many high-risk organizations (e.g., construction, offshore oil production, and aviation maintenance; Flin, Mearns, O'Connor, & Bryden, 2000).

Over the past two decades, researchers have demonstrated relationships between a variety of safety climate factors and mishap rates across a range of high-risk industries. Such studies have shown that the degree of safety program development and workers' safety initiative were related to lower work mishap and injury rates (e.g., Simard & Marchand, 1994; Donald & Canter, 1994; Mearns, Rundmo, Flin, Gordon, & Fleming, 2004; Zohar, 2000).

Other studies have compared high- and low-accident-rate plants (or evaluated plants with outstanding safety records) as their criteria upon which to base judgments of effectiveness (e.g., Braithwaite, 1985; Cohen & Cleveland, 1983; Cohen, Smith, & Cohen, 1975; Diaz & Cabrera, 1997; Hale & Hovden, 1998). However, it is difficult to identify those variables that are crucial to their outstanding performance, as opposed to those that are simply associated with it. The differences highlighted between good and bad companies may only be a fraction of the total, and these differences may change over time (Mearns, Flin, & O'Connor, 2001). Furthermore, the causal relationship between variables and outcomes are not proven; it is often difficult to say which is the independent variable and which the dependent variable (Hale & Hovden, 1998). Nonetheless, the early studies that compared high- and low-accident-rate organizations did identify issues that have been supported by the other more objective approaches to validation—such as incident and incident rates or self-reported safety behavior. The purpose of the study reported in this paper is to assess whether the responses to items from the Command Safety Assessment Survey (CSAS) are associated with a higher risk of a mishap in U.S. Naval aviation.

For a safety climate questionnaire to be useful, it requires both construct and discriminate validity. Construct validity is concerned with the extent to which the questionnaire measures the underlying theoretical construct it intends to measure. The identification of a reliable factor structure, that is consistent with theory, helps the researcher substantiate the construct validity of the questionnaire. Discriminate validity refers to the ability of the questionnaire to differentiate between organizations or personnel with different levels of safety performance. The discriminate validity can be assessed by correlating the data from the questionnaire with a criterion variable such as accidents, or other safety-related behavior (Guldenmund, 2007). In a review of 23 studies of aviation-specific safety climate tools, it was concluded that the reviewed questionnaires had some construct validity (O'Connor, O'Dea, Kennedy, & Buttrey, 2011). However, O'Connor et al. (2011) found insufficient evidence to support the discriminate validity of the reviewed questionnaires. If a questionnaire is unable to differentiate between organizations or personnel with different levels of safety performance, then it is of limited usefulness.

1.1 THE COMMAND SAFETY ASSESSMENT SURVEY (CSAS)

The CSAS was developed by researchers at the Naval Postgraduate School in Monterey, California to assess the safety climate of Naval aviation squadrons (Desai, Roberts, & Ciavarelli, 2006). The theoretical background underpinning the CSAS is based upon a conceptual framework of Models of Organizational Safety Effectiveness (MOSE). This framework identified five major areas that are critical to high reliability organizations (HROs) in managing risk and reducing mishaps (Libuser, 1994).

The five MOSE areas are: process auditing (a system of ongoing checks to monitor hazardous conditions), reward system (expected social compensation or disciplinary action to reinforce or correct behavior), quality assurance (policies and procedures that promote high quality performance), risk management (how the organization perceives risk and takes corrective action), and command and control (policies, procedures, and communication processes used to mitigate risk). The CSAS was used continuously by the U.S. Navy from 2000 until 2009 (the CSAS is still used in

U.S. Naval aviation, but some changes were made to the content and structure of the questionnaire in early 2010).

1.2 CONSTRUCT VALIDITY OF THE CSAS

The first attempt to establish the construct validity of the CSAS was reported by Buttrey, O'Dea, O'Connor, and Kennedy (2010). They used 110,014 responses to the CSAS collected over eight years. Utilizing a combination of exploratory and confirmatory factor analysis, Buttrey et al. (2010) were unable to identify a stable factor structure for the 61-item CSAS. This finding was attributed to the presence of substantial nonrandom response bias associated with the data (the reverse-worded items had a unique pattern of responses; there was an increasing tendency over time to provide only a modal response; the responses to the same item towards the beginning and end of the questionnaire did not correlate as highly as might be expected; and the faster the questionnaire was completed, the higher the frequency of modal responses). It was suggested that the nonrandom responses bias was due to a number of factors (questionnaire design, lack of a belief in the importance of the response, participant fatigue, and questionnaire administration; Buttrey et al., 2010), which had a negative effect on participant motivation.

Fortunately, since 2006, data was collected on the time taken by respondents to complete the survey. This "time to complete" data was then used as a metric to identify the respondents expending sufficient cognitive effort to provide a thoughtful response to each item. Using time to complete, a total of 23,442 responses were retained for analysis. Buttrey et al. (2010) also elected to discard the CSAS items that had low variability. This left 12 items from the original 61-item questionnaire. Using the truncated dataset, exploratory and confirmatory factor analyses resulted in a stable, 12-item, two-factor model (see Buttrey et al., 2010, for details of the analysis).

One of the factors was labeled as "personnel leadership" and consisted of the following seven items:

4. *My command closely monitors proficiency and currency standards to ensure aircrew are qualified to fly*;

9. *My command makes effective use of the flight surgeon to help identify and manage high risk personnel*;

13. *In my command, we believe safety is an integral part of all flight operations*;

16. *Leaders in my command encourage everyone to be safety conscious and to follow the rules*;

17. *In this command, an aviator who persistently violates flight standards and rules will seriously jeopardize his/her career*;

19. *My command has a reputation for high-quality performance*; and

48. *My command does not hesitate to temporarily restrict from flying individuals who are under high personal stress*.

The second factor was labeled as "availability of resources" and consisted of the following five items:

31. *I am provided adequate resources (time, staffing, budget, and equipment) to accomplish my job*;

32. *My command provides the right number of flight hours per month for me to fly safely*;

50. *Morale and motivation in my command are high*;

52. *Crew rest standards are enforced in my command*; and

55. *Within my command, good communications flow exists up and down the chain of command*.

The two factors identified are consistent with the broader safety climate literature. For example, management, or factors under the control of management, is reported to make up approximately 75% of safety climate factors (Flin et al., 2000). Additionally, O'Connor et al. (2011) found "resources" to be a key factor in aviation safety climate literature.

1.3 DISCRIMINATE VALIDITY OF THE CSAS

There has been one previous attempt to establish the discriminate validity of the CSAS reported in the peer reviewed literature. Desai et al. (2006) used a retrospective approach (i.e., the measurement of safety climate occurred after the mishap had taken place) to assess whether there was a link between 6,361 responses to the CSAS from 147

U.S. Navy squadrons between July 2000 and December 2001. Aviation mishap information was collected from the U.S. Naval Safety Center. Desai et al. (2006) hypothesized that safety climate would improve after a mishap. Moreover, the improvement would be greater following an extremely severe mishap than after a minor mishap. They postulated that after a major mishap "managers may be motivated to direct more resources toward safety than are managers in groups with less severe accident records" (Desai et al., 2006, p. 642). Under this hypothesis, as a result of the increased investment in safety after a mishap, the safety climate improves.

The dependent variable was a safety climate perception construct developed by aggregating each individual's responses to the complete 61-item CSAS. Desai et al. (2006) regressed the safety climate construct by tracking the occurrence of mishaps, grouped by their severity, in periods roughly one year prior to survey measurement and two years prior to survey measurement. A positive association was found between minor or intermediately severe mishaps and future safety climate scores, although no effect was found for major mishaps. These findings suggest a generally positive association between minor or intermediately severe mishaps and perceived safety climate after the incident. However, there are a number of limitations to this study.

Firstly, in a meta-analysis of studies that used a retrospective design to evaluate the relationship between accidents and safety climate, no link was found (Clarke, 2006). Clarke (2006) attributes lack of a relationship to the problem of reverse causation—the individuals' experience of accidents influences the safety climate of the organization. It could be argued that the rationale that safety climate will improve after a mishap is flawed. For example, if the squadron personnel believe that the causes of the mishap have not been addressed, the safety climate may worsen rather than improve (O'Connor et al., 2011).

Another issue relates to the nonrandom variability in the questionnaire responses to the CSAS; this problem makes it difficult to establish the criterion or discriminate validity of the questionnaire itself. Finally, the number of mishaps was likely to be fairly low (the actual number was not reported) and the level of detail regarding the mishaps was very limited. For example, no information was provided beyond the level of severity

of the mishap, and no details were provided to the researchers on the date of the mishap, beyond the financial year in which it occurred.

1.4 PURPOSE OF THE CURRENT PAPER

The study described in this paper used the 12 CSAS items identified by Buttrey et al. (2010) to assess whether the responses to the items can be used to differentiate between squadrons who have had a mishap within a recent period of time, and those that have not, using a prospective design (i.e., the safety climate was measured before the mishap and this data is used to predict which squadrons are more at risk of a mishap). Therefore, the experimental hypothesis is that the responses to the 12 CSAS items can be used to distinguish squadrons which have had a mishap from those that had not.

Thompson, Hilton, and Witt (1998) identified four difficulties with using mishap rates as an indicator of an organization's safety performance. However, we believe that due to the quantity and level of detail we were able to obtain in both the safety climate data and the mishap data, it was possible to address these issues more comprehensively than is generally the case in safety climate research.

- *Mishaps are rare occurrences, and can make the data unreliable*. Although U.S. Naval aviation can be regarded as an HRO, due to the type of flying conducted by the military, mishaps occur with a greater frequency than in other HROs. To illustrate, for U.S. commercial aviation, the accident rate was 0.2 per 100,000 flight hours from 2000 until 2009 (National Transportation Safety Board, 2010), compared to 1.5 major accidents per 100,000 flight hours in U.S. Naval aviation during the same time period (Naval Safety Center, 2010). As there is a relatively high incidence of mishaps in U.S. Naval aviation, this provides a pool of data for assessing the discriminate validity of the CSAS.
- *Mishaps are random events that are not under the direct control of personnel*. "No matter how compliant employees may be with the safety procedures, extraneous random influences can cause or contribute to accidents" (Thompson et al., 1998, p. 17). This is an issue that we cannot completely address with the data set described in paper. However, as the researchers were given access to the mishap causal factors, we were able to use this information to identify subgroups

of mishaps that were caused by specific factors related to safety climate such as personnel leadership and the availability of resources.

- *Mishaps may not be consistently recorded across organizations, and over and under recording causes unreliability*. In U.S. Naval aviation there is a strict criterion for classifying aviation mishaps that is applied across the organization. At the time the mishap data were collected, a class A mishap was classified as one in which the total cost of damage to property or aircraft exceeded $1,000,000, a naval aircraft was destroyed or missing, or any fatality or permanent total disability resulting from the direct involvement of Naval aircraft. A class B mishap was defined as one which did not meet the class A criteria, and in which the total cost of damage to property or aircraft, was more than $200,000, but less than $1,000,000, or which resulted in a permanent partial disability or the hospitalization of three or more personnel. A class C mishap was one that was neither class A nor class B, in which the total cost of damage to property or aircraft, was $20,000 or more, but less than $200,000, or which resulted in an injury requiring five or more lost workdays (Chief of Naval Operations, 2007).
- *The measurement of mishap severity is often a highly subjective issue and, therefore, also causes unreliability*. As discussed above, U.S. Naval aviation has strict criteria for measuring the severity of aviation mishaps.

This study is unprecedented in safety climate literature in terms of the quantity of safety climate data, the level of detail available regarding the causes and circumstances surrounding the mishaps, the fact that the climate data and the mishaps can be linked to specific squadrons and time periods, and the use of a prospective design. Therefore, this study represents a strong chance of identifying a clear link between the safety climate and mishaps in U.S. Naval aviation squadrons, should that link exist.

THIS PAGE INTENTIONALLY LEFT BLANK

2. METHOD

2.1 SURVEY DATA DESCRIPTION

The complete data set used for analysis consisted of 23,442 responses by U.S. Naval aviators to the CSAS. Of the 61 items in the survey, 12 items were used to assess the discriminate validity of the questionnaire (see Buttrey et al., 2010, for the rationale for the selection of these items).

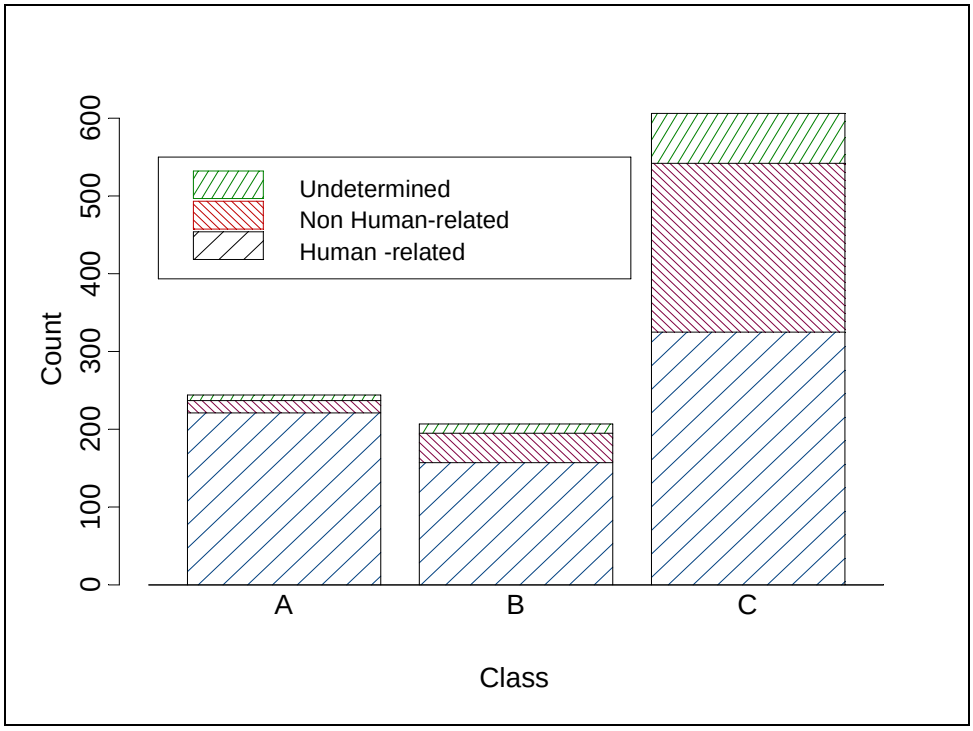
As discussed in Buttrey et al. (2010), many respondents give the very same numeric response to almost all of the items in the survey. In other words, they tended to answer "on mode." Consequently, the variability within the dataset, and the ability to interpret the true meaning of each response, is severely reduced. In order to introduce some variability and to highlight those responses that are meaningful, it was decided to adjust each item by replacing it with the difference from the mode for that individual. For example, one respondent may have answered "3" to almost every item in the questionnaire and then gave a response of "4" to a particular item; another respondent may have a modal answer of "2," but gave one item a "4." We believed that the difference between these response patterns is meaningful and could be used to introduce variability into the dataset. So, Buttrey et al. (2010) replaced each response by the difference between that response and the mode for that respondent. The modally adjusted item data was used in the analysis in this report.

One limitation inherent in this dataset is that not all of the cases are independent; we know that some of the respondents must have answered the survey multiple times. Unfortunately, because the survey is anonymous, that information is unavailable to us. Thus, for the purpose of this analysis, every case is treated as independent. Indeed, it is reasonable to take this stance since the survey methodology is based on the principle that every survey is an independent reflection of the squadron's safety climate at that point in time—a climate which can be expected to change across time.

2.2 MISHAP DATA DESCRIPTION

The Naval Safety Center provided a mishap summary and the identified causal factors of 244 class A, 207 class B, and 606 class C mishaps that took place between October 1999 and August 2009. Every mishap was classified as either "human-related," "not human-related," or "undetermined" based on the identified causal factors of the mishap provided by the Naval Safety Center. A human-related mishap can be defined as any mishap in which at least one of the causal factors was identified as a human error or failure. Human-related mishaps made up 91% of all class A mishaps, 76% of class B, and 54% of class C mishaps. Figure 1 shows the distribution of mishaps by class and human factor categorization.

Figure 1. Mishaps by Class and Human-Related Categorization.



In addition, two researchers independently reviewed the causal factors identified by the U.S Naval Safety Center for each mishap to identify whether the mishap was due to a failure in one or both of the two factors of "personnel leadership" (Factor one) and "availability of resources" (Factor two). For Factor one, a Cohen's kappa of 0.91 resulted (97.3% agreement); and for Factor two, a Cohen's kappa of 0.80 resulted (95.3%

agreement). There were 33 mishaps for which there was not full agreement between researchers; in each incidence, the case was discussed until a consensus was reached. A total of 30.7% of the class A and 10.6% of the class B mishaps were attributed (at least in part) to poor personnel leadership (Factor one), and 21.3% of the class A and 5.3% of the class B mishaps were attributed (at least in part) to lack of availability of resources (Factor two).

It is important to indicate that the rigor and reporting requirements decreased as mishap severity decreased. As a result, class C mishaps were not used in the analysis. In what follows, we will sometimes combine classes A and B, and other times, made clear by the context, we will consider only class A mishaps.

In the original survey data, a total of 309 unique squadrons were identified. About 4% of responses did not identify a specific squadron at all, and those responses could not be used for this analysis. Additionally, we found that 26 class A mishaps and 27 class B mishaps could not be associated with squadrons in the survey data. This is because the events were associated with organizations other than squadrons, or were labeled with names of squadrons that do not exist. Therefore, these mishaps were omitted from further analysis.

We did not have access to data regarding flight hours, missions, carrier landings, or other attributes of squadron makeup that might be partly responsible for differing mishap rates. Mishap rates differ by aircraft class: TACAIR (Tactical Aviation, which includes multirole fighter aircraft such as the F/A-18 Hornet and E/A-8 Prowler) squadrons account for about 47% of the mishaps in our data, with rotary (helicopters, such as the SH-60 Seahawk) and "other" accounting for another 20% each. "Big wing" (large transport and surveillance aircraft, such as the C-130 Hercules and P-3 Orion) squadrons (4%) had fewer mishaps than training squadrons (7%). We have proceeded as if all squadrons are interchangeable within a class in terms of mishap risk.

2.3 MEASURING INTERVALS

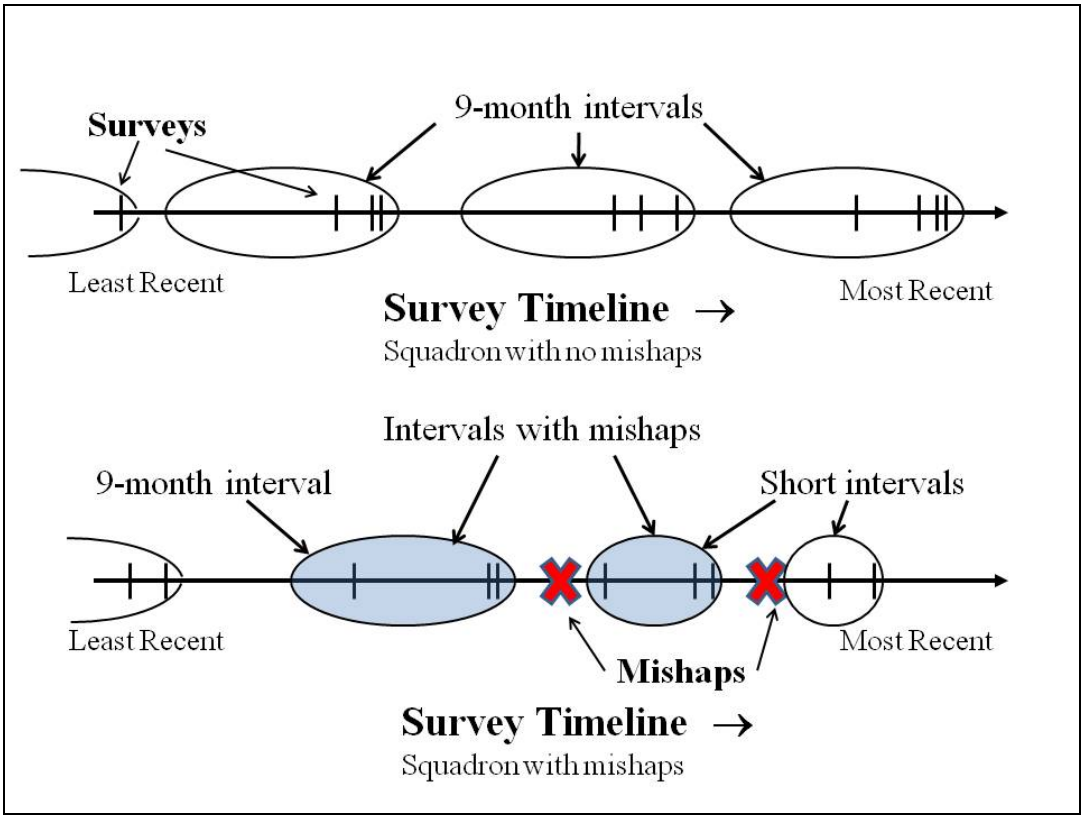
Interval definition. The unit of analysis for this investigation is not the individual response, since these must be examined at the group (i.e., squadron) level, nor the mishap, since we are interested in the safety climate at squadrons that experienced no

mishaps as well as in those that did. In order to be able to investigate the safety climate of each squadron during specific time intervals, we created a unit of analysis we refer to as an "interval." An interval is associated with a particular squadron, it allows for the analysis of squadrons in specific chunks of time for which we have safety climate data and an indication as to whether a mishap occurred. It allows us to attempt to predict the likelihood or risk of mishaps on the basis of the climate data for that squadron, in that period of time. We have chosen nine months as the maximum length of an interval, somewhat arbitrarily, reasoning that changes in the safety climate probably occur over a period of time of about this length.

Interval construction. It is perhaps easiest to think of an interval proceeding backwards in time, rather than forwards, so that the "start" date is more recent than the "end" date. Within each squadron, the first (that is, most recent) interval starts on the date of the most recent survey or the date of the most recent mishap, whichever is most recent. It proceeds backwards in time for 270 days or until one day more recent than the date of the next mishap for that squadron, whichever is less. The next interval for that squadron starts with the more recent survey taken farther into the past than those in the current interval. Each interval refers to a single squadron, and no two intervals overlap.

To recap, an interval begins on the date of a survey (either the squadron's most recent survey, or the most recent survey farther into the past than the last interval determined). It proceeds backwards in time 270 days or until one day more recent than the next mishap. Figure 2 shows examples of intervals constructed for squadrons with no mishaps (upper panel) and for those with at least one mishap (lower panel). In this figure, vertical bars represent dates on which surveys were collected. Every survey belongs to an interval, but some intervals have so small a number of surveys that they are discarded for purposes of analysis.

Figure 2. Intervals as Constructed for Squadrons With No Mishaps (Upper Panel) and for Squadrons With Mishaps (Lower Panel).



Interval Classification. Intervals that consist of a set of surveys followed (in the usual time sense) by a mishap are classified as "Yes," meaning that a mishap took place after the surveys in that interval were collected (these are the red "X"s in Figure 2; note that the 9-month period of the interval "starts" not on the date of the mishap, but on the date of the first survey following the mishaps). All other intervals are classified as "No." The goal of the analysis then becomes to predict which intervals will have mishaps, based on the safety climate data of the associated squadron. To the extent that accurate predictions can be made, it may be possible to identify squadrons at higher risk of a mishap and to take corrective action.

Assumptions. It is worth remarking on the assumptions underlying interval construction here. We assume that the safety climate in a squadron is roughly constant over the length of the interval. Observations are thought of as independent within an

interval and between intervals. Safety climate is "reset" after a mishap; surveys before a mishap and surveys after a mishap are never included in a single interval.

As noted above, we assume that mishap rates in squadrons are related to safety climate. Different sorts of squadrons—having different aircraft, missions, numbers of carrier landings and so on—can be expected to have different mishap rates. In our model, that difference is reflected by the inclusion of a "community" predictor. That is, our model treats communities as differing, in mishap rate, by a number that is constant on the log-odds scale, if other indicators of safety climate are all alike.

3. ANALYSIS

3.1 DATA SETS

Our data can take on several specific forms: first, we can either include only class A mishaps, or include both classes A and B. (As mentioned above, we have not used class C mishaps in this analysis.) Second, we can use (1) all the events of those types, or (2) restrict consideration just to those mishaps determined to be human-factors related, or (3) consider only mishaps associated with a failure in Factor one (personnel leadership), or (4) consider only mishaps associated with a failure in either Factor one or Factor two (availability of resources). We do not consider Factor two mishaps by themselves because they are few in number. Therefore, there are eight different data sets. A logistic regression model has been built with each.

3.2 RESPONSE VARIABLE

Each interval is either associated with a following mishap or it is not. The existence of a mishap constitutes the response variable. The object of the analysis is to produce a model that is capable of determining squadrons at elevated risk of mishap, i.e., squadrons whose predicted probability of mishap is high.

3.3 PREDICTOR VARIABLES

To predict mishaps we start with the squadron's "Community," a categorical variable that includes "Big Wing," "rotary," "TACAIR," "Training," and "Other." Every model included community as a predictor. We then consider for inclusion into the model the average scores, for each survey in the interval, on each of 12 items identified in Buttrey et al. (2010). We also considered the two factor scores described in Buttrey et al. (2010). The two factors (Factor one: personnel leadership; items 4, 9, 13, 16, 17, 19, and 48; Factor one: lack of availability of resources; items 31, 32, 50, 52, and 55) were based on exploratory and confirmatory factor analyses using the 12 CSAS items listed earlier. These two factors measure the extent to which perceptions of safety culture vary across respondents in the squadron, which might itself be an indicator of concern.

3.4 BUILDING LOGISTIC REGRESSION MODELS

We analyze the data with logistic regression models. The logistic regression model supposes that the squadron in interval i experiences a mishap with probability p_i, where p_i is related to the set of predictors $X_1, X_2, \ldots, X_k$ through $\log(p_i/(1 - p_i)) = \beta_0 + \beta_1 X_{i1} + \ldots + \beta k X_{ik}$. The quantity $p_i/(1 - p_i)$ is interpreted as the odds of an event in interval i (e.g., an odds of 0.111 is read "1 to 0.111," more familiarly expressed as "9 to 1," and corresponds to a probability of 0.10). For these models, we assume that mishaps occur independently from one squadron to another, and independently within a single squadron in non-overlapping intervals.

Each model was constructed starting with the "Community" variable as a predictor. Subsequent terms were added sequentially, selecting at each stage the variable that reduced the value of the Akaike Information Criterion (AIC) (Hilbe, 2009), as long as the p-value associated with that term (from the usual χ^2 test for changes in deviance) was also smaller than 0.05. The process stopped when no term both reduced AIC and also had a p-value < 0.05. The drawbacks of sequential model-building have been documented (see, for example, Harrell, 2001, p. 56 in the linear regression context). We emphasize here that our goal is not to select the "correct" set of variables, nor to obtain accurate estimates of individual regression coefficients—we do not claim our models are "right"—but rather to produce simple models that make reasonable predictions of mishap probabilities.

3.5 EVALUATING THE MODELS

We use a number of measures to evaluate the logistic regression models. In this section, we describe each of these briefly; the actual results are found in the "Results" section.

Hosmer-Lemeshow test p-value. The Hosmer-Lemeshow test (Hosmer and Lemeshow, 1989) presents an intuitively appealing comparison of observed outcomes with predicted probabilities. Intervals are sorted from smallest predicted probability of mishap to largest, and then divided into, say, ten groups. Within each decile group, the proportion of intervals with mishaps is computed, as is the average predicted probability

of the intervals in the decile group. If the model is performing well, the group averages of the predicted probabilities ought to be quite close to the actual proportions of intervals with mishaps. Hosmer and Lemeshow describe an approximate test for the null hypothesis that the model is performing adequately, based on the chi-squared measure of closeness of the two sets of numbers. If the *p*-value associated with the chi-squared test is larger than 0.05, we can describe the model as "adequate" under this test.

Decile odds ratios. As an additional piece of information from the Hosmer-Lemshow test, we computed the ratio of the (actual) mishap odds in the 10% of intervals judged, by the model, to most likely be associated with a mishap to the odds in the least likely 10%. This is one measure of the "lift" of the model, since it measures the extent to which the model can discriminate between intervals of higher and lower risk. In Table 1, this ratio is infinite, since there were no mishaps in the decile of intervals with the lowest predicted probability. Therefore, we reported the ratio of the expected odds: in this case, that ratio is 0.648/0.016 = 40.5. This suggests, then, that the set of intervals identified as highest risk by the model has odds of a mishap perhaps about 40 times greater than the set of intervals identified as lowest risk.

Area under Receiver Operating Characteristic (ROC) curve. One commonly used measure of the discriminatory power of a logistic regression is the area under the ROC curve. Consider declaring a threshold predicted probability, so that any interval whose predicted probability is greater than the threshold is predicted to have a mishap, while those whose probabilities are lower, are predicted to not have a mishap. Each threshold has an associated true positive rate (the "sensitivity," the proportion of intervals with predicted mishaps that, in fact, do have one) and false positive rate (the proportion of intervals with predicted mishaps that do not actually have one, one minus the specificity). The curve traced out by these two rates, as the value of the threshold probability changes, is called the ROC curve.

In a poor model, predictions will be close to random, so these two rates will be similar for any value of the cutoff. Therefore, the ROC curve will follow the line $Y = X$, and the area under the curve (above $Y = 0$ and between $X = 0$ and $X = 1$) will be 0.5. Larger values indicate more discriminatory power; Hilbe (2009, p. 258) notes that

"typical values…range from about 0.6 to 0.9" and "values of 0.95 and greater are highly unlikely." We compute the area under the curve using the lrm() function in S-Plus's Design library, an add-in due to Harrell (see Harrell, 2001). Harrell (2001, p. 247) also notes that "a model having [area under the ROC] greater than about 0.8 has some utility in predicting the responses of individual subjects."

The area under the ROC curve has a second, fairly intuitive interpretation. Imagine picking an interval associated with a mishap at random, and independently a second interval, with no mishap. Then the probability that the model's prediction for the first interval is higher than the prediction for the second is precisely the value of the area under the ROC curve (Harrell, 2001).

4. RESULTS

4.1 RESULTS FOR ALL MODELS

Table 1 shows the results from each of the eight models considered. Column one names the set of mishaps considered in each model (A only, or A and B). The second column gives the subset considered (all data, human-related mishaps only, Factor one mishaps only, or Factors one and two mishaps). "Omit" denotes the fact that for some models certain communities did not experience any mishaps at all. Those communities (B for Big Wing, T for TACAIR) are excluded from those models. The "n/m" column gives the number of intervals and the number of mishaps used in each model. (These numbers can change slightly between similar models because an interval's endpoint can be defined by the presence of a mishap and, in some subsets, some mishaps are omitted.) DF denotes the number of degrees of freedom used (that is, parameters estimated) by the model; HL10p gives the p-value associated with the Hosmer-Lemeshow test, based on ten groups. DecRat is the ratio of the observed odds in the tenth decile to the observed odds in the first. (In cases where the intervals in the first decile had no mishaps, we have substituted the corresponding values from the expected column, and appended an asterisk.) Finally, AUC shows the area under the ROC curve.

Table 1. Results from All Models, Showing Hosmer-Lemeshow p-value (HL10p), Tenth-Decile-to-First-Decile Observed Odds Ratio (* = Expected), Area Under Curve (AUC), # B= Big Wing, T=TACAIR.

Mishaps	Sub	Omit	n/m	DF	R^2	HL10p	DecRat	AUC
A+B	All		531/70	6	0.117	0.77	13.5	0.709
A	All		521/29	5	0.097	0.81	20.4*	0.721
A+B	HF		514/53	5	0.095	0.50	8.3	0.701
A	HF		518/26	5	0.108	0.81	22.6*	0.735
A+B	1, 2		513/52	6	0.123	0.56	8.3	0.723
A	1, 2	B, T#	243/9	3	0.130	0.93	52.3*	0.765
A+B	1		512/51	6	0.122	0.61	9.2	0.722
A	1	B, T	242/8	4	0.206	0.98	183*	0.831

Table 2 shows, for each of the models, the set of terms that entered into the model (not including Community, which is present in every model). These additional terms are sorted from greatest to least contribution to the model in terms of AIC (i.e., the first term

listed is the one whose removal would cause the largest increase in AIC). Also shown are some actual proportions of intervals containing mishaps. The "P10,10" column shows the proportion of intervals with mishaps in the tenth decile of predicted probability. In other words, among the intervals in the tenth of intervals with the highest predicted probability, about 34% experienced a class A or class B mishap (top row). Only 3.7% of the intervals among the tenth with the lowest predicted probability (column "$P_{10,\ 1}$") experienced a class A or class B mishap. The final two columns show the corresponding proportions for the first and third terciles (the top 33% and bottom 33% in predicted probability). Note that these numbers are proportions, whereas the numbers in Table 4 are in terms of odds. Also note that while the deciles and terciles are constructed using predicted probabilities from the models, the proportions reported in this table are the actual mishap rates across all relevant intervals. In this context, the mishap rate is the proportion of intervals that contain mishaps.

Table 2. Terms and Predictive Probabilities for Each Model.

Mishaps	Subset	Terms	$P_{10,10}$	$P_{10,1}$	$P_{3,3}$	$P_{3,1}$
A+B	All	μ17, μ4	.339	.037	.237	.057
A	All	μ17	.154	.000	.101	.006
A+B	HF	μ17	.250	.038	.199	.040
A	HF	μ17	.173	.000	.097	.012
A+B	1, 2	μ17, μ4	.250	.038	.193	.047
A	1, 2	μ17	.160	.000	.074	.000
A+B	1	μ17, μ4	.269	.038	.193	.041
A	1	μ17, μ13	.120	.000	.074	.000

4.2 ITEM ANALYSIS

Only three items in the questionnaire were shown to be predictive of mishaps—items 4, 13, and 17. We would expect that respondents who answered the three predictive items with a score higher than their median score would be expressing a strong belief that there was a good safety culture in the squadron, and so we would expect fewer mishaps in squadrons in which this question was frequently given high scores. As can be seen from Table 2, this was the case for items 4 (*My command closely monitors proficiency and currency standards to ensure aircrew are qualified to fly*), and 13 (*In my command, we believe safety is an integral part of all flight operations*). Table 3 shows the numbers

of intervals and combined number of class A and B mishaps, and the mishap rate, by community, for squadrons whose average score was in the lowest 75% (left columns) and for those in which it was in the highest 25% (right columns). The difference in the overall mishap rates (bottom row) was not statistically significant for item 4 (chi^2= 0.87, n.s., df=1), or item 13 (chi^2= 2.66, n.s., df=1).

For item 17 (*In this command, an aviator who persistently violates flight standards and rules will seriously jeopardize his/her career*) the pattern of responses was not as expected. Squadrons in which the average score for item 17 was in the lowest quartile had lower mishap rates than where it was not (the reverse of what would be expected; see Table 3).

Table 3. Mishap Rate by Community and Quartile of Squadron's for Items 4, 13, and 17.

Item 4	**Lower three quartiles**			**Fourth quartile**		
Community	**Total**	**Mishaps**	**Rate**	**Total**	**Mishaps**	**Rate**
BigWing	94	6	.064	33	0	0
Rotary	125	16	.128	39	6	.154
TACAIR	97	18	.186	47	7	.149
Training	36	7	.194	2	1	.500
Other	45	9	.200	13	0	0
Total	**397**	**56**	**.141**	**134**	**14**	**.104**
Item 13	**Lower three quartiles**			**Fourth quartile**		
Community	**Total**	**Mishaps**	**Rate**	**Total**	**Mishaps**	**Rate**
BigWing	100	6	.060	27	0	0
Rotary	116	19	.164	48	3	.062
TACAIR	105	20	.190	39	5	.128
Training	30	6	.200	8	2	.250
Other	43	7	.163	15	2	.133
Total	**394**	**58**	**.147**	**137**	**12**	**.088**
Item 17	**Lower three quartiles**			**Fourth quartile**		
Community	**Total**	**Mishaps**	**Rate**	**Total**	**Mishaps**	**Rate**
BigWing	105	5	.048	22	1	.045
Rotary	139	18	.129	25	4	.160
TACAIR	83	9	.108	61	16	.262
Training	33	4	.121	5	4	.800
Other	38	4	.105	20	5	.250
Total	**398**	**40**	**.101**	**133**	**30**	**.226**

4.3 UNPREDICTIVE ITEMS

Table 4 shows the mean response to the 12 items for those squadrons that had a mishap in the interval, and those that did not have a mishap (the factor scores for the two factors identified by Buttrey et al. [2010] were also not predictive of mishaps). As discussed above, the mode is calculated from the entire set of responses from that respondent in that survey. A positive value means the aviators in the squadron gave a higher score to the item than their modal response; a negative value means they gave a lower score to the item than their modal response.

Table 4. Mean Responses to the 12 Items
(Note: items in italics represent those that were predictive of mishaps).

Item	No Mishap Interval	Mishap Interval	Difference in Mean	Overall Mean
4	*0.094*	*0.049*	*–0.045*	*0.088*
9	–0.280	–0.242	0.038	–0.275
13	*0.200*	*0.201*	*0.001*	*0.200*
16	0.158	0.173	0.015	0.160
17	*0.010*	*0.088*	*0.077*	*0.021*
19	0.058	–0.002	–0.060	0.050
31	–0.807	–0.849	–0.042	–0.812
32	–0.481	–0.528	–0.047	–0.487
48	–0.176	–0.164	0.012	–0.175
50	–0.636	–0.575	0.061	–.0628
52	–0.206	–0.210	–0.004	–0.207
55	–0.443	–0.420	0.023	–0.440

5. DISCUSSION

5.1 PREDICTING MISHAP LIKELIHOOD

In this paper, we analyzed data from the CSAS in an effort to determine whether there is a relationship between survey responses and mishap probability. Since surveys are distributed across time, our unit of analysis is a time interval during which surveys were taken, which may or may not end with a mishap. We built logistic regression models using the survey data. We built models that included (1) class A mishaps by themselves; (2) class A and class B mishaps; (3) a subset of the mishaps for which there were human-related causes; (4) a subset of the mishaps that had a causal factor related to "personnel leadership"; and (5) a subset of mishaps, which had a causal factor related to either "availability of resources" or "personnel leadership."

In all cases, the models that attempted to predict both class A and class B mishaps together, performed better than the models that used only subsets of the mishaps' data. This may be because the larger number of mishaps in the former case makes higher precision possible. Only one model has an AUC of 0.8 or more, which suggests (by Harrell's rule) that the utility of these models lies not in predicting individual intervals, but rather in identifying squadrons that appear, because of recent survey results, to be at high risk. Here the results of the decile odds ratios are striking. In several cases, the models were able to accurately identify intervals with no mishaps at all. It is also important to indicate that despite using the factors' scores calculated from the factor analysis of the items reported by Buttrey et al. (2010), these scores were not found to be predictive of mishap risk.

The fact that the squadrons with low probabilities of mishaps have personnel who more strongly agree with the need to both monitor personnel and integrate safety and operations than squadrons with a higher probability of mishaps is to be expected. However, the fact that aircrews in low-probability-of-mishap squadrons agreed less strongly than aircrews in higher-probability-of-mishap squadrons that persistent rule violators will jeopardize their careers as compared to personnel in low-probability-of-mishap squadrons, on the face of it, is unexpected. However, the unusual pattern of

responses to this item supports Reason's (1997) premises of an effective reporting and just culture.

In a reporting culture, the workforce willingly report their errors and near-misses. Their willingness to report is dependent on how the organization handles blame and punishment. In a "no blame" culture, there is no accountability for one's actions. This is clearly neither feasible, nor desirable. However, neither is a "blame" culture in which individuals are identified and punished for errors and violations. A "blame" culture encourages members of the workforce to cover up errors and does not encourage organizational learning. Therefore, what is required is a just culture, "an atmosphere of trust in which people are encouraged, even rewarded, for providing essential safety-related information, but in which they are also clear about where the line must be drawn between acceptable and unacceptable behavior" (Reason, 1997, p. 195). The pattern of responses to item 17 seem to suggest that the higher-probability-of-mishap squadrons may adopt more of a "blame" culture, compared to the lower-probability-of-mishap squadrons, which may have more of a "just culture" in which there is not a blanket punishment of all violators.

5.2 NONPREDICTIVE ITEMS

Although this fact is not useful for the purposes of assessing the risk of mishaps, it was found that respondents tended to answer below their modal response for the questionnaire when asked about the availability of resources (item 31- *I am provided adequate resources (time, staffing, budget, and equipment) to accomplish my job*), flight time (item 32- *My command provides the right number of flight hours per month for me to fly safely*), and communication within the command (item 55- *Within my command, good communications flow exists up and down the chain of command*).

Desai et al. (2006) postulated that after a major mishap, managers may be motivated to direct more resources toward safety than are managers in groups with less severe mishap records. However, it was found that lack of resources was an issue for all squadrons. It is possible that, due to the current economic climate, the availability of resources has become increasingly relevant. The lack of flight time is a related issue to lack of resources. With skill-based errors being the most commonly cited causal factor

for both F/A-18 and H-60 class A mishaps (70.2% and 81.3%, respectively; O'Connor, Cowan, & Alton, 2010), the need for aviators to remain proficient is clearly crucial for both safety and performance. Finally, the responses to item 55 highlight the importance of interactions between different levels of seniority. Specifically, senior leadership participation and involvement in work and safety activities, as well as frequent, informal communications between workers and management, are recognized as critical behaviors. Therefore, although not found to be predictive of mishaps, these are areas that should be addressed.

THIS PAGE INTENTIONALLY LEFT BLANK

6. CONCLUSIONS

The data analysis carried out in this paper was a comprehensive attempt to examine the strength of the link between safety climate questionnaire data and mishap probability. Within the context of previous work that has been carried out in the safety climate literature the sample size was much larger than is typical, the number of mishaps was greater, and the detail available on the causes of the mishaps was unparalleled. Additionally, a prospective research design was employed. Nevertheless, only three of the CSAS items were found to have some limited value in predicting mishap risk. Our findings would seem to support the premise of other researchers (e.g., Clarke, 2006; Guldenmund, 2007; Nahrgang, Morgeson, & Hofmann, 2007) that safety climate and safety performance (as measured by mishap data) are, at best, weakly related.

Mishaps are rare events, and they describe only part of the spectrum of risks pertaining to a work system. We suggest that measuring workers' self-reported safety attitudes and behavior is an alternative way to assess the discriminate validity of safety climate. In fact, Thompson et al., (1998) use self-reported safety behavior as their preferred criteria for safety research. Though less "objective" than accident and incident rates, members of the workforce are likely to be sensitive to the safety behaviors of coworkers.

Two different aspects of behavior have typically been distinguished—task behaviors and contextual behaviors. Task behaviors, such as rule compliance, are prescribed as part of the job and are fundamental to safety. Contextual behaviors, such as safety proactivity or safety initiative, support the broader organizational context and are thought to have the potential to enhance the safety of the organization as a whole. Such behaviors represent a more advanced form of safety awareness.

Contextual behaviors underlie the safety culture of an organization and may be more critical as indicators of an organization's safety performance than mishaps alone. Herein lies the greatest strength of the concept. Safety climate introduces the notion that the likelihood of accidents occurring can be predicted on the basis of certain organizational factors. These organizational factors can be used as leading indicators to

identify, in advance, the strengths and weaknesses within an organization that influence the likelihood of accidents occurring. Once weaknesses are identified, remedial actions can be taken.

The safety climate literature is dominated by reports of the development and use of different questionnaires, with some attempts to establish the construct validity of the instrument. However, if the goal is to make advances in the safety climate, researchers must also attempt to establish the discriminate validity of their tools. The research reported in this paper, in addition to the findings from others, suggests that establishing the discriminate validity through attempting to link safety climate with mishaps would appear to be a largely futile effort and may actually be detrimental in terms of realizing the other benefits that climate surveys can bring. Therefore, safety climate researchers would be better advised to attempt to establish the discriminate validity of their questionnaire through self-reported safety attitudes and behaviors.

LIST OF REFERENCES

Braithwaite, J. (1985). *To punish or persuade*. Albany: State University of New York Press.

Buttrey, S. O'Dea, A., O'Connor, P., & Kennedy, Q. (2010). *An evaluation of the construct validity of the command safety assessment survey.* Monterey, CA: Naval Postgraduate School.

Chief of Naval Operations. (2007). *Naval aviation safety program.* OPNAV Instruction 3750.6 R. Washington, D.C.: Author.

Clarke, S. (2006). Contrasting perceptual, attitudinal and dispositional approaches to accident involvement in the workplace. *Safety Science*, *44*, 537-550.

Cohen, H., & Cleveland, R. (1983, March). Safety program practices in record-holding plants. *Professional Safety*, 26-33.

Cohen, A., Smith, M., & Cohen, H. (1975). *Safety program practices in high versus low accident rate companies- and interim report* (Publication no. 75-185). Cincinnati: National institute for Occupational Safety and Health: U.S. Department of Health Education and Welfare.

Desai, V.M., Roberts, K.H., & Ciavarelli, A.P. (2006). The relationship between safety climate and recent accidents: Behavioral learning and cognitive attributions. *Human Factors*, *48*, 639-650.

Diaz, R.T., & Cabrera, D.D. (1997). Safety climate and attitude as evaluation measures of organizational safety. *Accident Analysis and Prevention*, *29*(5), 643-650.

Donald, I., & Canter, D. (1994). Employee attitudes and safety in the chemical industry. *Journal of Loss Prevention in the Process Industries*, *7*(3), 203-208.

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Safety climate: Identifying the common features. *Safety Science*, *34*, 177-192.

Guldenmund, F. (2007). The use of questionnaires in safety culture research – An evaluation. *Safety Science*, *45*(6), 723-743.

Hale, A.R., & Hovden, J. (1998). Management and culture: The third age of safety. A review of approaches to organizational aspects of safety, health and environment. In A.M. Feyer & A. Williamson (Eds.), *Occupational injury: Risk prevention and intervention*. (pp. 117-119) London: Taylor and Francis.

Harrell, Jr., F.E. (2001). *Regression modeling strategies*. New York: Springer.

Hilbe, J.M. (2009). *Logistic regression models*. Boca Raton, FL: Chapman & Hall.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.

Libuser, C.B. (1994). Organizational structure and risk mitigation (Ph.D. Dissertation). Los Angeles, CA: University of California at Los Angeles.

Mearns, K., & Flin, R. (1999). Assessing the state of organizational safety – Culture or climate. *Current Psychology*, *18*(1), 5-17.

Mearns, K., Flin, R., & O'Connor, P. (2001). Sharing "worlds of risk": Improving communication with crew resource management. *Journal of Risk Research*, *4*(4), 377-392. doi:10.1080/13669870110063225.

Mearns, K., Rundmo, T., Flin, R., Gordon, R., & Fleming, M. (2004). Evaluation of psychosocial and organizational factors in offshore safety: A comparative study. *Journal of Risk Research*, *7*(5), 545-561.

Nahrgang, J.D., Morgeson, F.P., & Hofmann, D.A. (2007). Predicting safety performance: A meta-analysis of safety and organizational constructs. Paper presented at the 22nd Annual Conference of the Society for Industrial and Organizational Psychology, New York, NY.

National Transportation Safety Board. (2010). Aviation accident statistics. Retrieved from www.ntsb.gov/aviation/Table5.htm.

Naval Safety Center. (2010). Class A flight mishaps. Retrieved from www.public.navy.mil/navsafecen/Documents/statistics/execsummary/Daily_Mishap_Stats.ppt#506,1,Slide 1.

O'Connor, P., Cowan, S., & Alton, J. (2010). A comparison of leading and lagging indicators of safety in Naval aviation. *Aviation, Space and Environmental Medicine*, *81*, 677-682.

O'Connor, P., O'Dea, A., Kennedy, Q., & Buttrey, S. (2011). Measuring safety climate in the aviation industry: A review and recommendations for the future. *Safety Science, 49*, 128-138.

O'Dea, A., O'Connor, P., Kennedy, Q., & Buttrey, S. (2010). *A review of the safety climate literature as it relates to Naval aviation.* Monterey, CA: Naval Postgraduate School.

Reason, J. (1997). *Managing the risks of organizational accidents.* Aldershot, UK: Ashgate.

Simard, M., & Marchand, A. (1994). The behaviour of first line supervisors in accident prevention and effectiveness in occupational safety. *Safety Science*, *17*, 169-185.

Thompson, R.C., Hilton, T.F., & Witt, L.A. (1998). Where the safety rubber meets the shop floor: A confirmatory model of management influence on workplace safety. *Journal of Safety Research*, *29*(1), 15-24.

Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on micro-accidents in manufacturing jobs. *Journal of Applied Psychology*, *85*(4), 487-596.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Research Office (Code 09).....1
   Naval Postgraduate School
   Monterey, CA 93943-5000

2. Dudley Knox Library (Code 013).....2
   Naval Postgraduate School
   Monterey, CA 93943-5002

3. Defense Technical Information Center.....2
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, VA 22060-6218

4. Richard Mastowski (Technical Editor).....2
   Graduate School of Operational and Information Sciences (GSOIS)
   Naval Postgraduate School
   Monterey, CA 93943-5219

5. Defense Safety Oversight Council.....1
   110 Army Pentagon, Room 3E464
   Washington, D.C. 20310-0110

6. Associate Professor Samuel E. Buttrey 1
   Operations Research Department
   Naval Postgraduate School
   Monterey, CA 93943-5219

7. Lecturer Quinn Kennedy.....1
   Operations Research Department
   Naval Postgraduate School
   Monterey, CA 93943-5219

8. Dr. Paul O'Connor.....1
   Room 342
   J.E. Cairns School of Business & Economics
   National University of Ireland
   Galway, Ireland

9. Associate Professor Nita L. Shattuck.....1
   Operations Research Department
   Naval Postgraduate School
   Monterey, CA 93943-5219

10. Lt Col Valerie E. Martindale 1
    Plans and Programs Office
    711 Human Performance Wing
    2610 7th Street, Bldg 441
    Wright-Patterson AFB, OH 45433