2010

An evaluation of the construct validity of the command safety assessment survey

Buttrey, Samuel L.; O'Connor, Paul; O'Dea, Angela; Kennedy, Quinn

Monterey, California. Naval Postgraduate School

NPS-OR-10-004

NAVAL
POSTGRADUATE
SCHOOL

MONTEREY, CALIFORNIA

An Evaluation of the Construct Validity of the Command Safety Assessment Survey

by

Samuel L. Buttrey
Paul O'Connor
Angela O'Dea
Quinn Kennedy

December 2010

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, CA 93943-5001

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, 4000 Defense Pentagon, Washington, D.C. 20301-4000 and funded by the Defense Safety Oversight Council.

This report was prepared by:

SAMUEL L. BUTTREY
Associate Professor of
Operations Research

PAUL O'CONNOR
Assistant Professor of Operations Research

ANGELA O'DEA
Research Associate Professor

QUINN KENNEDY
Lecturer 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 VAN BIBBER
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)* 12-2010	**2. REPORT TYPE** Technical Report	**3. DATES COVERED** *(From - To)*

4. TITLE AND SUBTITLE
An Evaluation of the Construct Validity of the Command Safety Assessment Survey

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Samuel L. Buttrey, Paul O’Connor, Angela O’Dea, and 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 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-OR-10-004

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense OSD Readiness Programming and Assessment, Defense Safety Oversight Council, 4000 Defense Pentagon, Washington, D.C. 20301-4000.

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
The purpose of this study was to evaluate the construct validity of the 61-item command safety assessment survey (CSAS) using the responses of 110,014 U.S. Naval aircrew. Utilizing a combination of exploratory and confirmatory factor analysis, we were unable to identify a stable factor structure from the CSAS data. We believe that this finding was because of the effect of the nonconstant variance of the data, which was due to the large proportion of participants using a satsificing strategy (respondents interpret each question superficially and select what they believe to be a reasonable answer). Fortunately, since 2006, the amount of time taken by respondents to complete the survey was collected. This “time to complete” data was then used as a metric to identify the respondents utilizing an optimizing strategy (respondents generating the optimum answer). A total of 2,344 responses were retained for analysis. We also elected to discard the CSAS items that had low variability. Using the truncated dataset, we carried out an exploratory and confirmatory factor analysis and were able to establish a stable, 12-item, two-factor (named personnel leadership, and integration of safety and operations) model. Based on the analysis, recommendations for improving the CSAS were made.

15. SUBJECT TERMS Safety Climate, Aviation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	47	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The purpose of this study was to evaluate the construct validity of the 61-item command safety assessment survey (CSAS) using the responses of 110,014 U.S. Naval aircrew. Utilizing a combination of exploratory and confirmatory factor analysis, we were unable to identify a stable factor structure from the CSAS data. We believe that this finding was because of the effect of the nonconstant variance of the data, which was due to the large proportion of participants using a satsificing strategy (respondents interpret each question superficially and select what they believe to be a reasonable answer). Fortunately, since 2006, the amount of time taken by respondents to complete the survey was collected. This "time to complete" data was then used as a metric to identify the respondents utilizing an optimizing strategy (respondents generating the optimum answer). A total of 2,344 responses were retained for analysis. We also elected to discard the CSAS items that had low variability. Using the truncated dataset, we carried out an exploratory and confirmatory factor analysis and were able to establish a stable, 12-item, two-factor (named personnel leadership, and integration of safety and operations) model. Based on the analysis, recommendations for improving the CSAS were made.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The Command Safety Assessment Survey (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 61-item CSAS is completed on-line, and responses are obtained for each item on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). The responses are anonymous. In 2004, Vice Admiral Zortman declared it mandatory for all squadrons to complete the CSAS semiannually and within 30 days following a change of command (Zortman, 2004). The results of a squadron's survey are only available to the Commanding Officer (CO). However, aggregated data is made available to all COs for comparison of their squadron's performance with their peers.

Safety climate describes employees' perceptions, attitudes, and beliefs about risk and safety (Mearns & Flin, 1999). The theoretical background underpinning the CSAS is based upon the work carried out by Roberts et al on high reliability organizations (HRO; Desai et al., 2006). Libuser (1994) developed a theoretical Model of Organizational Safety Effectiveness (MOSE) that identified five major areas relevant to organizations in managing risk and developing a climate to reduce accidents. 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
- **Command and control** – policies, procedures, and communication processes used to mitigate risk

Due to legal reasons, we cannot list the CSAS items in this report. For a complete list of the CSAS items, along with the MOSE areas from which they are drawn, see

Adamshick (2007). For a more detailed discussion of the CSAS, and a general discussion of the use of safety climate surveys in aviation, see O'Dea, O'Connor, Kennedy, and Buttrey (2010).

The purpose of the analysis discussed in this report is to establish the construct validity of the CSAS. Construct validity is concerned with the extent to which the measurement instrument measures what it is intended to measure. In other words, to what extent does this safety climate survey actually measure the perceived safety climate? Identifying a reliable factor structure, that is consistent with theory, helps researchers substantiate claims regarding the validity of the questionnaire.

II. STUDY 1: ANALYSIS OF ALL CSAS DATA FROM 2000 UNTIL 2008

A. STUDY 1: METHOD

Following approval by Commander Naval Air Forces, all of the CSASs that were administered from July 2000 until July 2009 were obtained (N=110,014 surveys) for research. A total of 6% of respondents were Navy aircrew (people whose job involves flying in aircraft such as naval aviators and enlisted aircrew), 31% were Marine Corps aircrew, and the remainder were identified as civilians or "other." Of the respondents, 67% were officers and 33% were enlisted personnel (about 0.3% of respondents were civilians or warrant officers). Table 1 provides a summary of the numbers of respondents across year and aircraft type: TACAIR (Tactical Aviation, which includes multirole fighter aircraft such as the F/A-18 Hornet and E/A-8 Prowler), rotary (helicopters such as the SH-60 Seahawk), and big wing (large transport and surveillance aircraft such as the C-130 Hercules and P-3 Orion).

Community	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
BigWing	455	1,549	833	2,624	3,133	5,480	5,534	7,195	6,083	1,625	**34,511**
Helo	575	2,361	1,357	2,994	2,789	4,887	5,558	6,599	6,613	1,671	**35,404**
Tacair	261	934	522	1,687	1,492	2,786	2,960	3,243	3,123	771	**17,779**
Training	132	593	503	1,107	1,844	1,821	2,041	2,941	2,838	760	**14,580**
Missing	0	400	298	867	804	1,215	1,209	1,149	1,444	354	**5,942**
Total	**1,423**	**5,837**	**3,513**	**9,279**	**10,062**	**16,189**	**17,302**	**21,127**	**20,101**	**5,181**	***110,014***

Table 1. Number of Respondents, By Year and Community

Given the frequency with which the questionnaire is completed, in the majority of cases, the same respondent will have answered the survey more than once. Indeed, it is possible that some respondents may have answered the survey eight times. Responses from a single individual might be expected to be correlated, even in the face of organizational evolution; however, since responses are not individually identifiable, they have been treated as independent observations.

B. STUDY 1: ANALYSIS

Data Screening. Prior to carrying out factor analysis, the proportion of missing values and variability of responses were examined for each item. The preanalysis of the data is one of the most important steps, and yet is also one of the most often overlooked. This step is crucial to ensure that the data is appropriate for exploratory factor analysis (EFA).

Factor Analysis. Factor analysis refers to a model, and a set of techniques, in which observed responses are presumed to be based on a set of underlying (and unobserved) factors (Harmon, 1976). Evidence from the safety climate literature suggests that number of factors is likely to be relatively small in number (i.e., less than 12; O'Dea et al., 2010). EFA was performed using data collected in years up to and including 2007, using routines built into the S-Plus statistical package (Insightful Corp., 2005). Data from 2008 and later was reserved to test the factor structure identified as part of the EFA, using confirmatory factor analysis (CFA). CFA seeks to determine whether the number of factors, and the loadings of measured variables on them, is consistent with theory or with previously determined structure. It is imperative that the construct validity of the CSAS is determined, in order to establish the usefulness of the tool in measuring safety climate. A linear structural relations approach to CFA, as implemented in EQS for Windows, was used.

C. STUDY 1: RESULTS

1. Data Screening

Missing Values. Of the 6.7 million total questions (110,014 respondents x 61 items), about 300,000 (4.5%) were missing. About 32% of the missing values (1.4% of the total number of responses) came from four items: 7 (*Human Factors Councils have been successful in identifying aircrew members who pose a risk to safety*), 8 (*Human Factors Boards have been successful reducing chances of an aircraft mishap due to high-risk aviator*), 56 (*my command has good two-way communication with external commands*), and 59 (*the Aviation Safety Officer position is a sought after billet in my command*). Because at least one of these items was missing in at least 17% of all responses, they were discarded for the remainder of this analysis. Missing data rates in the other items ranged from 1.0% to 12.7%, with a median of 2.4%. Figure 1 shows the average number of missing values per case by year. This figure shows that respondents, on average, omitted answers to one of the frequently missed items (blue bars) and to around two of the remaining 59 items (brown bars). Although the exact numbers vary, there is no evidence of a long-term trend. Missing values in the remaining items were replaced by the median of the nonmissing values for that case. From here forward, results are reported with those replacements included.

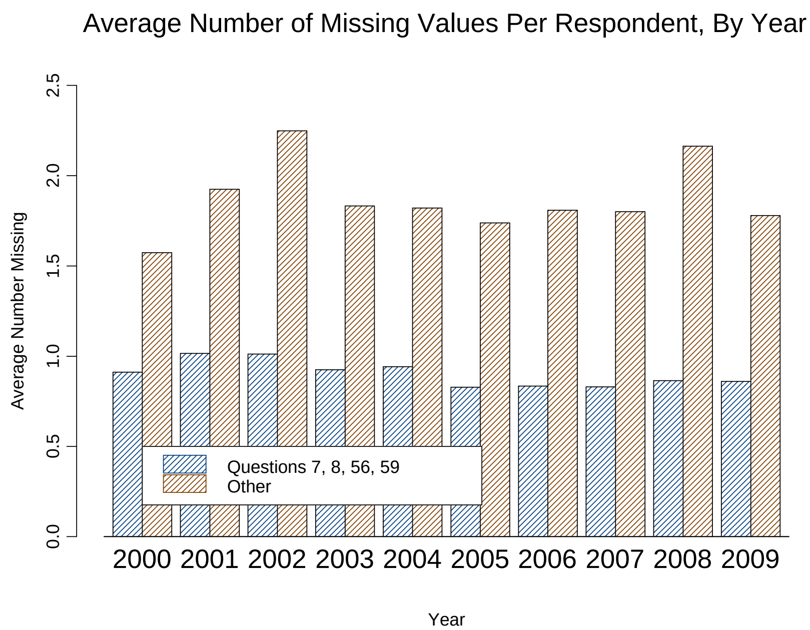

Figure 1. Missing Value Rate Per Year, For Four Items and For All Others

Negatively Worded Items. Five items from the CSAS (items 18, 23, 24, 30, and 34; see Adamshick, 2007) had a scale opposite to the other 56 items in that they were negatively worded (i.e., "strongly disagree" is indicative of a desirable response). As part of the data-screening process, the responses to these five items were reversed so that responses of "1" were recorded as "5" and so on; reverse scoring or negatively worded items is standard practice in survey research. However, despite the reversed scoring of these items, it is evident respondents were confused when it came to answering negatively worded items. Overall, about 4.5% of answers, across all items, were recorded as “1” or “2” (this computation includes all years and was done after missing values were replaced), but in the set of reversed items this proportion was 16.0%. Indeed, although these items accounted for only 8.8% of the total responses, they accounted for 33.9% of all “1” responses. (Recall that a “1” in this analysis corresponds to an original answer of “5” by the respondent.) The proportion of “1” and “2” responses can be seen in Figure 2. Red triangles show the reversed items. Note that for all five of the negatively worded items, more than 7.5% of respondents gave a “1” or a “2” response (horizontal line). Out of the other 52 items, only four of them (items 31, 32, 50, and 55) were rated “1” or “2” by more than 7.5% of respondents.

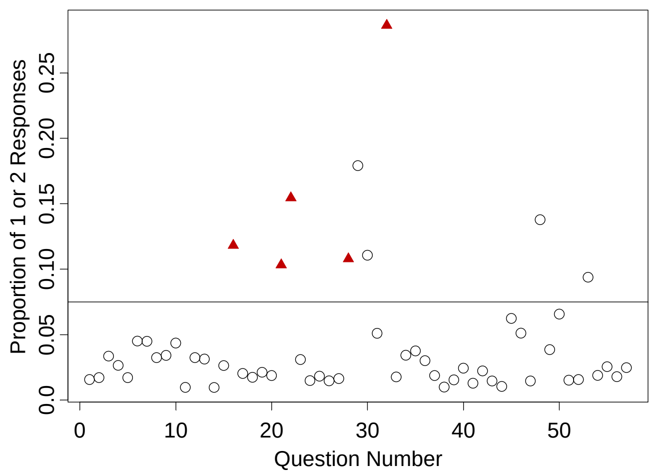

Figure 2. Proportion of “1” or “2” Answers, By Item

It is difficult to compare the large number of negative responses to items 31, 32, 50, and 55 with the negative responses to the negatively worked items; however, it is our belief that that the former probably reflects real dissatisfaction and the latter is most likely due, at least in part, to confusion on the part of the respondents. As a result, we omitted the negatively worded items from further analysis.

Adjusting for Modal Responses. A large number of respondents showed very little variability from one item to the next. Out of the 2000 through 2007 dataset including only the 84,732 surviving cases, and 52 surviving questionnaire items, 7.3% of respondents gave the very same answer to every item. An additional 24.5% of respondents used only two possible answers across the set of 52 items. Only 32% used four or all five of the options presented. Indeed, many respondents showed a pattern of answering almost all items with a single response option. For example, more than half of respondents gave a single response option at least 40 times. The modal response (the one most frequently given) was “4” in about 66% of cases, and “5” in another 29%.

On average, items were answered at the respondent’s mode 75% of the time, below the mode 15% of the time, and above the mode 10% of the time. The problem with the high number of “on-mode” responses is that it reduces the variability in the dataset and reduces the ability to conduct meaningful statistical analysis. Moreover, items that are answered at the respondents’ mode are less informative than items whose responses are different from the mode. We believe that the latter are more likely to indicate real satisfaction or dissatisfaction with the issue, while the former group represents a “nonopinion” on the issue.

In order to minimize the problem of limited variability, we adjusted each response to reflect its distance above or below the respondents’ mode. A score of “4” given by a respondent whose modal response is “5” was therefore coded as a “–1”, whereas a score of “4” given by a respondent whose modal response is “3” was coded as a “1.” Most adjusted responses were “0”s, because more than 75% of all responses were made by respondents assigning their own mode.

Figure 3 shows the proportion of respondents answering each item above and below their own modal response. Eight items stand out as eliciting responses that are unusually distant, on average, from the respondents’ modes. Items 31, 32, 50, and 55 (as discussed above), are in the bottom-right part of Figure 3. The items in the top-left part of Figure 3, each of which was frequently answered above respondents’ modes, are items 4 (*my command closely monitors proficiency and currency standards to ensure aircrew are qualified to fly*), 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*), and 19 (*my command has a reputation for high-quality performance*).

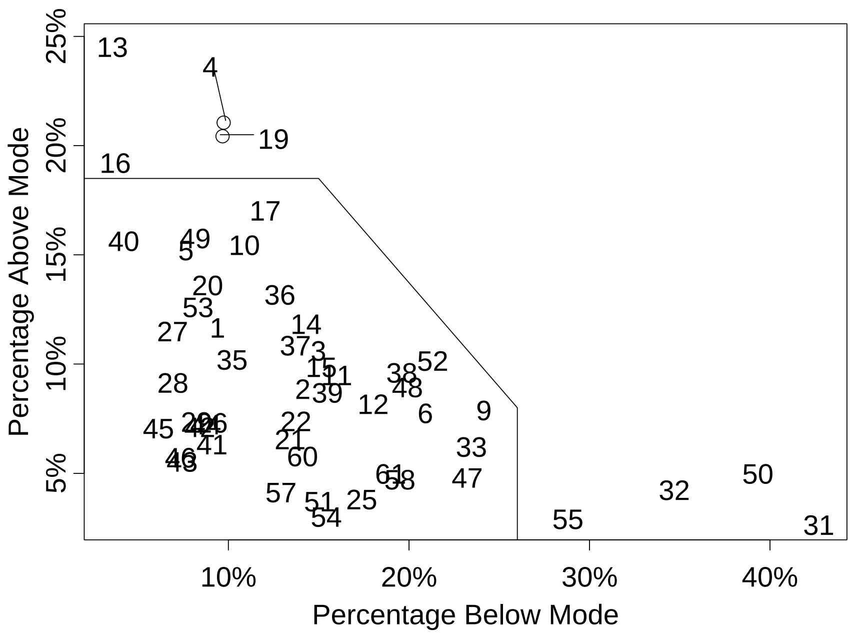

Figure 3. Rates of Responses Above and Below Respondents' Mode

Modal Stability. The idea of adjusting each response by reference to the respondents' mode makes sense if we feel that respondent-to-respondent differences in modal response are a major source of "noise" in the data. However, the adjustment has the potential to remove important "signal"-type information. This would be true if the sets of observed modes differ across communities or, from squadron-to-squadron within a community. In fact, different squadrons are associated with different observed modal values, suggesting that the modal response is a true reflection of satisfaction or dissatisfaction with issues at that squadron. However, this interpretation is complicated by the fact that respondents' modes are also correlated with their demographics, and different squadrons have different demographic compositions.

Figure 4 shows that responses are more positive among senior personnel than among juniors, more positive in training and TACAIR squadrons than in big wing and rotary wing squadrons, and more positive among pilots than among aircrew. Of course, these categories are associated: about 51% of respondents in big wing squadrons and 54% in rotary-wing squadrons were officers, compared to 99% for TACAIR and training (counting, for the purpose of computing these percentages, 192 warrant officers as officers and one civilian as enlisted). Furthermore, only 3% of aircrew, but 94% of Naval Flight Officers (NFOs; officers who specializes in airborne weapons and sensor systems, but do not actually fly the aircraft) and 99% of pilots were officers.

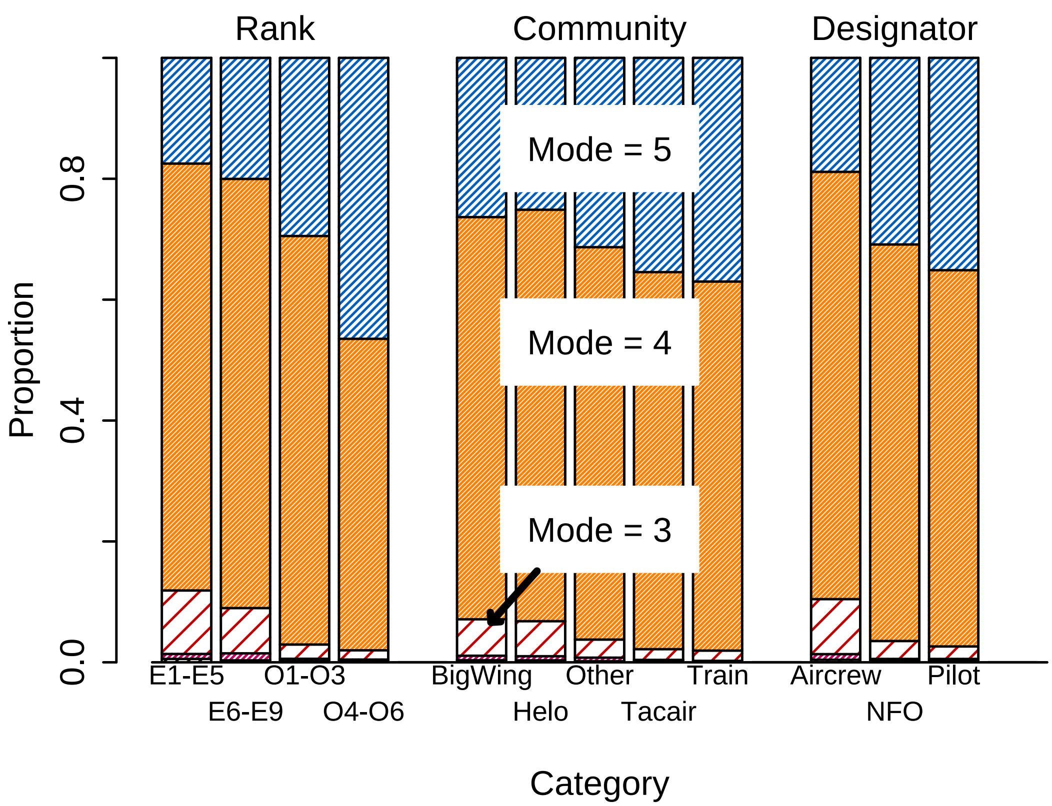

Figure 4. Proportion of Modal Response, By Demographic Category

The fairly similar makeup of Navy and Marine Corps respondents is worth noting here. In each case, aircrew make up about a third of respondents (and, not coincidentally, enlisted respondents make up a third of total respondents as well). The Navy has more NFOs (19% of respondents) than the Marine Corps (8% of respondents).

Identical Items. Items 5 and 43 are exactly the same (they both read "Command leadership is actively involved in the safety program and management of safety matters"). In the pre-2008 data, the correlation between the sets of responses for those two items in the adjusted data was found to be 0.69, high in comparison to many other pairwise correlations (it is in the 97th percentile of the set of 1,326 pairwise correlations), but still not the largest among the 1,326 pairs. In the mode-adjusted data, the correlation is much smaller (0.34), though still in the 96th percentile.

Analysis of the correlations between pairs of items produced another interesting result. In all but one pair of items, the correlation between the responses was higher in the 2008 and later data than in the data before 2008. This reflects the fact that more respondents gave more responses at their mode in later years, leading to more coincident pairs of responses.

2. Exploratory Factor Analysis (EFA)

We performed EFA on the adjusted dataset (i.e., the dataset with the nine items discarded and with each response having been adjusted for the respondents' mode). We elected to keep 12 factors because we felt that any more would render the analysis incomprehensible. Table 2 shows the set of items whose loadings on each factor exceeded a cutoff of 0.415; a cutoff we selected with interpretability in mind. The factors were named by two psychologists familiar with the safety climate literature. Other items have nonzero loadings on each of these factors. (This use of the cutoff led to the discarding of one factor with no such loadings; a second factor that overlapped with a previous one was also removed.)

Factor	Factor Name	Items						
1	Safety leadership	40	41	42	43	44	45	46
2	Safety monitoring	1	2	3	4	5		
3	Risk management	26	27	28	29			
4	Quality leadership	47	50	55				
5	Violations	14	15	17				
6	Safety department effectiveness	38	39	58				
7	Maintaining standards	19	20	21				
8	Safety training effectiveness	57	60	61				
9	Availability of resources	31	32	33				
10	Reporting culture	10	11					

Table 2. Factors From EFA and Their Associated Items

Using the factor loadings for each factor, we can then produce a set of 10 scores for each respondent. These are the scores we use to characterize respondents. The factor analysis was performed in S-Plus 7.0 (Insightful Corp., 2005); the built-in “factanal()” and related commands produce the factor loadings and the scores.

Analyzing the Factor Score. Although factor scores range from around –10 to around +11.5, the vast majority (97%) are between –2 and 2, 84% are between –1 and 1, and about half are between –0.33 and +0.33. Of course, these low factor scores are associated with the fact that most responses have adjusted values of zero. Table 3 shows the counts of respondents falling in each quintile of scores on factors 1 and 2. This table shows that the two factors appear to measure quite different things.

	Factor 2 →	**First**	**Second**	**Third**	**Fourth**	**Fifth**
Factor 1	First	4,592	3,083	1,007	3,378	4,887
↓	Second	1,271	2,730	8,086	2,313	2,546
	Third	2,364	3,804	5,310	3,263	2,205
	Fourth	4,163	4,036	1,539	4,775	2,433
	Fifth	4,557	3,293	1,000	3,221	4,876

Table 3. Counts of Respondents, By Factor Score Quintiles

Scores tend to differ a little between groups by service, rank and designation, and additionally in many cases an interaction effect can be seen. We attribute some of these differences to the very high power brought about by our relatively large sample sizes. An example is given in the left-hand part of Table 4. These columns display proportions of respondents by service (Navy and Marine Corps only) and by quintile of score on factor 2. Although the sets of numbers are quite similar, the hypothesis of homogeneity of the Marine Corps and Navy populations from which our samples are assumed to have been drawn is rejected (χ^2 = 59.0 on 4 d.f., $p = 0$). Other differences, though, are presumably important. The four left-hand columns of Table 4 show quintiles of factor one broken down by rank. Here the pattern is clear: enlisted personnel, particular the junior ones, are much more inclined to have low factor scores than the officers, particularly senior ones. In fact, the effect of rank seems to be the biggest for every factor.

	Factor 1				Factor 2	
	E1-E5	**E6-E9**	**O1-O3**	**O4-O6**	**USMC**	**USN**
First	23.3	21.6	19.1	17.8	20.4	19.9
Second	25.9	19.5	19.0	16.6	21.0	19.5
Third	17.6	19.2	20.7	21.4	20.3	19.9
Fourth	16.9	20.6	20.6	21.4	19.6	20.2
Fifth	16.4	19.1	20.6	22.9	18.7	20.6

Table 4. Proportions of Counts in Quintiles of Factor 1, By Rank, and Factor 2, By Service

Evaluating the Fit. There are techniques by which the "fit" of a factor analysis can be measured. However, these approaches generally rely on assumptions about the distributions of responses that cannot be justified here. Our belief is that, if our factors were largely the result of the model fitting the noise rather than real signal, the factors would fit substantially less well on a separate dataset (i.e., data that was not used in the model-building process).

3. Confirmatory Factor Analysis (CFA)

As described above, the 2008 dataset (n= 20,101) was not included in the EFA process. The 2008 dataset was used to test whether the 36-item, 10-factor model identified through the EFA resulted in an acceptable fit to another dataset. The fit of the EFA model to the 2008 data did not prove to be acceptable (χ^2 = 285903, df = 630, p>0.05; Comparative Fit Index (robust; CFI)= 0.71; Goodness of Fit (GFI) = 0.75; and root mean square error of approximation (robust; RMSEA)= 0.07). As part of the CFA process, researchers often carry out post-hoc fitting. However, even with substantial post-hoc fitting, arguably beyond what was based in theory, an acceptable fit of the model to the data could not be achieved (χ^2 = 285903, df = 630, p>0.05; CFI(robust) = 0.87; GFI = 0.92; and RMSEA(robust)= = 0.05). Although the adapted model did have a fit that was better than that obtained with the original model, it was still below that which is generally accepted in the literature (see Byrne, 2006, for a discussion).

Given these findings, a CFA process was used to assess the fit of the original and adapted models for the 2000 and 2007 datasets. The fit between the models and the data was found to be unacceptable for both the original (χ^2 = 885993, df = 630, p>0.05; CFI (robust) = 0.50; GFI = 0.64; and RMSEA (robust) = 0.09), and adapted (χ^2 = 885993, df = 630, p>0.05; CFI (robust) = 0.72; GFI = 0.86; and RMSEA (robust) = 0.07) models.

4. Post EFA and CFA Item Analysis

Given the failure to establish a stable factor structure, additional analysis was carried out to identify the reasons for the lack of stability. We believe that the lack of stability is due to changes in response characteristics over time, the high levels of intercorrelation among all of the CSAS items, and differences in demographics between the 2000 to 2007 data set and the 2008 data. Evidence supporting these conclusions is provided below.

Changes Across Time. From its inception until October 2004, the CSAS survey was voluntary. After that time, response was mandatory. It is reasonable to suspect that respondents being compelled to answer might do so in a manner that differs from those volunteering. For every item, a larger proportion of post-October 2004 respondents replied at their mode than was the case among the pre-October 2004 respondents. This was also true separately at each rank (except among senior officers ranked O4 and above). In fact, for the senior officers 49 of the 52 items showed an increase in the number of participants answering at their own mode.

In fact, there are a number of changes that appear to take place as years go by, not just as the boundary between voluntary and mandatory reporting is crossed. Two changes in the characteristics of the responses across time were apparent. First, there was an increased frequency of respondents whose modes were 5 as time went by, and also of respondents whose modes were 3 or less. This pattern can be seen, to at least some extent, in individual groups as well, so it is not merely a consequence of changing demographics within the population being surveyed. Figure 5 shows the distribution of modal responses other than 4; the sum of these went from a low of 29.5% in 2002 to a high of 39.8% in 2007.

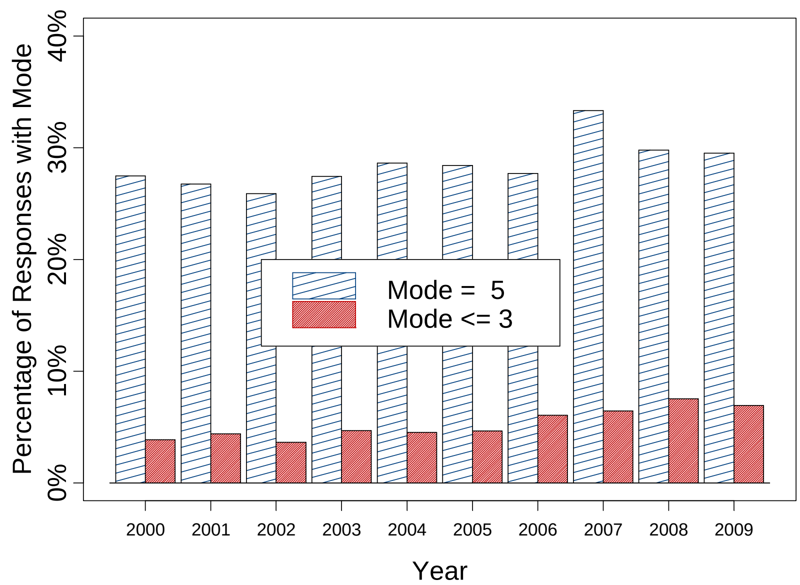

Figure 5. Proportion of Respondents with Different Modes, By Year

Second, and perhaps more importantly, there was an increase in the frequency with which respondents answer at their own mode. The left panel of Figure 6 shows the average number of different-from-mode responses provided by respondents, by year. The downward trend indicates that more and more respondents are using only a few of the possible response options. The right panel shows the average number of items answered at the respondents' mode, by year. In 2009, the average respondent gave their modal response to more than 40 of the 51 items, up from around 36 in 2000.

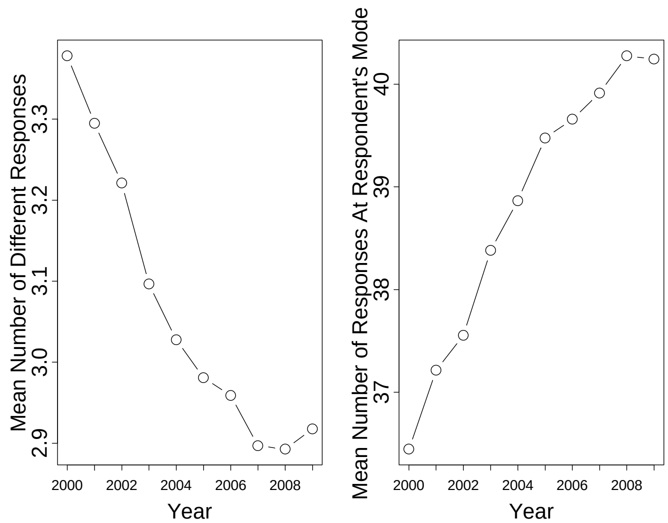

Figure 6. Decrease in Number of Different Responses (left) and Increase in the Number of Responses at the Mode (right), By Year

These two shifts have opposing effects on variance—the first acting to increase it and the second to decrease it. The overall effect is nonconstant variance. The effect of nonconstant variance is to render invalid the standard statistical tests, such as factor analysis, that rely on constant variance.

Factor Reliability. Cronbach's alpha is often used a measure of reliability in surveys. An alpha statistic is computed from the set of response items that are to be combined into a single factor. Alpha coefficient ranges in value from 0 to 1; the higher the calculated Alpha score, the more reliable the scale. Nunnally (1978) indicated 0.7 to be an acceptable reliability coefficient. Alpha should be large when the set of items elicits similar responses, so that the variance of the sum of scores for these items is small, compared to the sum of individual variances for these items. However, an unusual effect is seen in our data. First, many of the factors produced high alpha values. In fact, 9 out of the 10 factors had alphas greater than 0.7 for each year using the nonmodally adjusted dataset. Adjusting for modes removed some of the correlation between items and therefore produces fewer alpha in excess of 0.7, although most factors still reach this level in most years using the adjusted data. However, the assumption that this means the factors are reliable can be proven false using simulation.

We selected a random set of 15,000 cases and a random set of seven items and computed the alpha for those responses. Because the items chosen at random would not be expected to be particularly coherent, we expected smaller values of alpha than were observed for the "real" factors. However, the alpha from this simulation were uniformly greater than 0.8. The "coherence" among the items seems to be largely due to the fact that almost every respondent gives almost the same answer to every item.

Demographic Differences between the 2000 to 2007 Data and the 2008 Data. In addition to changes in variance, the demographic makeup of the 2008 sample is somewhat different from that of the respondents in the earlier data. There is a lower proportion of officers in the later group: the ratio of respondents in the earlier group to respondents in the later is 3.04 for E1-E5s (junior enlisted), 3.18 for E6-E9s (senior enlisted), 3.47 for junior officers (O1-O3), and 3.60 for senior ones (O4 and above). Of course, because most enlisted personnel have the designator "aircrew" and vice versa, a similar change in designator can be seen.

D. STUDY 1: DISCUSSION

The exploratory factor analysis did not replicate the five MOSA factors. However, more importantly, we were also unable to establish a factor structure that was stable across time using exploratory and confirmatory factor analysis techniques. This finding calls into question the construct validity of the questionnaire. It could be argued that the failure to establish the construct validity of the CSAS is due to flaws in the questionnaire construction. Given the findings from this study, it is not possible to rule this out. Nevertheless, the CSAS items are not dissimilar from those that are typically included in other safety climate questionnaires (see O'Dea et al., 2010). It could also be argued that the lack of our ability to find a stable factor structure can be more correctly attributed to changes in the response characteristics over time.

Krosnick (1999) differentiates between two different strategies participants use for responding to questionnaire items: optimizing and satisficing. When optimizing, the respondent must interpret the item and deduce its intent. Next, they must conduct retrieval by searching their memory for the relevant information, and integrating it into a single judgment. Finally, they complete the judgment stage by translating the judgment into a response by selecting the appropriate alternative from those offered. If the respondent is attempting to generate the optimal answer, then this process requires a large cognitive load. Conversely, when using a satisficing strategy, respondents compromise their standards and use a strategy that does not require as much cognitive effort as optimizing, and a quicker and less thoughtful response is given. Krosnick (1999) states that if satisficing is done subtly, respondents still interpret the item, retrieve the information, make a judgment, and make a response selection. However, less effort is put into these steps (weak satisficing). More extreme satisficing results in skipping the retrieval and judgment stages entirely. The respondent interprets each item superficially and selects what they believe to be a reasonable answer. In even more extreme cases, the respondent omits reading the item altogether and provides an arbitrary response (perhaps even the same response) to each item. So, what is the evidence that large numbers of CSAS respondents are using a satisficing strategy, and that the frequency with which this strategy has been used has increased over time?

Misinterpretation of the Negatively Worded Items. The five negatively worded items (after being reversed to align with the other items) had response patterns that indicate that respondents are more likely to misinterpret the item, as compared to the other questionnaire items. As described in the results, overall about 4.5% of answers across all items were recorded as “1” or “2.” However, in the negatively worded items (after reversal) this proportion was 16.0%. If a satisficing strategy was being used, this finding would be expected as the participants are simply responding in the same way they did to the other 56 items, and do not notice the negative wording of the five items in the CSAS.

Increase in the Number of Modal Responses Over Time. A large proportion of responses were given at each respondent’s mode (75% of responses). Moreover, the proportion of responses at the mode has increased over time. This suggests that the respondents are increasingly using a satisficing strategy and not expending the cognitive effort necessary to give a considered response.

Decrease in Deviation from Modal Responses Over Time. Related to the increased frequency with which respondents are giving their modal response, they also appear to be decreasing in the frequency with which respondents are prepared to deviate from their mode over time. Again, this result is suggestive of a decrease in cognitive effort being put into questionnaire completion.

In sum, there are multiple sources of evidence indicating a prevalence of satisficing response strategy among CSAS respondents. There are a number of possible explanations for this increase.

- As described in the introduction, since 2004, Naval aircrew are mandated to complete the questionnaire. It is suggested that "CSAS malaise" may have set in because participation is mandatory and the questionnaire is completed a minimum of once every two years. Therefore, instead of those who do not wish to participate simply opting out, it may be that a large proportion of aircrew are using a strategy of questionnaire completion that requires as little cognitive effort as possible, but still allows them to "do their duty."
- The CSAS is also part of an increase in the number of behavioral based safety programs that have been introduced in naval aviation. O'Connor and O'Dea (2008) identified 15 individual elements of the naval aviation safety program that focus on addressing the human causes of mishaps. Therefore, it is possible that there may be a larger overall climate of 'safety fatigue' within naval aviation.
- The CO is not required to share the CSAS responses with squadron members. There is anecdotal evidence to suggest that some COs do share the CSAS findings with squadron personnel. It is suggested if the CSAS findings were shared more often with squadron personnel, this may decrease the number of aircrew utilizing a satisficing strategy. To illustrate, Ward (1994) found that General Practitioners are less likely to participate in future surveys if they are given insufficient feedback.

E. STUDY 1: CONCLUSION

"Recent research has shown that surveys with very low response rates can be more accurate than surveys with much higher response rate" (Krosnick, 1999, p. 540).

The representativeness is more important than the sample size. Therefore, to establish whether the CSAS has construct validity, there is a need to separate the respondents who are using a satisficing strategy, and discard these individuals from further analysis. Malhorta (2008) states that extremely quick completion time may be a valuable criterion in identifying individuals who are utilizing a satisficing strategy. He does indicate that completion time, in and of itself, may not be an optimal filtering criterion because it is only a proxy for satisficing. Nevertheless, if the questionnaire is completed very rapidly, we can be very sure that the respondent is not giving the items too much thought.

Fortunately, since 2006, data has been collected on the time taken by respondents to complete the CSAS. In the next study, the response time will be used as a criterion for discarding the respondents we suspect of using a satisficing questionnaire completion strategy. We will retain only a subset of items for which there was a large degree of variance. We will then attempt to establish a factor structure with acceptable construct validity once again.

THIS PAGE INTENTIONALLY LEFT BLANK

III. STUDY 2

A. STUDY 2: INTRODUCTION

In this section, we limit the data analysis to a subset of respondents and questionnaire items from the original dataset. The original questionnaire posed 61 items, plus 10 items used for demographics, and 2 open-ended (essay-style) questions. The criteria used to select the respondents to be included in the remaining analysis is the time taken to complete the questionnaire, the criteria to select the questionnaire items to remain in the analysis is the percentage of different-from-mode responses.

Based on a detailed analysis of the time taken to complete the questionnaire, and the proportion of modal responses, we selected 10 minutes as the smallest reasonable time in which the survey could be filled out.

Table 5 shows the distribution of the proportion of items each respondent answered at his or her mode, broken down by whether the elapsed time was less than 10 minutes, missing, or greater than 10 minutes. (The rows break the numbers of items answered at the mode into groups.) The first and third columns show the substantial differences in response patterns between those who completed the survey quickly and those who did not; the middle column (including the untimed data prior to 2006) appears, not surprisingly, to be a blend of the two types of responders.

	Time to Complete Survey		
# At Mode	**≤ 10 min**	**Not Timed**	**>10 min**
13-29	12.1	17.7	20.3
30-39	24.2	32.6	33.3
40-44	15.1	17.7	18.2
45+	48.6	31.9	28.3
Sample Size	24,154	62,418	23,442

Table 5. Proportion of Responses by Number at Mode at Time to Completion

Table 5 shows that when responders who finished in less than 10 minutes, or for whom no time was recorded, were excluded, the resulting data set contains 23,442 observations. Even in this group, though, a substantial proportion of respondents give large numbers of responses at their own mode, particularly for certain items. Therefore, we selected the subset of items that were most frequently answered away from the mode. These items included 4, 13, 16, and 19, which were answered positively more frequently than any others; items 31, 32, 50, and 55, which were the items most frequently answered negatively; and items 9, 17, 48, and 52, which had large differences between the number of positive and the number of negative respondents (see Figure 3). The rationale is that the inclusion of items for which there is little variation is not useful in identifying differences between groups of respondents. After this action, our data set included 23,442 cases and 12 questionnaire items.

B. STUDY 2: METHOD

As in study one, we divided the data into two parts. We performed an exploratory factor analysis on the 47% of the data (10,968 cases) collected prior to January 1, 2008. The factor structure from the EFA was then used to carry out a CFA with the remaining 12,476 cases from 2 January 2008 through July 2009. Once a stable factor structure was identified, comparisons were made of the factors scores on the basis of rank, type of aircraft flown (big wing, TACAIR, rotary), and branch of service (Navy versus Marine Corps). The factor scores were calculated using the "GLS" function that is calculated by EQS for windows. GLS factor scores are described in Bentler and Yuan (1997).

C. STUDY 2: RESULTS

Exploratory Factor Analysis. Because only 12 items remained under consideration, we sought two factors. (Because the numbers of cases and items has been reduced, we refer to this analysis as the "reduced" EFA.) Items loaded onto factors in the manner depicted in Table 6. Factor 1 captures the items frequently answered positively; those answered negatively load onto factor 2; items that were "controversial" were split. The usual χ^2 test for model adequacy suggests that two factors are insufficient; it proposes using seven factors, which is the maximum possible. However, we chose to keep a two-factor solution for purposes of interpretability.

Factor	Factor Name	Items						
1	Personnel leadership	4	13	16	19	9	17	48
2	Integration of safety & operations	31	32	50	55	52		

Table 6. Items Loading Onto Factors of Reduced EFA

Confirmatory Factor Analysis. The two factors identified from the EFA were entered into a CFA. The initial fit of the data was not acceptable (χ^2 = 24540, df = 66, p>0.05; CFI (robust) = 0.73; GFI = 0.93; and RMSEA (robust) = 0.08). However, by allowing the two factors to correlate, and allowing the error terms between items 31 and 32, and items 13 and 16 (as recommended by the Wald test), an acceptable fit resulted (χ^2 = 24540, df = 66, p>0.05; CFI (robust) = 0.91; GFI = 0.97; and RMSEA(robust) = 0.047). The standardized solution is shown in Figure 7. The two-factor model was also found to be an acceptable fit for the 23,442 cases analyzed. (χ^2 = 45647, df = 66, p>0.05; CFI (robust) = 0.91; GFI = 0.97; and RMSEA(robust) = 0.047).

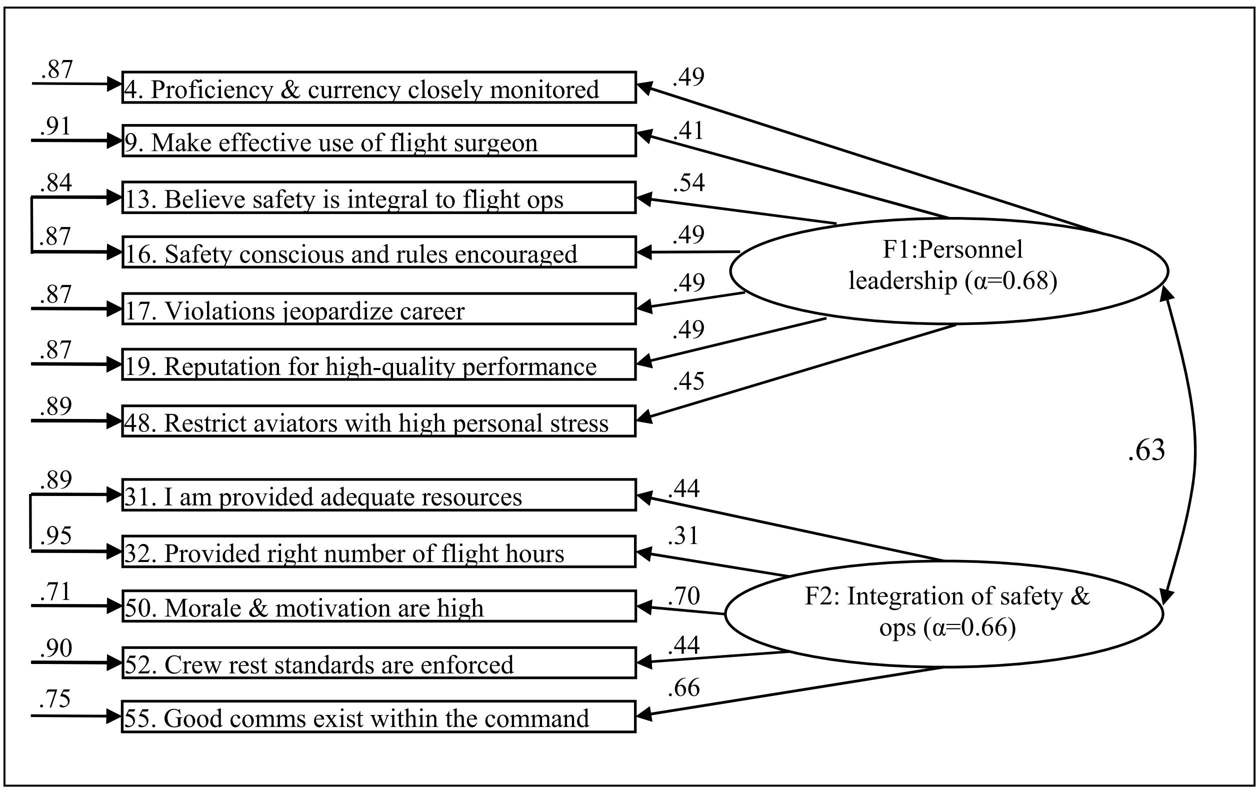

Figure 7. CFA Standardized Solution

Between Groups Comparison. The factors scores were used to examine whether there were differences on the basis of rank, type of aircraft flown, or branch of service. We used Analysis of Variances ANOVAs, acknowledging that the assumptions for this test are met only approximately. Figure 8 shows the average factor scores for factor one (left) and factor two (right), by aircraft flown. Vertical lines extend ± 2 standard deviations above and below the means. Although some deviation in the means can be detected, and an ANOVA suggests that the means are statistically significantly different for factor one (ANOVA *F*-test *p*-value ≈ 0), the spread of the responses is large compared to the range of the means, and the two linear models both have R^2 values < .01.

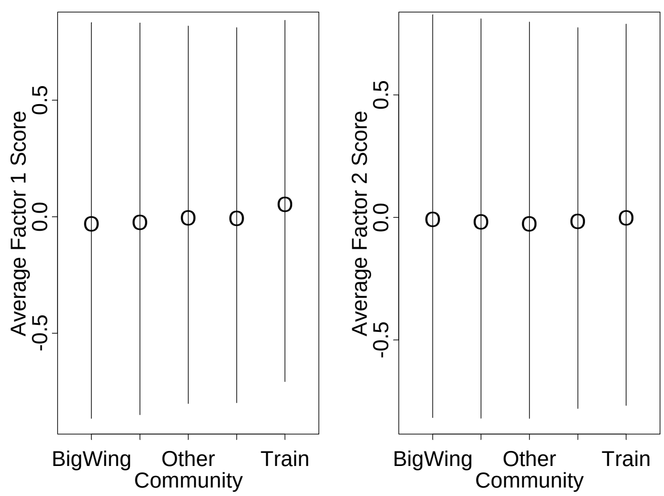

Figure 8. Factor Scores for Factors 1 (left) and 2 (right), By Community

D. STUDY 2: DISCUSSION

It was possible to establish a stable, two-factor structure for the 12 CSAS items that had reasonable levels of variability. The two factors identified (personnel leadership and integration of safety and operations) is consistent with the safety climate literature (see O'Dea et al., 2010). From examining the items that make up the two factors, it can be seen that their focus is on those individuals in leadership positions. A factor concerned with management is identified about 75% of the time in safety climate research (see O'Dea et al., 2010). However, the identification of the two factors was at the expense of an enormous amount of data. Of all the data collected, only 4.2% of the original data set was retained in the second study. This may seem very wasteful; however, we can justify the decision making. Firstly, we felt it was necessary to identify those respondents utilizing an optimizing technique to complete the questionnaire. The metric we used for this was completion time. As time was only available from 2006 onwards, we could not use the complete data set. Secondly, we discarded the majority of items for which there was low levels of variance. DeVellis (1991) states that the discarding of items is a normal part of questionnaire development, and it is not unusual to begin with a pool of items that is three or four times as large as the final scale. The authors were interested in discarding those items for which there were little variance, as this was necessary for next stage of this research effort.

In future work, we will use the factor scores from the two factors identified in study two to assess whether the responses from individuals from squadrons in which mishaps took place differ from the responses from those individuals in squadrons with no mishaps. If the factor scores really reflect aspects of safety climate that contribute to mishaps, then it may be possible to detect (and alert) squadrons at higher risk of mishap by examining their survey responses.

Although the usefulness of items with little variance is low for the purposes of establishing the predictive validity of the questionnaire, it may be that it is useful information for the CO to know that, for example, the vast majority of the squadron personnel agree that all unit members are responsible and accountable for safe flight operations. Nevertheless, as discussed below, it seems that quality is being compromised at the expense of quantity.

E. STUDY 2: CONCLUSION

After identifying those respondents that we were confident were using an optimizing CSAS completion strategy, and discarding those items for which there was little variance, we were able to identify a stable, two-factor structure, both of which have safety leadership at their core. These two factors, and the items they contain, will be used to establish the predictive validity of the CSAS in the next phase of this research effort.

F. GENERAL CONCLUSIONS AND RECOMMENDATIONS

In essence, a safety climate questionnaire provides senior leadership with a "snapshot" of the safety climate at a particular moment in time. The goal, however, is to provide information that may be useful in identifying in advance issues that may increase the likelihood of an accident occurring, and thus, allowing leadership the opportunity to rectify those situations before an accident occurs. Our analysis of the CSAS, as discussed in this report has identified serious concerns about the usefulness of the CSAS data in fulfilling this goal. The evidence supporting these concerns with the validity of the data is summarized below.

- Despite the reversal of the answers to the negatively worded items, there is evidence that respondents were confused when it came to answering them. Overall, about 4.5% of answers, across all items, were recorded as "1" or "2" (this computation includes all years and was done after missing values were replaced), but in the set of reversed items this proportion was 16.0%.
- Correlation between items 5 and 43. Items 5 and 43 are exactly the same. Seventy-five percent of respondents gave the same answer to the two identical items. It is in the 97th percentile of the set of 1,326 pairwise correlations; nevertheless, this level of agreement is not as high as might be expected.
- Large proportion of "satisfied" responses. The modal response (the one most frequently given) was "agree" in about 66% of cases, and "strongly agree" in another 29%. Therefore, there is a tendency for aircrew to respond positively to the majority of the items.
- Large proportion of respondents answering at their mode, with an increase in the frequency of this behavior over time. On average, items were answered at the respondents' mode 75% of the time, below the mode 15% of the time, and above the mode 10% of the time. Further, over time, there was an increased frequency of respondents whose modes were 5 (strongly agree, indicative of a positive view of safety climate), and also of

respondents whose modes were 3 or less (indicative of a neutral to negative view of the safety climate. There has also been an increase over time in the frequency with which respondents answer at their own mode.

- The quicker respondents complete the questionnaire, the more likely they are to answer at their mode. For example, among those who completed the questionnaire in 10 minutes or less, 48.6% responded at their mode for at least 45 of the 61 items, compared to 28.3% of respondents who took more than 10 minutes.

It is difficult to draw conclusions about the construct validity of the CSAS when there are clear data validity issues. Of great concern is that genuine safety issues are not being identified, as the thoughtful and considered responses are being "washed out" by those adopting a satisficing strategy. We are certainly not suggested that the Navy abandon the periodic assessment of safety climate. However, it is recommended that steps should be taken to increase the proportion of respondents who are willing to provide considered responses, and screen out those that are likely using a satisficing strategy. We propose a number of recommendations for improving the quality of the safety climate data being collected by the CSAS.

1. Recommendation 1. Develop a short, or adaptive, version of the CSAS

As would be expected, length of questionnaire generally has a negative effect on response rate (e.g., Bogen, 1996; Dillman, Sinclair, & Clark, 1993; Sheehan, 2001; Smith, Olah, Hansen, & Cumbo, 2003). To illustrate, Smith et al. (2003) found nearly a doubling of response rate when they compared a one-page survey with a three-page version of the same survey. Asiu, Antons, and Fultz (1998) asked U.S. Air Force Academy cadets what the thought should be the ideal length of a survey. On average, the students stated that the ideal length should be 22 items that take 13 minutes or less to complete. Therefore, although the CSAS length is typical of other safety climate questionnaires (O'Dea et al., 2010) it is suggested that to reduce the proportion of satisficing responses, the length of the questionnaire should be drastically reduced and a short version of the CSAS developed. The development of short-form versions of standard questionnaires is common in the health and psychology research. Our recommendation would be to develop a 20-item version of the CSAS. It is suggested that the 12 items analyzed in the second study should be given strong consideration for inclusion, as well as one or two other items from each of the five MOSE areas.

An alternative, and arguably more preferable, method to developing a short-form version, is to develop a questionnaire that adapts the items that are asked based on responses to previously asked items. The respondents do not respond to every item, as is the case with the current version of the CSAS, but proceeds through the questionnaire by skipping sections according to responses given to previous items. To illustrate, if the responses suggest that a respondent feels strongly that if they raise a safety issue with senior leadership then it will be acted on, then it is unnecessary to ask them 10 items along this theme. Given that the questionnaire is Web-based, the use of this type of methodology would be seamless to the respondent.

The Navy/Marine Corps School of Aviation Safety in Pensacola, Florida is well placed to aid in the development of a short version of the CSAS. The school has access to students at the aviation safety officer and aviation safety commander course, which represent a cross section of aircrews. Involving these personnel would allow information to be obtained on exactly what they want to know about safety climate, and how they are using the information collected. Anecdotal evidence suggests that COs are very interested in the open-ended comments, but spend little time examining the data from the items analyzed as part of this paper. It is suggested that involving the user population will be helpful in identifying the critical issues.

2. Recommendation 2. The development of a rigorous data screening tool

A rigorous methodology should be applied to screen out those suspected of using a satisficing strategy. It is suggested that the data-screening methods used in this report should be used to screen the collected day. Time-to-complete would seem to be a particularly important metric.

3. Recommendation 3. Consideration should be given as to whether CSAS should be mandatory, or highly encouraged

Consideration should be given to allowing individuals to anonymously decline to take the questionnaire. As demonstrated, forcing people to complete the questionnaire has had a detrimental effect on the quality of the data collected.

4. Recommendation 4. More closely align the CSAS program with safety culture workshops

The United States Navy has another mechanism for providing COs with information on safety culture, called safety culture workshops. Its purpose is to identify potential hazards that might interfere with mission accomplishment. They also identify command strengths. A safety culture workshop is facilitated by specially trained senior Naval aircrew. The facilitators spend time looking around the squadron, watching people working, and having informal conversations with a cross section of squadron personnel. Following the informal phase of the workshop, the facilitators carry out focus group discussions with squadron personnel. The information gleaned from the workshop is then summarized and given back to the squadron's CO. The CO should use this information to focus on areas that require better risk assessment and risk controls.

It is suggested that, as is currently the case, a safety survey be carried out six months after a new CO has taken command. However, it is recommended that a safety culture workshop should be carried out shortly after the survey is administered. The CSAS data should be shared with the safety workshop team, in order that the workshops can be tailored to the specific needs of the squadron. Combining these two safety programs will help the squadron personnel see the benefit of the CSAS, and improve the effectiveness of the safety culture workshop.

5. Conclusion

It is important to indicate that we are in no way saying that a number of naval aircrew are deliberately providing misleading information in their CSAS responses. Rather, as is the case in the majority of organizations, naval aircrew have a high workload. To illustrate, O'Connor, Cowan and Alton (2010) reported a review of the safety concerns of 68 squadrons from two U.S. Naval aviation communities (helicopter and fighter/attack). It was found that workload and operational tempo were identified as a major safety concern for 85% of the squadrons. Therefore, it unsurprising that naval aircrew may not give the CSAS the time and attention required to provide an accurate view of safety climate at their squadron. This sentiment is likely to be particularly true if the participants feel there are no major safety issues in the squadron, or they have not seen changes that resulted from the survey in the past.

The analysis in this paper has demonstrated the truth in Krosnick’s (1999) assertion that representativeness does not increase monotonically with response rate. Given that most military personnel follow orders, it can be assumed that the CSAS database discussed in this paper represents close to a 100% response rate. However, as shown, this does not mean that all of the data is accurate and useful. It is argued that more ‘truthful’ data could have been obtained from a shorter questionnaire completed by a smaller number of motivated individuals.

LIST OF REFERENCES

Adamshick, M. H. (2007). Leadership and safety climate in high-risk military organizations. Ph.D. dissertation. University of Maryland, College Park, MD. Available from: www.lib.umd.edu/drum/bitstream/1903/6808/1/umi-umd-4294.pdf.

Asiu, B. W., Antons, C. M., & Fultz, M. L. (1998). Undergraduate perceptions of survey participation: Improving response rates and validity. Paper presented at the annual meeting of the Association of Institutional Research. Minneapolis, MN.

Bentler, P. M., & Yuan, K.-H. (1997). Optimal conditionally unbiased equivariant factor score estimators. In M. Berkane (Ed.), *Latent variable modeling with applications to causality* (pp. 259-281). New York: Springer-Verlag.

Bogen, K. (1996). The effect of questionnaire length on response rates: A review of the literature. *Proceedings of the Section on Survey Research Methods*. Alexandria, VA. 1020-1025.

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.

DeVellis, R. (1991). *Scale development: Theory and applications*. London: SAGE Publications.

Dillman, D. A., Sinclair, M. D., & Clark, J. R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant addressed census mail surveys. *The Public Opinion Quarterly*, *57*(3), 289-304.

Harmon, H. H. (1976). *Modern factor analysis*. Chicago: University of Chicago Press.

Insightful Corp. (2005). *S-Plus 7 for windows users' guide*. Seattle: Insightful Corp.

Krosnick. J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*, 213-236.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*, 537-567.

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

Malhorta, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, *72*(5), 914-934.

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

Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.

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. (2007). The U.S. Navy's aviation safety program: A critical review. *International Journal of Applied Aviation Studies*, *7*(2), 312-328.

O'Dea, A., O'Connor, P., Kennedy, Q., & Buttrey, S. (2010, March). A review of the safety climate literature as it relates to Naval aviation. NPS-OR-10-002. Monterey, CA: Naval Postgraduate School.

Sheehan, K. B. (2001). E-mail survey response rates: A review. *Journal of Computer-Mediated Communication*, *6*(2). Available at http://www.ascusc.org/jcmc/vol6 /issue2/sheehan.html.

Smith, R., Olah, D., Hansen, B., & Cumbo, D. (2003). The effect of questionnaire length on participant response rate: A case study in the U.S. cabinet industry. *Forest Products Journal*, *53*(11), 33-36.

Ward, J. (1994). General practitioners' experience of research. *Family Practice*, *11*, 418-23.

Zortman, J. M. VADM. (2004). CNAF commanders training symposium safety wrap-up. Unclassified General Administrative Naval Message: R 240054Z NOV 04.

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. Major General Thomas Travis...1
   59th Medical Wing, Lackland AFB
   San Antonio, TX 78236

6. Colonel Lex Brown..1
   711th Human Performance Wing
   Brooks-City Base, TX 78235

7. Professor Nita Lewis Shattuck...1
   Department of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5219