



Calhoun: The NPS Institutional Archive
DSpace Repository

Faculty and Researchers

Faculty and Researchers' Publications

2011

Identifying and addressing the limitations of safety climate surveys

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

Journal of Safety Research

OConnor, P., Buttrey, S, ODea, A., & Kennedy, Q. (2011). Identifying and addressing the limitations of safety climate surveys. *Journal of Safety Research*, 42, 259-265.
<http://hdl.handle.net/10945/48012>

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun



Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

<http://www.nps.edu/library>

**Cite as: O'Connor, P., Buttrey, S, O'Dea, A., & Kennedy, Q. (2011).
Identifying and addressing the limitations of safety climate surveys.
Journal of Safety Research, 42, 259-265.**

Identifying and Addressing the Limitations of Safety Climate Surveys

Paul O'Connor¹, Samuel E. Buttrey², Angela O'Dea³, & Quinn Kennedy⁴

1. **Corresponding Author:** Centre for Innovation and Structural Change, J.E. Cairnes School of Business & Economics. National University of Ireland, Galway. Ireland. Email: poc73@hotmail.com. Tel: +353 (0) 91 492897 Fax: +353 (0) 91 495524.
2. Operations Research Department, Naval Postgraduate School, 1411 Cunningham Rd, Monterey, California 93943, USA. Email: buttrey@nps.edu. Tel: 831 656 2618.
3. 21 Gleann na Coille, Barna Road, Galway, Ireland. Email: aodea69@hotmail.com, Tel: +353 (0) 91 449687.
4. Operations Research Department, Naval Postgraduate School, 1411 Cunningham Rd, Monterey, California 93943, USA. Email: mqkenned@nps.edu. Tel: 831 656 3035.

Identifying and Addressing the Limitations of Safety Climate Surveys

Abstract

There are a variety of qualitative and quantitative tools for measuring safety climate. However, questionnaires are by far the most commonly used methodology. This paper reports the descriptive analysis of a large sample of safety climate survey data ($n=110,014$) collected over ten years from U.S. Naval aircrew using the Command Safety Assessment Survey (CSAS). The analysis demonstrated that there was substantial non-random response bias associated with the data (the reverse worded items had a unique pattern of responses, there was a increasing tendency over time to only provide 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 is suggested that the non-random responses bias was due to the negative effect on participant motivation of a number of factors (questionnaire design, lack of a belief in the importance of the response, participant fatigue, and questionnaire administration). Researchers must consider the factors that increase the likelihood of non-random measurement error in safety climate survey data and cease to rely on data that are solely collected using a long and complex questionnaire.

KEY WORDS: Safety climate, safety culture, survey, questionnaire, measurement error.

1. Introduction

Safety climate describes employees' perceptions, attitudes, and beliefs about risk and safety (Mearns & Flin, 1999). Over 40 different safety climate measures have been developed (Yule, O'Connor, & Flin, 2003), and have been used in a wide range of high-risk industries (see Flin, Mearns, O'Connor, & Bryden, 2000 for a discussion). Further, in much of the published research on safety climate, a questionnaire is the single method used to evaluate safety climate (e.g., Desai, Roberts, & Ciavarelli, 2006; Gibbons, von Thaden, & Wiegmann, 2006; Singer et al., 2003).

There has been an ongoing debate within the safety literature regarding the use of the terms 'culture' and 'climate,' and whether they represent the same or different concepts. The general consensus is that culture represents the more stable and enduring characteristics of the organization, and has been likened to its dimensions or 'personality.' Safety culture is a complex and enduring trait, reflecting fundamental values, norms, assumptions, and expectations, which, to some extent, reside in societal culture (Mearns & Flin, 1999). Safety climate, on the other hand, is thought to represent a more visible manifestation of the culture, which can be seen as its 'mood state,' at a particular moment in time (Cox & Flin, 1998). It is generally accepted that the questionnaires provide a measure of the safety climate of the organization at the time of the study.

Guldenmund (2007) describes the safety climate questionnaire as a quick but 'dirty' instrument to assess an organization's state of safety. It is quick because a large amount of data can be collected with fairly minimal effort required by either the researchers, or the participants. It is 'dirty' because the survey only gives a broad perspective of the safety climate of an organization. Questionnaires are also appealing as they allow for statistical comparisons between different variables such as type of job or seniority. However, despite the prevalence and appeal of questionnaires, a number of researchers are beginning to

question the usefulness of the methodology for collecting valid information on safety climate (e.g., Antonsen, 2009; Guldenmund, 2007).

Guldenmund (2007) states that *“a self-administered questionnaire is a valuable tool in (social scientific) research. In organizational culture research, however, certain conditions apply which might make the self-administered questionnaire less useful”* (p.725).

These conditions contribute to the measurement error (the difference between what is true and what is measured) of the survey. The conditions identified by Guldenmund (2007) are:

- in most of the safety climate literature, there are insufficient participants to average out random influences;
- the scales used to record the responses (e.g., Likert scales) are assumed to be interval, when they are actually ordinal, rendering parametric multivariate statistical techniques invalid;
- there is confusion in safety climate research in distinguishing between attitudes and perceptions; and
- organizations are not a closed system, and there are outside conditions (e.g., national culture, socio-economic circumstances) that influence the workforce (particularly if an organisation has not had any serious mishaps).

Both reliability and validity are affected by measurement error. Random error has an impact on reliability, while non-random measurement error directly affects validity. The first two conditions identified by Guldenmund (2007) contribute to random measurement error and can be addressed with a sufficiently large number of responses. However, as will be shown, the presence of non-random measurement error in safety climate survey data also is a concern, but is rarely discussed in the literature.

3. Objectives

The purpose of this paper is to provide evidence, using the descriptive analysis of a large sample of safety climate questionnaire responses, for the presence of non-random measurement error in safety climate questionnaire data. The issues identified with the data reported in this paper are certainly not confined to safety climate questionnaires, and have been shown to be weaknesses with questionnaire methodology more generally (e.g., Krosnick, 1999; Marsden & Wright, 2010). Nevertheless, despite the almost exclusive focus on the use of a survey methodology for collecting safety-climate data, the limitations of this technique are rarely discussed within the context of the safety climate literature. Reasons why safety climate survey data are at particular high risk of non-random measurement error will be presented, and the implications for future research delineated.

4. The Command Safety Assessment Survey (CSAS)

Just as has been the case in other high-risk industries, U.S. Naval aviation recognized that there was a need to develop a tool to assess safety climate, and provide feedback to a squadron's (a military aviation unit with a total of 12 to 24 aircraft, depending on aircraft type) Commanding Officer (COs) on the state of the safety climate. To achieve this goal the Command Safety Assessment Survey (CSAS) was developed by researchers at the Naval Postgraduate School (Desai et al., 2006). The 61-item CSAS is completed on-line, and responses are obtained for each item on a five point Likert scale from 1 (strongly disagree) to 5 (strongly agree)¹. A small number of additional questions elicit demographic and organizational data. The responses are anonymous.

The theoretical background underpinning the CSAS is based upon a conceptual framework of Organizational Safety Effectiveness (MOSE) that identified five major areas

¹ A list of the CSAS items is available in the appendix of Adamshick (2007).

relevant to organizations in managing risk and developing a climate to reduce mishaps in high reliability organizations (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 Command Safety Assessment Survey (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). The specific results of a squadron's CSAS are only available to the CO. However, aggregated data are made available to all COs for comparison of their squadron's performance with that of their peers. The CSAS is typical of many of the safety climate questionnaires reported in the literature in terms of the length of the survey and the questions that are asked. However, there are two important differences in the administration of the CSAS that distinguish it from the typical method that is reported in the literature.

Until 2004, completion of the CSAS was voluntary. However, in 2004 Vice Admiral Zortman declared the CSAS mandatory for all squadrons to complete semiannually and within 30 days following a change of command (a change of command is when a new CO takes over a squadron, and occurs approximately every two years; Zortman, 2004). Thus, in practice a particular aviator is likely to be required to complete the CSAS at least twice in any given two year period. For typical safety climate surveys participation is voluntary and the study is generally a 'single shot' assessment.

To date three published studies that have utilized CSAS data (Desai et al., 2006; Gaba, Singer, Sinaiko, Bowen, & Ciavarelli, 2003; Singer, Rosen, Zhao, Ciavarelli, & Gaba, 2010). However the factor structure of the CSAS was not examined. We recognized the lack of a validated factor structure as a hindrance to the advancement of the instrument itself, and importantly, it's potential to prevent mishaps. Given this fact, it was our original intention to conduct a confirmatory factor analysis in order to establish the construct validity of the CSAS. Once a stable factor structure was established, researchers could begin to evaluate the predictive validity of the instrument or model the factor structure for comparison within and across industries. However, during the initial data screening process (described in this paper), it quickly became evident that there were some serious threat to the validity of the data collected using the instrument.

5. Dataset

Following approval by Commander Naval Air Forces, all of the responses to the CSAS from 2000 until 2009 were obtained. The data contains 110,014 responses to each of the 61 items in the CSAS. A total of 65% responses were obtained from Navy aircrew, and 31% from Marine Corps aircrew, the remainder being identified as "other." Of the respondents 67% were officers, and 33% were enlisted personnel (predominately aircrew). The officers primarily consisted of pilots and Naval Flight Officers (NFOs; officers who specialize in airborne weapons and sensor systems, but do not actually fly the aircraft). Only 3% of aircrew were officers, but 94% of NFOs and 99% of pilots were officers.

A total of 16% of respondents flew in TACAIR aircraft (Tactical Aviation, which includes multirole fighter aircraft such as the F/A-18 Hornet and E/A-8 Prowler), 32% in rotary wing aircraft (helicopters such as the SH-60 Seahawk), 31% in big wing aircraft (large transport and surveillance aircraft such as the C-130 Hercules and P-3 Orion), 13% in

training aircraft, and the remaining 5% did not provide information on the aircraft flown. The response rate from the entire population of U.S Naval aviators was 75.5%.

Given the frequency with which the questionnaire was completed, in the majority of cases the same respondent will have answered the survey on multiple occasions. However, as the responses are anonymous it was not possible to identify the repeated survey responses of an individual respondent. As a result of this fact, it was necessary to treat each set of responses to the CSAS as independent observations. Although not ideal, it is proposed that the assumption of independence is not a major limitation to the study for the following reasons. Firstly, there is constant turnover of personnel in a U.S Naval aviation squadron. Aviators will generally only spend two to three years in a particular squadron before they move on to a different squadron or other non-flying assignment. Therefore, even within the same squadron there is likely to be changes in the safety climate as personnel rotate in and out. Secondly, the CSAS generally is completed twice over two year span of time. Therefore, a respondent is unlikely to recall how he or she responded to the CSAS items the last time they completed the survey.

6. Evidence of substantial non-random measurement error

Below we provide five different pieces of evidence that support our claim of substantial non-random measurement error in the responses to the CSAS.

6.1 Reversed items displayed a unique response pattern

Five items from the CSAS had negatively worded items. In contrast to the majority of the CSAS items, a Likert rating of 1 “strongly disagree” to a negatively worded item is indicative of a desirable response. These items were:

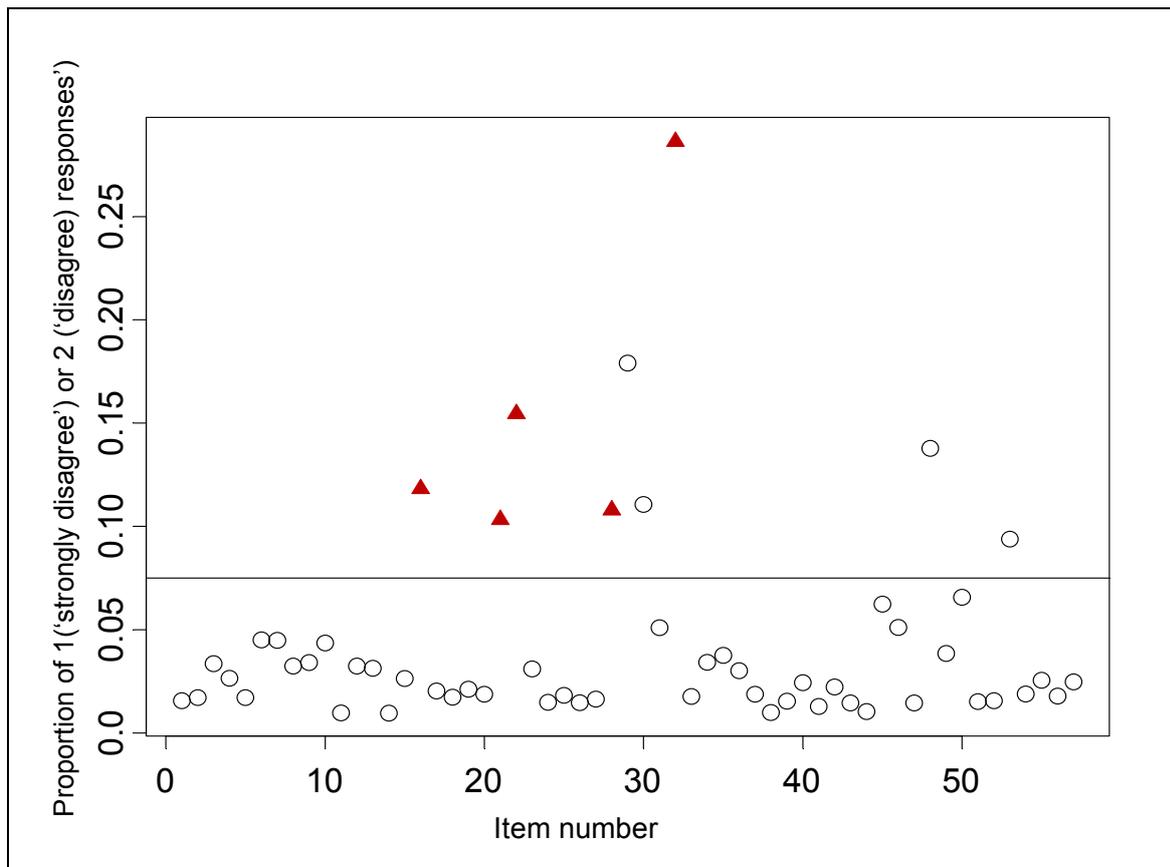
- Item 18 — *I am not comfortable reporting a safety violation, because people in my command would react negatively toward me;*
- Item 23 — *Command leaders permit cutting corners to get a job done;*
- Item 24 — *Lack of experienced personnel has adversely affected my command's ability to operate safely;*
- Item 30 — *My command has increased the chances of a mishap due to inadequate or incorrect risk assessment; and*
- Item 34 — *Based upon my command's personnel and other assets, the command is over-committed.*

As part of the data screening process, the responses to these five items questions were reversed so that Likert ratings of '1' were recorded as '5' and so on by subtracting the response from six. Despite the reversal of the answers for these questions there is some evidence that respondents were confused when it came to answering them. Overall, across all questions, about 4.5% of items were rated a '1' or '2' but 16% of the reversed questions were given a rating of '1' or '2'. Put another way, although reversed questions accounted for only 8.8% of total responses, they accounted for 33.9% of all '1' ratings. (Recall that a '1' in this analysis corresponds to an original answer of '5' by the respondent.) The proportion of '1' and '2' Likert ratings can be seen in Figure 1. Note that for all five reversed questions, more than 7.5% of ratings to each of these items is a '1' or a '2' (the horizontal line indicates 7.5%). Out of the other 52 items, only four were rated '1' or '2' more that 7.5% of the time. These items were:

- Item 31 — *I am provided adequate resources, time, staffing, budget, and equipment, to accomplish my job;*
- Item 32 — *My command provides the right number of flight hours per month for me to fly safely;*

- Item 50 — *Morale and motivation in my command are high*; and
- Item 55 — *Within my command, good communications flow exists up and down the chain of command*.

Figure 1. Proportion of "1" or "2" Likert responses by item (triangles indicate the negatively worded items and the circles indicate the positively worded items).



It is difficult to discern whether the large number of “1” and “2” Likert ratings to items 31, 32, 50, and 55 is paralleled in the large number of those responses in the reversed questions. Our judgment is that the former set of “1” and “2” Likert ratings probably reflect real dissatisfaction and that the latter is most likely due, at least in part, to confusion on the part of the respondents.

6.2 Little item-to-item variability in responses

As discussed in section five, although we had to treat each set of responses to the 61 CSAS items as independent, we were able to examine the pattern of responses to the items on each occasion the questionnaire was completed. It was found that there was very little variability in the range of responses provided to the items. Out of the 110,014 times the questionnaire was completed, for 8.0% the very same Likert rating was given to all items (this percentage is based upon 52 items- it does not include the nine items identified in section 6.1). Additionally, for 24.9% of the times the questionnaire was completed, only two possible Likert ratings were provided across the set of 52 items. On only 31% of occasions were four or all five possible Likert ratings used. Indeed on many of the occasions the questionnaire was completed, almost all items were answered with the same Likert rating. For example, on more than half of the occasions the questionnaire was completed, the same Likert rating was provided to at least 40 items.

Little variability in the responses to large numbers of questionnaire items is a threat to both construct and discriminate validity (see O'Connor, O'Dea, Kennedy, & Buttrey, 2011 for a detailed discussion of these concepts within the context of safety climate surveys). Construct validity is the extent to which the questionnaire measures what it is intended to measure. Identification of a reliable factor structure, that is consistent with theory, helps the researcher substantiate claims regarding the construct validity of the questionnaire. Establishing a reliable factor structure is generally carried out using exploratory and/or confirmatory factor analysis techniques. However, a large number of items with little variability leads to a lack of normality among the data and invalidates exploratory and confirmatory factor analysis methods. Another issue is that low variability tends to result in principle components analysis (often used to carry out exploratory factor analysis in the

literature; O'Connor et al., 2011) identifying only one principle component including all of the items.

Discriminate validity is the ability of the questionnaire to differentiate between organizations or personnel with different levels of safety performance. If a large proportion of respondents agree (or disagree) with a particular item, then the item is not useful in distinguishing between high- and low- safety performing groups. The danger of retaining a large number of nondiscriminating items when exploring differences between different groups of respondents is that the discriminating items can become “washed out” when they are averaged with nondiscriminating items.

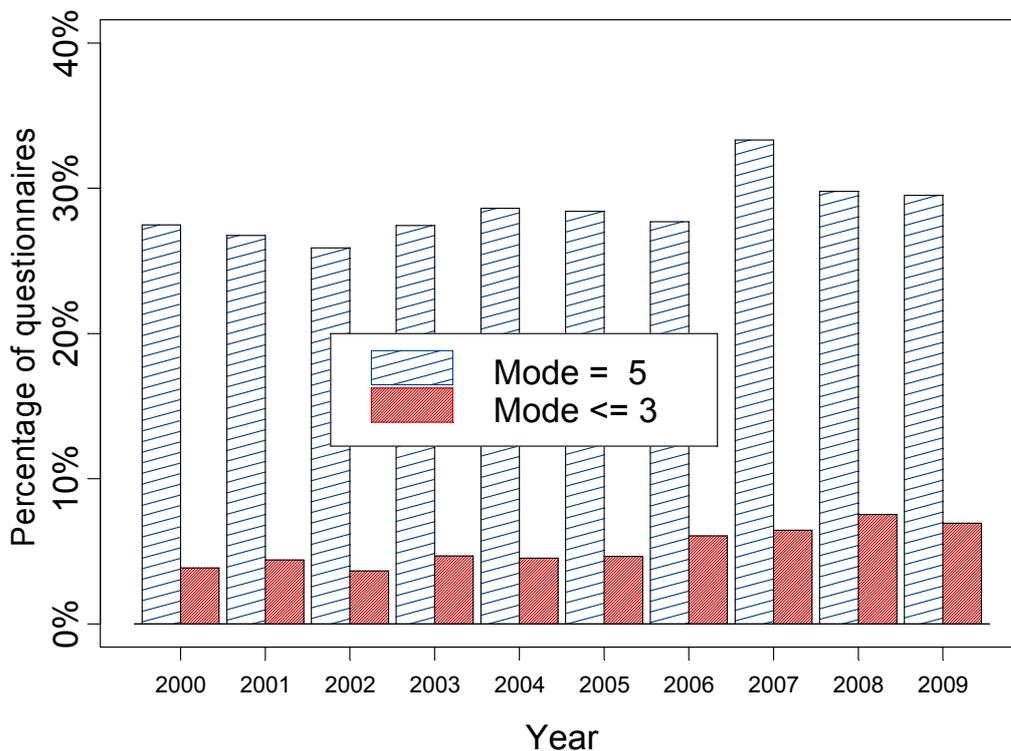
6.3 Changes in response pattern across time

As described above, participation in the survey was voluntarily until October 2004, after which time it became mandatory. The mode of the Likert ratings to the 52 items retained for analysis was calculated for each individual occasion that the questionnaire was completed. If a Likert rating to a specific item was made at that mode, this will be described as the ‘modal response.’ To illustrate, if the mode of the Likert rating on a particular occasion that the questionnaire was completed was ‘4’, with the participant using a rating of ‘4’ for 40 items, then the modal response was given 40 times out of 52. Similarly, if on another occasion the questionnaire was completed the mode of the Likert rating was ‘5’ with a rating of ‘5’ given for 35 items, then the modal response was given 35 times. It was found there was an increase in the frequency of modal responses for those questionnaires that were completed post-October 2004 as compared to pre-October 2004.

In fact, there are two changes that appear to take place in the response patterns to the questionnaire items as years go by, not just as the boundary between voluntary and mandatory reporting is crossed. First, there was an increase over the years in the frequency of

questionnaires for which the mode of the ratings was ‘5’, and also of the frequency of questionnaires for which the mode of the ratings was ‘3’ or less. Figure 2 shows the distribution of questionnaires for which the mode of the ratings was other than ‘4’; the sum of these went from a low of 29.5% in 2002 to a high of 39.8% in 2007. The effect of this change is to decrease the variance in the ratings, leading to a lack of homogeneity of variance. The effect of this non-constant variance is to render invalid the standard statistical tests that rely on it being constant (e.g. factor analysis, analysis of variance).

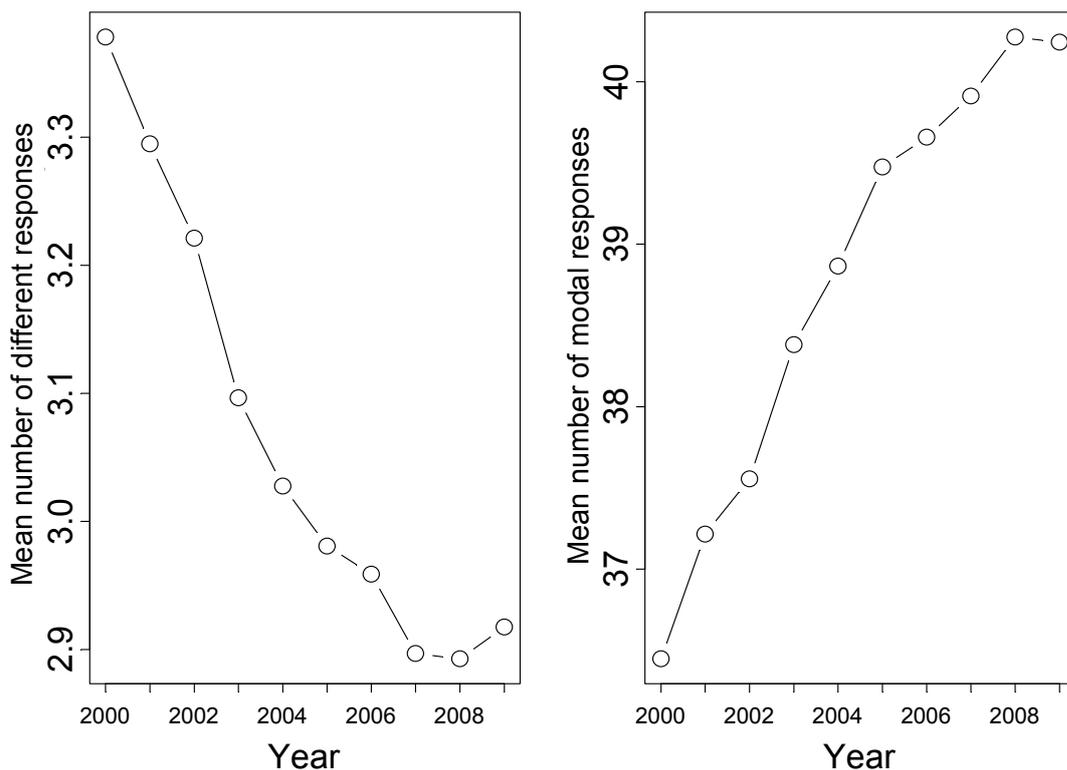
Figure 2. Percentage of questionnaires for which the modal response was less than or equal to 3, and those equal to 5, by year.



Second, and perhaps more importantly, there was also a different kind of change in the variance over time that acted to decrease it. Figure 3 (left) shows the average number of

different Likert ratings to the items on each occasion the questionnaire was completed, by year. The downward trend indicates that increasingly over the years only a few of the possible range of Likert ratings were being used. The right panel shows the average number of modal ratings each time the questionnaire was completed, by year. In 2009 a modal rating was given to more than 40 of the 54 questions, up from around 36 in 2000. Therefore, although these two shifts have opposing effects on the variance, the overall outcome is a lack of homogeneity of variance of the data.

Figure 3. Decrease in number of non-modal responses (left) and increase in the mean number of modal responses (right) each time the questionnaire was completed, by year.



6.4 Lower than desirable correlation between responses to 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 data collected prior to 2008, the correlation between the sets of responses for these two items 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.

6.5 Relationship between time to complete and number of modal responses.

From 2006 onwards data were collected on the length of time taken to complete the CSAS. Table 1 shows the proportion of modal responses, broken down by whether the time to complete was less than ten minutes, unrecorded, or greater than ten minutes. Ten minutes was selected as the cut-off through a detailed analysis of the proportion of modal responses across 60 second intervals.

Table 1. Proportion of modal responses for each occasion the CSAS completed, by time taken to complete.

# of responses at mode	Time to complete survey		
	≤ 10 mins.	>10 mins.	Not timed
13-29	12.1	20.3	17.7
30-39	24.2	33.3	32.6
40-44	15.1	18.2	17.7
45+	48.6	28.3	31.9
Sample size	24154	23442	62418

Table 1 highlights substantial differences in response patterns when the survey was completed quickly as compared to when more time was taken to complete the survey. The

last column (the untimed data prior to 2006) appears, not surprisingly, to be a blend of the two types of responses.

7. Summary of descriptive analysis

This descriptive data analysis indicates non-random measurement error associated with the CSAS response set. The reverse worded items had a unique pattern of responses, there was an increasing tendency over the years to only provide modal responses to the items, 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.

An individual who is motivated to participate in a survey will be more likely to provide a response that is an accurate reflection of their true attitude. However, if a respondent lacks the motivation to participate then he or she is more likely to adopt a satisficing response strategy to select a reasonable answer. Satisficing is a strategy in which the respondent interprets each question superficially and selects what he or she believes to be a reasonable answer. The answer *“is selected without referring to any internal psychological cues relevant to the attitude, belief, or event of interest. Instead, the respondent may look to the wording of the question for a cue, pointing to a response that can be easily selected and defended if necessary. If no such cue is present, the respondent may arbitrarily select an answer”* (Krosnick, 1999: 548).

Motivation to participate will be higher if the questionnaire is constructed in such a way that it appears to be quick and easy to complete (Dillman, 1978). Motivation is also affected by the degree to which the topic of a question is personally important to the respondent, beliefs about whether the survey will have useful consequences, respondent fatigue, and aspects of questionnaire administration that encourage, or discourage

participation (Krosnick & Presser, 2010). Each of these influences will be discussed specifically within the context of safety climate surveys.

8. Questionnaire design

When designing safety climate questionnaires, researchers must consider the effect on respondent motivation of item complexity, the number of items in the questionnaire, and the use of reverse worded items. These factors are discussed in detail below.

8.1 Item complexity.

In comparison to, for example, personality or health questionnaires with items such as *'I often feel blue'* or *'to what degree do you experience pressure due to your job?'* the items that are typically included in safety climate surveys are very complex and open to different interpretations by respondents. To illustrate, consider the CSAS item, *'within my command, good communications flow exists up and down the chain of command.'* It may be that communications are good down the chain of command, but not up. It is also possible that communications about some issues is good, but it is not the case for others. The item may be better written as *"within my command there is good communication about safety"*, and/or separating the item to distinguish between communication 'up', and communication 'down' the chain of command.

Another example of a typical safety climate questionnaire item is, *'my management care about the negative effect that job uncertainty has on safety'* (taken from the offshore safety questionnaire; Mearns, Whitaker, Flin, Gordon, & O'Connor, 2003). It may be that some managers do, and others do not. The term 'management' is also very ambiguous in the context of the offshore industry as the majority of the workforce are contractors and so management could refer to either the contract company managers or the operating company

mangers. The item would be less ambiguous if it referred to a specific management role such as the Oil Instillation Manager (the most senior manager on an oil platform).

These two exemplar items are typical of those included in safety climate surveys. If a respondent is not motivated, does not understand the question, or does not feel he or she can answer the question then it is unlikely that the required cognitive effort will be used or the person simply may be unable to provide an accurate and thoughtful response. Therefore, researchers should make efforts to ensure that the items in the questionnaire are simple, and readily understood by participants.

8.2 Number of items.

As would be expected, as the number of items in a questionnaire increases, there is a corresponding reduction in the number of people willing to complete it (Sheehan, 2001; Smith, Olah, Hansen, & Cumbo, 2003). However, safety climate surveys, the CSAS included, generally consist of a large numbers of items. The Veteran Health Administration survey has 112 items (Singer et al., 2003). In the 18 safety climate questionnaires reviewed by Flin et al., 2000, the mean number of items was 60.8, (st dev=28.8). Studies have reported more missing data, higher levels of agreement, and less differentiation among items when they appear later in a questionnaire as compared to when the same items are placed earlier in the questionnaire (e.g., Johnson, Sieveking, & Clanton, 1974). Therefore, researchers should give careful consideration to the number of items in safety climate questionnaires, and make efforts to keep the questionnaire as short as possible.

8.3 Reverse worded items.

The responses to the reverse worded items in the CSAS demonstrated a unique response pattern. There is a large literature on issues with internal consistency, factor

structures, and other statistics when negatively worded items are used in combination with positively worded items (e.g., Barnette, 2000; Herche & Engelland, 1996). Nevertheless, despite the recognition that negatively worded items can introduce measurement error, they are used in many safety climate questionnaires (e.g., Antonsen, 2009; Mearns et al., 2003; Gibbons et al., 2006) and little mention is made of the potential problems with negatively worded items beyond the fact that the responses were reversed for analysis. Therefore, careful consideration should be given to the inclusions of reverse worded items in safety climate surveys, and these items should be screened to ensure they do not have a unique response pattern.

9. Belief in the importance and usefulness of the questionnaire

If a survey participant believes his or her response is important, has an interest in the topic area, and believes that the response will have consequences, then the participant will be more likely to expend the effort required to provide a thoughtful response. Squadron COs are not required to share the CSAS findings with squadron members, and information is not available about the prevalence with which COs share the survey results. However, if similar organizational surveys have been carried out in the past with no feedback to participants, and no evidence of changes resulting from the survey, then it stands to reason that participants will be less likely to exert effort in completing the survey. To illustrate, Ward (1994) found that General Practitioners are less likely to participate in future surveys if they are given insufficient feedback from previous participation. Although this is not likely to be within the control of the researcher, senior leadership must demonstrate support for the survey. Support could be demonstrated by providing time during the work day to complete the questionnaire, explaining the purpose of the survey, and providing assurances the information is important and the findings will be acted upon.

Another issue that is likely to have a negative impact on beliefs about the importance and usefulness of survey response is whether the participants themselves believe they have attitudes towards safety climate worthy of reporting. As can be seen from Figures 2, the vast majority of respondents either ‘agree’ or ‘strongly agree’ with most of the CSAS items. If the respondents are generally satisfied with the state of the safety climate in the squadron (as measured by the survey), then they may not feel that it is necessary to carefully consider each and every survey item before responding.

10. Respondent fatigue

As discussed earlier, questionnaire length and complexity are likely to have a negative effect on respondent motivation, and will also contribute to respondent fatigue. U.S. Naval aviators complete a large number of mandatory surveys (e.g., equal opportunities, command climate). Moreover, the CSAS is one of an increasing number of behavior-based safety programs that have been introduced in naval aviation in the last decade. 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” and “survey fatigue” within naval aviation. These types of fatigue are not specific to U.S Naval aviation, and are likely to be prevalent in other organizations in which safety climate surveys are utilized. Therefore, researchers must attempt to avoid periods of particularly high workload, or immediately following another survey when distributing safety climate surveys.

11. Questionnaire administration.

11.1 Mandatory completion.

As described in the introduction, since 2004, naval aviators are mandated to complete the questionnaire. The analysis showed that there was an increase in the number of modal responses after completion was made mandatory as compared to when participation was voluntary. As the respondents who do not wish to participate cannot opt out, a large proportion of aircrew may use a strategy of questionnaire completion that requires as little cognitive effort as possible (i.e. adopt a satisficing strategy), but still allows them to “do their duty”. This argument is supported by the finding that on 50% of the survey occasions, the CSAS was completed in less than 10 minutes.

The risk is that those respondents who have strong opinions about the state of the safety climate are being ‘washed out’ by the high numbers adopting a satisficing strategy. So, although a large proportion of naval aviators completed the CSAS, the quality of the data were compromised. Therefore, of greater importance than having a large proportion of a population completing the questionnaire is the extent to which the responses obtained are a true representation of that population’s attitudes and beliefs. It would seem that researchers should examine the selection bias of volunteers as compared to respondents who are ‘compelled’ to participate in survey research..

11.2 *Survey frequency*

Apart from U.S. Naval aviation, there are few examples of the use of the same safety climate survey on more than one occasion within the same organizations (Mearns et al., 2003 being a notable exception). Pronovost and Sexton (2005) recommend annual safety climate surveys of hospital staff, and we believe that the use of safety climate questionnaires on multiple occasions within the same organization is likely to become more commonplace. From both an operational and research perspective, measuring the safety climate of an organization on more than one occasion is desirable. Multiple measurements allows an

assessment to be made of how safety climate is changing over time and provides data that allows for the evaluation of the effectiveness of programs designed to impact safety climate.

It is surprising that there is almost no research on the impact of multiple survey requests on survey response reported in the literature (Porter, Whitcomb, & Weitzer, 2004). However, as shown in Figure 3, the increase in the number of modal responses over time suggests that quality of information is being sacrificed for quantity. Therefore, researchers should give careful consideration on the frequency with which safety climate surveys are administered in order to reduce the likelihood of survey-taking fatigue.

12. Implications of the findings for future safety climate research

We believe safety-climate data collected using a complex and lengthy survey leads to substantial non-random measurement error. Although the requirement for mandatory completion is unique to the CSAS, the other risk factors discussed above are common across the majority of safety climate surveys reported in the literature. We are not suggesting the abandonment of surveys as a method for assessing safety climate, but researchers must make efforts to limit non-random measurement error in the data through survey design, survey administration, and motivating respondents. In agreement with Antonsen (2009), a triangulation of quantitative and qualitative is recommended for assessing the safety climate of an organization. Recently, qualitative assessments of safety climate have begun to be reported in the literature (e.g., Atak & Kingma, 2011; Haukelid, 2008; Richter & Koch, 2004). Other researchers also agree that a combination of quantitative and qualitative techniques provides a comprehensive evaluation of the safety climate of an organization (e.g., Antonsen, 2009; Gibbons et al., 2006).

In the U.S Navy, squadron COs are able to obtain a qualitative assessment of the safety climate of his/her squadron by requesting a safety culture workshop. The purpose of

the safety culture workshop identifies potential hazards that might interfere with mission accomplishment. They also identify command strengths. A safety culture workshop is facilitated by specially trained senior naval aviators. 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. Currently, safety culture workshops are not mandatory, and are run independently from the CSAS. It is suggested that these two programs should be integrated to allow a more complete picture of the safety climate of the squadron to be obtained.

13. Conclusion

The descriptive analysis of the CSAS data reported in this paper has demonstrated the presence of non-random measurement error in safety climate survey data. We readily acknowledge that the issues identified in this paper are not new, nor unique to safety climate questionnaires. However, although the threat of non-random measurement error would seem to be particularly high with safety climate surveys, it is largely ignored in the safety climate literature.

We believe that the evidence presented here supports the conclusion that the use of long and complex surveys should cease to be *the* measurement method for assessing safety climate. It is recommended that a triangulation of quantitative and qualitative methods should be used to obtain insights into the true safety climate of organizations.

Acknowledgement

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 paper was supported by funding from the Defense Safety Oversight Council.

References

- Adamshick, M.H. (2007). *Leadership and safety climate in high-risk military organizations*. PhD Dissertation, College Park, MD: University of Maryland.
- Antonsen, S. (2009). Safety culture assessment: A mission impossible? *Journal of Contingencies and Crisis Management*, 17(4), 242-254.
- Atak, A., & Kingma, S. (2011) Safety culture in an aircraft maintenance organisation: A view from the inside, *Safety Science*, 49(2), 268-278.
- Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, 60, 361–370.
- Cox, S., & Flin, R. (1998). Safety culture: Philosopher's stone or man of straw. *Work and Stress*, 12, 189-201.
- 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.
- Dillman, D.A. (1978). *Mail and telephone surveys*. New York, NY: Wiley
- Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Safety climate: Identifying the common features. *Safety Science*, 34, 177-192.

- Gaba, D.M., Singer, S.J., Sinaiko, A.D., Bowen, J.D., & Ciavarelli, A.P. (2003). Difference in safety climate between hospital personnel and naval aviators. *Human Factors*, 45, 173-185.
- Gibbons, A.M., von Thaden, T.L., & Wiegmann, D.A. (2006). Development and initial validation of a survey for assessing safety culture within commercial flight operations. *International Journal of Aviation Psychology*, 16(2), 215-238.
- Guldenmund, F. (2007). The use of questionnaires in safety culture research – an evaluation. *Safety Science*, 45(6), 723-743.
- Hauklid, K. (2008). Theories of (safety) culture revisited- An anthropological approach. *Safety Science*, 41 (8), 413-426.
- Herche, J., & Engelland, B. (1996). Reverse-polarity items and scale unidimensionality. *Journal of the Academy of Marketing Science*, 24(4), 366-374.
- Johnson, W. R., Sieveking, N. A., & Clanton, E.S. (1974). Effects of alternative positioning of open-ended questions in multiple-choice questionnaires. *Journal of Applied Psychology*, 6, 776-778.
- Krosnick, J.A. (1999). Survey research. *Annual Review of Psychology*, 50, 537-567.
- Krosnick, J.A. & Pressner, (2010). Question and questionnaire design. In: Marsden, P.V. & Wright, J.D. (Eds.). *Handbook of survey research*, (pp. 263-314). Bingley, UK: Emerald.
- Libuser, C.B. (1994). *Organizational structure and risk mitigation*. Ph.D. Dissertation, Los Angeles, CA: University of California at Los Angeles.
- Marsden, P.V., & Wright, J.D. (2010; Eds.). *Handbook of survey research*. Bingley, UK: Emerald.
- Mearns, K., & Flin, R. (1999). Assessing the state of organizational safety – culture or climate. *Current Psychology*, 18(1), 5-17.

- Mearns, K., Whitaker, S., Flin, R., Gordon, R., & O'Connor, P. (2003). *Factoring the human into safety translating research into practice. Volume 1: Benchmarking human and organisational factors in offshore safety*. Norwich, England: Her Majesty's stationary office.
- 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'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.
- Porter, S.R., Whitcomb, M.E., & Weitzer, W.H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, 121, 63-73.
- Pronovost, P., & Sexton, B. (2005). Assessing safety culture: guidelines and recommendations. *Quality and Safety in Health Care*, 14, 231–233.
- Richter, A., & Koch, C. (2004). Integration, differentiation and ambiguity in safety cultures. *Safety Science*, 42 (8), 703-722.
- Sheehan, K. B. (2001). E-mail survey response rates: A review. *Journal of Computer-Mediated Communication*, 6(2). Retrieved April 16, 2010, from <http://www.ascuse.org/jcmc/vol6/issue2/sheehan.html>
- Singer, S.J, Rosen, A., Zhao, S., Ciavarelli, A., & Gaba, D. (2010). Comparing safety climate in naval aviation and hospitals: Implications for improving patient safety. *Health Care Management Review*, 35(2), 134-146.
- Singer, S.J., Gaba, D.M., Geppert, J.J., Sinaiko, A.D., Howard, S.K., & Park, K.C. (2003). The culture of safety: results of an organization-wide survey in 15 California hospitals. *Quality and Safety in Health Care*, 14, 231–233.

- 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.
- Yule, S., O'Connor, P., & Flin, R. (2003). Testing the structure of a generic safety climate survey instrument. Paper presented at the 5th Australian Industrial/Organisational Conference, Melbourne, Australia.
- Zortman, J.M. (2004). *CNAF Commanders training symposium safety wrap-up*. Unclassified General Administrative Naval Message: R 240054Z NOV 04.