Cite as: O'Connor, P., Jones, D., McCauley, M., & Buttrey, S. (2012). An evaluation of the effectiveness of the U.S Navy's crew resource management program. *International Journal of Human Factors and Ergonomics, 1(1)*, 21-40.

AN EVALUATION OF THE EFFECTIVENESS OF THE CREW RESOURCE MANAGEMENT PROGRAM IN NAVAL AVIATION

Paul O'Connor[1], Doug Jones[2], Michael McCauley[2], & Samuel Buttrey[2]

1. **Corresponding Author:** Centre for Innovation and Structural Change, J.E. Cairnes School of Business & Economics. National University of Ireland, Galway. Ireland. Email: paul.oconnor@nuigalway.ie; poc73@hotmail.com. Tel: +353 (0) 91 492897 Fax: +353 (0) 91 495524.
2. Naval Postgraduate School, 1411 Cunningham Rd, Monterey, California 93943, USA.

AN EVALUATION OF THE EFFECTIVENESS OF THE CREW RESOURCE MANAGEMENT PROGRAM IN NAVAL AVIATION

ABSTRACT

Although every U.S. Naval aviator receives annual Crew Resource Management (CRM) training designed to improve mission effectiveness, the program has not been formally evaluated in the last decade. To assess the adequacy of the training, a multi-faceted method was used. Reactions were evaluated by analyzing 51,570 responses, collected over nine years, to an item pertaining to CRM that is part of a safety climate survey. A total of 172 responses were obtained on a multiple-choice knowledge test. The attitudes of 553 naval aviators were assessed using a measure based upon the Cockpit Management Attitudes Questionnaire. The CRM mishap rate from 1997 until 2007 was evaluated. It was found that naval aviators appear to think than CRM training is useful, are generally knowledgeable of, and display positive attitudes towards, the concepts addressed in the training. However, there is a lack of evidence to support the view that CRM training is having an effect on the mishap rate. The results suggest that there is a need to focus training on behaviors, with a particular emphasis on performance under stress. Further, a behavioral marker system should be used evaluate whether the CRM skills taught are translating into the appropriate behaviors in the aircraft.

KEY WORDS: Crew Resource Management, training evaluation, safety

RUNNING HEADER: Effectiveness of the CRM program in Naval aviation

INTRODUCTION

The U.S. Navy and Marine Corps instituted Crew Resource Management (CRM) in the late 1980s.The goal of the U.S. Navy's aviation CRM training is to "*improve mission effectiveness by minimizing crew preventable errors, maximizing crew coordination, and optimizing risk management*" (Chief of Naval Operations, 2001). The Navy's CRM training is based upon seven critical skills (decision making, adaptability/ flexibility, situational awareness, mission analysis, communication, assertiveness, and leadership; Prince & Salas, 1993). Every naval aviator must receive ground training and a CRM evaluation during an actual or simulated flight, by a CRM instructor, or facilitator, once a year (for a detailed discussion see O'Connor, Hahn, & Salas, 2010).

The Navy's CRM program is governed by a Chief of Naval Operations (CNO) Instruction – OPNAVINSTR 1542.7C. The CNO instruction outlines a rudimentary foundation of CRM program academics and the behaviors the program aims to achieve. This instruction also sets out how the CRM program must be implemented by each aviation squadron. The U.S Navy's CRM program is centrally controlled, but each aviation community is given the latitude to administer its own CRM program tailored for its particular aircraft and mission. In a sense, the Navy and Marine Corps have 48 separate, but closely related, CRM programs (one for each type model aircraft the Navy and Marine Corps flies). Each program is inspected to ensure that it is in compliance with the CNO instruction in the Program Managers' programs, and that the academic literature is consistent with the U.S Navy's CRM curriculum. Although there have been minor updates, the basic CRM instruction in naval aviation has not changed greatly in the last decade. The last systematic update of the training curriculum was in 1999 (Oser, Salas, Merket, & Bower., 2001).

The Federal Aviation Administration (FAA) states that for CRM training "*it is vital that each training program be assessed to determine if CRM training is achieving its goals. Each organization should have a systematic assessment process. Assessment should track the effects of the training program so that critical topics for recurrent training may be identified and continuous improvements may be made in all other respects*" (FAA, 2004: 12). Similarly, Nullmeyer, Spiker, Wilson, and Deen (2003) assert that CRM trainers struggle to maintain the resources necessary to support CRM training. They suggest that in the absence of measurable CRM training objectives, it is difficult to differentiate between effective and ineffective programs.

There have been a number of comprehensive reviews (O'Connor, Flin & Fletcher, 2002; Salas, Burke, Bowers & Wilson, 2001; Salas, Wilson, Burke, & Wightman, 2006), and a meta-analysis (O'Connor et al., 2008) of the effectiveness of CRM across a range of high-risk industries. There has also been a review of 27 studies that specifically examined the effectiveness of military CRM training (O'Connor, Hahn, & Nullemeyer, 2010). These reviews all used Kirkpatrick's (1976) training evaluation hierarchy to categorize the different methods of effectiveness used.

The lowest level of Kirkpatrick's (1976) hierarchy is an evaluation of reactions. Reactions are the equivalent to measuring satisfaction. For example, did the participants like the training? Learning is the second level in the hierarchy, and refers to "*the principles, facts, and skills which were understood and absorbed by the participants*" (Kirkpatrick, 1976: 11). Learning is made up of two components: attitudinal change and knowledge gains. Evaluation of behavioral changes is the assessment of whether knowledge learned in training actually transfers to behaviors on the job, or a similar simulated environment. The organizational level is the highest in Kirkpatrick's

(1976) hierarchy. The aim of any training program is to produce evidence of an organizational level effect such as an improvement in safety and/or productivity.

These reviews concluded that the reactions to CRM training were positive, and the training was found to have a positive effect on the attitudes, knowledge, and behaviors of participants. Although the review of military CRM training surmised that it had a positive effect on mishap rate, the other cross-industry reviews stated that there was a lack of evidence regarding the effect of CRM training on the organization

There have been eight published studies that have specifically examined the effectiveness of the U.S. Navy's CRM program (see O'Connor, Hahn, & Salas, ,2010 for a summary of these studies). In agreement with the other reviews it was found that U.S Navy CRM participants liked the training, were more knowledgeable as a result of the training, there was generally a positive shift in attitudes toward the CRM concepts, and the training resulted in improved team coordination. However, the effect of CRM training on the mishap rate in naval aviation was inconclusive. Although Alkov and Gaynor (1991) reported a reduction in the aircrew error mishap rate, Wiegmann and Shappell (1999) concluded that CRM training had not affected the mishap rate.

The results of these evaluations of the U.S Navy's CRM training are encouraging. However, we cannot draw conclusions about the effectiveness of the U.S. Navy's current CRM program based upon the findings of these studies because:

1. The studies lack currency. They were carried out more than a decade ago (except one study by O'Connor & Jones, 2009, which included a subset of the data analyzed in this paper), so the findings may not reflect on the current U.S. Navy CRM program.
2. The sample sizes of the studies were small. Not including the O'Connor & Jones (2009) study, the mean number of participants was 48.9 (sd = 23.8).

3. The participants in the studies were not representative of all types of naval aviator. The participants were all helicopter aviators (e.g. Salas, Fowlkes, Stout, Milanovich, & Prince, 1999) except for the studies described by Alkov & Gaynor (1991; aviation instructors) and Stout, Salas, & Kraiger. (1996; student naval aviators). The effectiveness of CRM training for aviators who fly other types of aircraft (e.g. fighters or transport) was not examined.
4. The CRM training that was evaluated was not consistent across the studies. To illustrate, Stout, et al. (1996) evaluated five hours of training on an introduction to CRM, communication, and assertiveness. In the first study reported by Salas et al. (1999), the program that was evaluated consisted of the following modules: introduction to CRM, communication, assertiveness, mission analysis, and situation awareness, and in the second study, the training consisted of all of the skills identified by Prince and Salas (1993). Further, Salas et al. (1999) did not report the length of the training.
5. The studies examined the effectiveness of the training immediately after it had been received. No analysis was carried out of the long term effectiveness of the training. The lack of this information is an issue as there is evidence in the literature to suggest that the benefits of CRM training decay over time. For example, in an evaluation of civilian pilots, Irwin (1991) found a decay in positive attitudes toward CRM over time. Helmreich, Merritt , & Wilhelm (1999) also reported a slippage in attitudes to CRM concepts, even with recurrent training.
6. The studies were evaluations of a training course delivered early in development, and not typical of the manner in which CRM training is now conducted in the U.S. Navy.
7. The type of aircraft used by U.S. Naval aviators is not the same now as it was in the early and mid-1990s. A number of new advanced aircraft have been introduced (e.g. V-22 Osprey), and older aircraft retired (e.g. F-14 Tomcat).

As a result of the limitations delineated above, the authors recognized a need to identify whether the Navy's current CRM program is meeting it's goal of improving mission effectiveness. This paper reports a multi-faceted evaluation of the effectiveness of a mature CRM training program. Similar to the reviews of CRM effectiveness, the evaluation methodologies are guided by Kirkpatrick's (1976) training evaluation hierarchy. This paper reports an evaluation of the reactions of course participants, learning (attitudes and knowledge), and organizational level effects (as assessed by examining the mishap rate). Each of the evaluations will be described as an independent study. The paper finishes with general conclusions.

A limitation of all four studies reported in this paper is the lack of a control group who did not receive the training. Every naval aviator is mandated to receive CRM training, and is exposed to CRM as soon as they start flight school. From a scientific perspective, it would have been desirable to withhold CRM training from a number of squadrons and use them as a control group. However, even in the very unlikely event of senior Navy leadership agreeing to this proposal, given the findings supporting the effectiveness of CRM training reported in the literature, it would not be ethical. Therefore, this lack of control groups represents a limitation in the conclusions that can be drawn from these studies.

STUDY 1: REACTIONS TO CRM TRAINING

Reactions: Introduction

Helmreich and Wilhelm (1991) claimed that it is axiomatic that training will have a greater impact if participants perceive it to be useful. In the three previous studies of reactions to the Navy's CRM program, the participants were generally positive, giving a mean usefulness rating of 4.4 on a five point scale (e.g. Salas, et al, 1999).

There is not a requirement for aviation squadrons to assess reactions to CRM training. However, one item from the 61-item naval aviation Command Safety Assessment Survey (CSAS; see O'Connor, O'Dea, Kennedy & Buttrey, 2011 for more information on the CSAS) pertains to the effectiveness of a squadron's CRM program. The question reads as follows: "*My command's Crew Resource Management (CRM) program is helping to improve mission performance and safety*." Responses are obtained using a 5-point Likert scale (1="disagree strongly" to 5="agree strongly").The CSAS is an online survey completed periodically by all members of a naval squadron (approximately every two years).

The responses to the CRM item from the CSAS were examined to assess whether reactions have changed over time, and if there were differences on the basis of aircraft flown or rank. The reason for examining these two variables is that they may exert an effect on the acceptance of the concepts addressed in CRM training.

For the purpose of all of the studies reported in this paper, naval aviation will be divided into three broad aviation communities based upon aircraft flown: TACAIR (Tactical Aviation, includes multi-role fighter aircraft such as the F/A-18 Hornet, E/A-6 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). The rationale for the three categories is that each of these aircraft types have different crew configurations (TACAIR aircraft such as the F/A-18 Hornet are single seat aircraft, whereas the P-3 Orion has a crew of 11 personnel), and different missions (e.g. search and rescue versus transportation). There is also some research to suggest that, at least in the U.S. Air Force, CRM was more readily embraced by aviators flying crew aircraft than by fighter and attack aircraft pilots (Karp, Condit & Nullmeyer, 1999). Anecdotal evidence suggests that this effect of aircraft type may also be the case within the U.S. Navy. Oser, et al. (2001) stated that the funding was not made available to fully implement the CRM program across naval aviation, and there was considerable variation in the execution of the training across naval aviation communities.

Aviation officers were divided into two groups based upon rank: junior officers (officer grades O2 to O3, with a minimum of one tour as an aviator, who do the bulk of the flying in the squadron), and senior officers (O4 to O6, with multiple tours as an aviator, who perform the leadership roles in a squadron). The rationale for splitting junior and senior officers was that issues of rank of personnel may play a greater role in military flight crews than in civilian aviation, and this may be at odds with the assertiveness taught in CRM training (Guzzo & Dickson, 1996). This finding was supported in a survey of the attitudes of 272 U.S Navy divers to CRM concepts. It was found that junior divers were significantly more sensitive to the effect of personal limitations on performance, and showed a significantly greater willingness to want to speak up than senior divers (O'Connor, 2007).

The data used in this study was the entire population of survey responses. However, the analysis will proceed as if the responses constitute a random sample from a (hypothetical) population of responses that could have been seen if, for example, surveys had been taken at different times or under different conditions. Statements of statistical significance therefore refer to the parameters of this hypothetical population.

Reactions: Method

Procedure. The responses to the CRM item from the CSAS from 2000 until 2008 were obtained electronically from the company that administers the survey for the U.S. Navy. It is mandatory for all squadrons to complete the CSAS semiannually and within 30 days following a change of command (O'Connor et al., 2011).

Sample. A sample of 51,570 responses were obtained from naval aviators. A total of 67.7% were from junior officers, and 32.3% from senior officers. Overall, 37.4% of responses were from TACAIR, 36.3% rotary wing, and 26.3% from big wing aviation communities.

Table 1. Percentage of respondents separated by rank and aviation community.

Rank	Aviation community		
	Rotary	TACAIR	Big wing
Junior officer	26.5	25.5	15.7
Senior officer	9.8	11.9	10.7

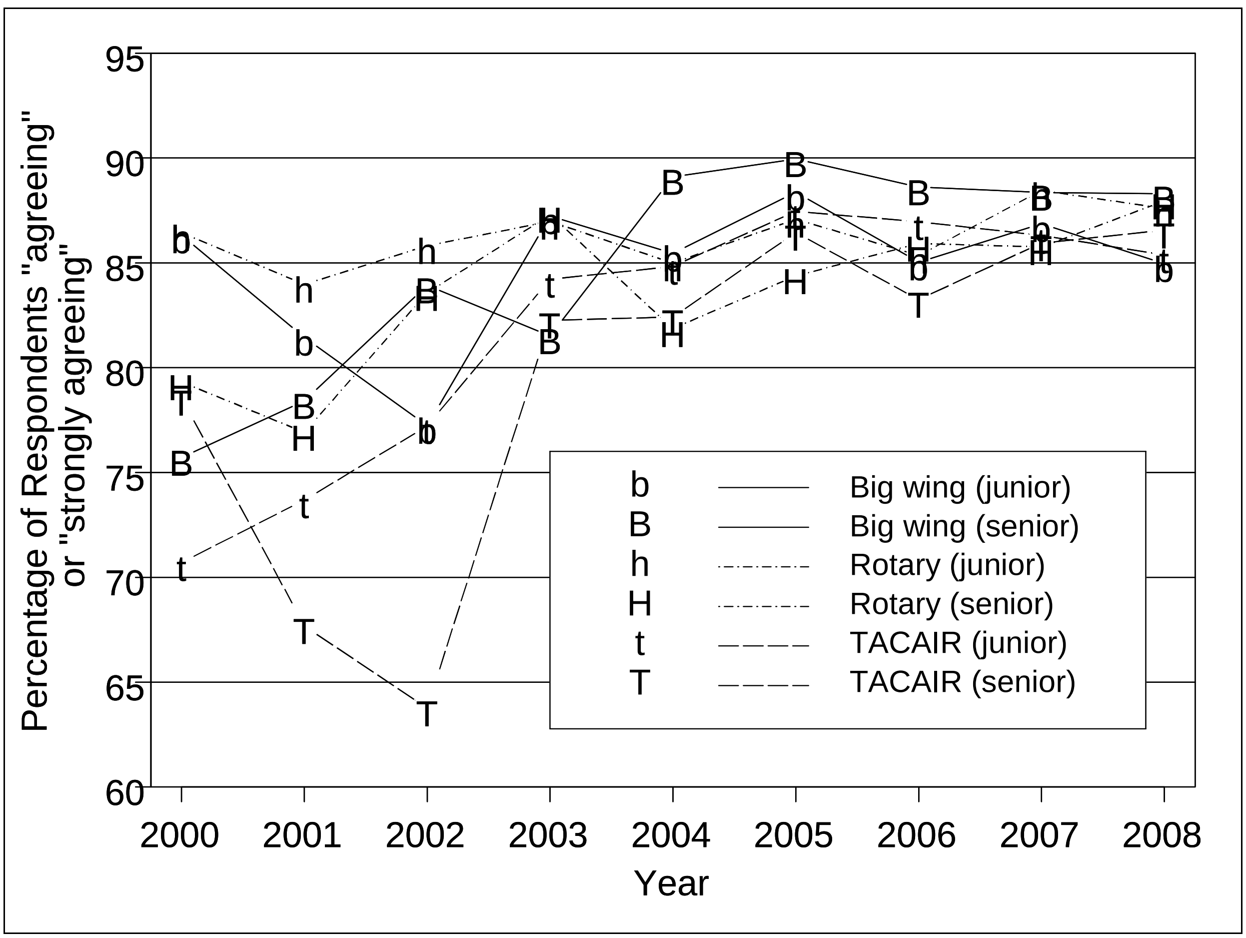
Analysis. A multinomial logistic regression model (Venables & Ripley, 2003) was fit to the response data. This model extends the ordinary binary logistic regression model to the case where the response has more than two categories. In this case, the response was the item (1 through 5) selected by the respondent. The predictors were year (considered as a categorical variable), rank and aviation community. The model then estimates the log-odds relative to the baseline (response 1) of the event that one of the other responses (2 through 5) is selected, with a separate estimation for each of those other responses. The model produces a set of predictions for each possible combination of predictor variables. Each of those predictions is an estimate of the probability of seeing a particular response when given that set of predictor variables. Since there are five possible responses, then, the model produces five predicted probabilities for each respondent, and those five probabilities add up to one.

Since this model is complicated and requires a large number of parameters, a second model used ordinary binomial logistic regression. In this model, the response was "positive" when a respondent chose either "4" (agree) or "5" (strongly agree), and "negative" for responses of "1" (strongly disagree), "2" (disagree), and "3" (neither agree nor disagree).

Reactions: Results

Across all ranks and aviation communities the average response rating was found to be positive (mean = 4.1, sd =0.7). About 83% of all responses were “4” or “5.” (A similar pattern held for other questions; across all responses to all questions, 85.3% were “4” or “5.”) The rate at which responses to the CRM item were “4” or “5,” separated by aviation community and rank, are shown in Figure 1.

Figure 1. The percentage of respondents ‘agreeing’ or ‘strongly agreeing’ that CRM is helping to improve mission performance and safety.



The Akaike Information Criterion (AIC; Venables & Ripley, 2003) was used as an aid in model selection. In both multinomial and binomial models, the model with the lowest (that is, best) AIC included all three two-way interactions but excluded the three-way interaction. Therefore, there is some evidence that the distribution of responses across the five categories is associated with rank and aviation community, and has changed across time in a way that is not the same as rank from the other aviation communities.

Reactions: Discussion

Overall it can be seen that naval aviators were positive about the usefulness of CRM training. However, it is important to indicate that this finding is based upon a response to a single item over which the researchers had no control. This is certainly a large limitation. Nevertheless, given the large number of responses, collected over eight years, the data provides an estimate of what naval aviators think about CRM training.

The positive reaction to CRM training of the non-TACAIR aviators across the years examined is consistent with the evaluations of reactions carried out by Salas and colleagues in the late 1990s. It is also worth noting that there has not been a decay in reactions over time. In one of the few longitudinal studies of CRM effectiveness, Irwin (1991) found a decay in positive attitudes of airline pilots over time.

The most interesting finding is that over the period of study, TACAIR aviators (particularly senior officers) have become more positive about CRM, and since 2003, had reactions more comparable with the other naval aviation communities. The lower reported levels of satisfaction with CRM training of TACAIR aviators during the early 2000s, as compared to aviators from other communities, is consistent with findings in the U.S. Air Force over a similar time period (Karp et al, 1999). The improvement in the reactions of TACAIR aviators is reflective of the work carried out during 2000 to tailor CRM training for the TACAIR aviation community, and demonstrates the benefit of allowing each aviation community to customize CRM training for their unique operational needs. Therefore, it can be concluded that the current naval CRM program is very well received across all naval aviation communities.

STUDY TWO: ASSESSMENT OF KNOWLEDGE OF CRM CONCEPTS

Knowledge: Introduction

According to Kirkpatrick (1996), in order to successfully gauge the amount of knowledge acquired, one must devise a method to measure the principles, facts and techniques learned from the program. Learning measures should be objective and quantifiable in nature, addressing aspects of declarative and procedural information retained as a result of the training. Declarative knowledge refers to the factual information that can be encoded and recalled from one’s memory. Procedural knowledge addresses applying a specific learned behavior to a situation or task (Holt, Boehm-Davis, & Beaubien, 2001).

The most commonly used technique to assess the knowledge acquired from CRM training is a multiple-choice test (O’Connor, et al., 2002). Three studies have examined the CRM knowledge gained by U.S. Naval aviators as a result of attending the training (e.g. Salas et al., 1999). Three of the four studies reported a significant increase in the knowledge of participants after training, as compared to before.

The current study will use a multiple-choice test to examine the knowledge of naval aviators towards the concepts addressed by CRM training, and assess whether there are differences between the knowledge of aviators from different communities, or between junior and senior aviators.

Knowledge: Method

Test development. A ten item multiple-choice test was developed to assess knowledge of the concepts taught in Naval aviation CRM. Questions for the test were derived from the material taught at the Naval CRM instructor course. The items addressed each of the seven skills that should be addressed in the Navy's CRM training program. The questionnaire was administered as a pilot test on ten naval aviators.

Procedure. The test was distributed to U.S. Navy and Marine Corps aviation officers attending aviation safety training at the Naval Aviation Safety School Command (NASC) in Pensacola, Florida, and studying for a Masters' degree at the Naval Postgraduate School (NPS) in Monterey, California during 2008 and 2009. The response rate was 73%.

Participants. A total of 172 responses were obtained from naval aviators (62% from NASC; and 32% from NPS).

Knowledge: Results

Item-total point biserial correlations were carried out for each of the 10 items in the knowledge test. As a result of this analysis, two items were discarded due to low correlations (less than 0.15). Table 2 summarizes the overall percentage of correct responses separated by rank and aviation community for the remaining eight items.

Table 2. Mean percentage of items correct separated by rank and aviation community.

Rank	Aviation community	Mean % correct	St. Dev.	#of responses
Junior	TACAIR	74	16	28
	Rotary	74	19	26
	Big wing	75	14	12
	Total	74	17	66
Senior	TACAIR	74	18	41
	Rotary	76	21	29
	Big wing	68	14	28
	Total	73	18	98
Total	TACAIR	74	18	69
	Rotary	75	20	55
	Big wing	70	14	40
	Total	73	18	164

A two way between subjects analysis of variance (ANOVA) was carried out to assess whether there were significant differences between rank (junior versus senior officers), or aviation community (TACAIR, rotary, or big wing). It was found that there were not significant main effects of rank ($F_{(1, 164)}$= 0.32, n.s.), aviation community ($F_{(2, 164)}$= 0.43, n.s.), or a significant interaction ($F_{(2, 164)}$= 0.71, n.s.). The two items for which the respondents were most commonly incorrect were: an item concerned with situation awareness (48% incorrect), and an item associated with decision making (48% incorrect).

Knowledge: Discussion

No differences were found between naval aviator's knowledge of CRM concepts on the basis of rank or type of aviation community. Despite the fact that declarative knowledge of CRM concepts is not routinely tested at naval squadrons, respondents performed reasonably well on the multiple-choice test, with an overall average of 73% correct. This average may not be considered reasonable within the context of higher education tests. However, the performance on the test by naval aviators represents the baseline CRM knowledge of participants who had not prepared to be evaluated, and had not just completed CRM training.

The low percentage of correct responses to the situation awareness and decision making items may be worthy of more detailed investigation. Given the disproportionate number of aviation mishaps that can be attributed to situation awareness and decision making errors, it is important that naval aviators have a thorough understanding of these concepts. It is recognized that, although typical of how knowledge of CRM concepts have been measured in the past, a multiple-choice test is a fairly coarse method for assessing the baseline CRM knowledge of aviators. Therefore, a more detailed analysis of knowledge of situation awareness and decision making concepts would be desirable.

STUDY 3: ASSESSMENT OF ATTITUDES TO CRM

Attitudes: Introduction

The most commonly used method for assessing aviator attitudes towards the concepts taught in CRM training is the Cockpit Management Attitudes Questionnaire (CMAQ). The CMAQ has formed the basis of a CRM attitude questionnaire in a number of industries (Flin, O'Connor, & Crichton, 2008). Five previous studies of naval aviators examined the effect of CRM training on the attitudes of course participants as compared to a control group. These studies used adaptations of the CMAQ to assess attitude change. Although significant changes in attitudes were not found, the studies generally reported a positive shift in the attitudes of CRM participants, (e.g. Salas et al., 1999). None of the studies reported an evaluation of the psychometric properties of the instrument that was used. In fact, this is true of the vast majority of the papers that have reported an evaluation of attitudes to CRM.

As with the evaluation of the knowledge of naval aviators reported earlier, comparisons were made based upon aviation community (TACAIR, rotary, or big wing), and rank (junior versus senior officers). An evaluation of the psychometric properties of the instrument was also carried out.

Attitudes: Method

Instrument development. The civilian aviation centric CMAQ questionnaire was adapted for naval aviation. It was necessary to change some of the language to ensure that it would make sense to naval aviators. A draft questionnaire was distributed to a group of 20 experienced naval aviators for comment. The comments from these aviators were used to develop the Naval Aviator Human Factors (NAHF) questionnaire. The NAHF consisted of 31 questions pertaining to five categories:

- My stress: 6 items. This scale emphasizes the consideration of- and possible compensation for- stressors in oneself.
- Stress of others: 6 items. This scale emphasizes the consideration of- and possible compensation for- stressors in other team members.
- Communication: 6 items. This scale encompasses communication of intent and plans, delegation of tasks and assignment of responsibilities, and the monitoring of crew members.
- Command responsibility: 9 items. Includes the notion of appropriate leadership and its implications for the delegation of tasks and responsibilities.
- Rules and order: 4 items. This subscale is concerned with adherence to rules and procedures.

The participants were asked to respond using a five-point Likert scale ranging from ‘strongly disagree’ to ‘strongly agree’. Although there were not explicit decision making or situation awareness categories, aspects of both of these CRM skills are addressed by a number of items (e.g. *‘my decision making is as good in emergencies as it is in normal situations’*, and ‘*junior aviators should not question the aircraft commander in emergencies*’).

Psychometric Analysis

1. Item analysis. The skewness, kurtosis, and correlation between the items pertaining to each factor were evaluated. Items with unacceptably high skewness or kurtosis were discarded from further analysis.-

2. Confirmatory factor analysis (CFA). A CFA technique as implemented by EQS for Windows was used to assess whether the a-priori model of the relationship between the items and factors was valid. CFA seeks to determine whether the number of factors, and the loadings of measured variables on them, conform to what is expected on the basis of pre-established theory. A linear structural relations approach to CFA, as implemented in EQS for Windows, was used. Tomas and Oliver (1999) recommend assessing model fit using the χ^2 statistic in association with, the Comparative Fit Index (CFI), the Goodness-of-Fit Index (GFI), and the Root Mean Square Error of Approximation (RMSEA).

The CFI is an incremental fit index that produces a statistic in the range of 0 to 1. The GFI is an absolute fit index. This means it directly assesses how well a model reproduces the sample data. A value of .9 is considered to be the minimum for model acceptance for both the CFI and GFI (Byrne, 2006). The RMSEA is computed based on sample size, the noncentrality parameter, and degrees of freedom for the target model (Browne & Cudeck, 1993). Models that are good descriptors of the data should produce RMSEA values of less than .05.

3. *Between group comparisons*. A comparison of the effects of rank, and aviation community on the factors was carried out use a multivariate analysis of variance (MANOVA).

Procedure. As with the knowledge assessment, the NAHF questionnaire was distributed to naval aviators attending training at two commands (NASC and NPS) during 2008 and 2009.

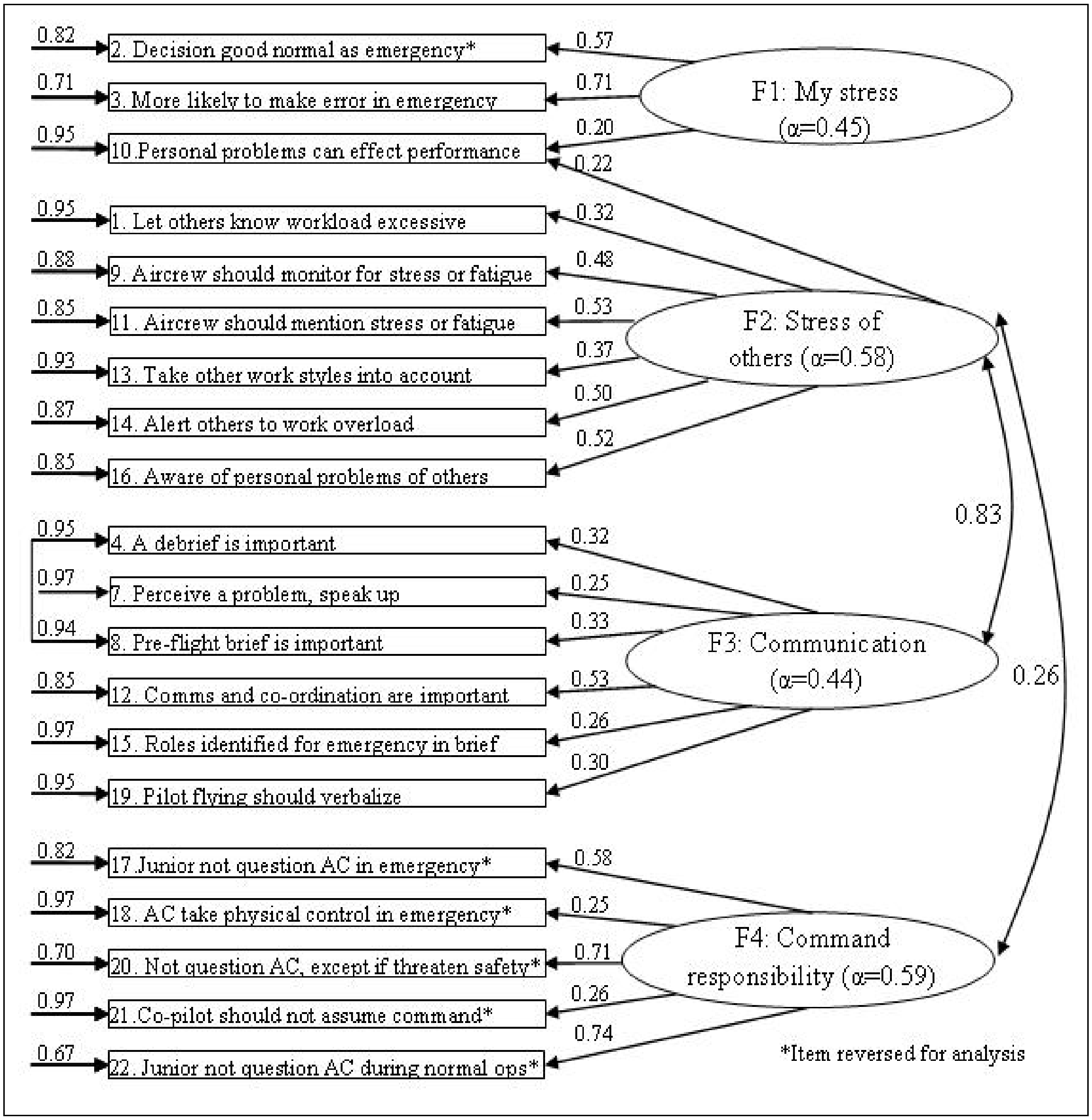
Participants. Responses were obtained from 526 U.S. naval aviators. A total of 72.7% were obtained from NASC students, and 27.3% from NPS. Separating by aviation community, 38% of participants were from TACAIR, 36% rotary wing, and 27% big wing communities. Examining the distribution of respondents by rank, 38% were junior officers, and 62% senior officers. The response rate was 78%.

Attitudes: Results

Psychometrics. The skewness and kurtosis of each item were assessed. Two items were dropped from further analysis from the ‘my stress’ factor, and three items from the ‘command responsibility’ factor due to excessive levels of skewness or kurtosis. The Cronbach’s alpha values of the proposed factors were also evaluated. Nunnally (1978) indicated 0.7 to be an acceptable reliability coefficient, but lower thresholds are sometimes quoted in the literature. It was found that the Cronbach’s alpha for the ‘rules and order’ factor (4 items) was unacceptably low (α=0), and could not be improved through discarding items. So, it was decided to drop this factor entirely from further analysis.

The 22 remaining items were entered into a CFA to test a measurement model comprising four factors (‘my stress’, six items; ‘stress of others’, four items; ‘communication’, six items; and ‘command responsibility’, six items). The fit was not found to be acceptable (χ^2 = 1431, df = 231, p>0.05; CFI = 0.75; GFI = 0.92; and RMSEA= 0.052). To achieve an acceptable fit, one item was dropped from the ‘my stress’ factor (*I am less effective when stressed or fatigued*), one item from the ‘communications’ and one item from the ‘command responsibility’ factor (*in abnormal situations, I rely on my superiors to tell me what to do*). The removal of these items was recommended by the Wald test, and resulted in an improved model fit. It was also necessary to load item 10 onto both factors 1 and 2 and allow covariance between factors two and three, and two and four, and one set of error terms (see figure 2). The final four-factor, 20-item model with the standardized solutions is shown in figure 3 and found to be an adequate fit for the data (χ^2 = 1321, df = 190, $p>0.05$; CFI = 0.91; GFI = 0.95; and RMSEA= 0.035).

Figure 2. Standardized solution for the four-factor model.



Between group comparison. As a result of the correlation between factors two and three, and two and four, a MANOVA was used to make comparisons between the two independent variables (rank and aviation community) for each factor. Consideration was given to using one of the refined methods for calculating the factor scores provided in EQS. However, it was decided to simply take the average of the item scores for each factor. The reason for choosing this method for calculating the factor scores was this was what was done in the previous studies of the CRM attitudes of naval aviators, non-refined factor scores are thought to be more stable across samples than refined methods (Grice & Harris, 1998), and this method is desirable when reliability or validity has not been established (Hair, Black, Babin, Anderson, & Tatham, 2006). The MANOVA did not result in any significant differences on any of the factors for aviation community or rank.

Attitudes: Discussion

Discussion: psychometric properties. Due to excessive skewness or kurtosis, and poor factor reliability, a total of nine items were discarded from the NAHF questionnaire. Two more items were discarded as part of the CFA process. The practice of discarding items is common in questionnaire development (DeVellis, 1991). As part of the post-hoc analyses changes were made to the original model. Byrne (2006) states that if the model is respecified in this manner, it is important to realize that the analysis is now framed in an exploratory rather than confirmatory mode. Consideration was also given to additional post-hoc changes to the model to improve the fit. However, given the lack of a specific model misfit, and to avoid overfitting the model, no further changes were made.

One issue that should be addressed is the relatively low Cronbach's Alphas for the factors. However, the values found are typical of this type of questionnaire. For example, in the CMAQ, Cronbach's Alphas between 0.47 and 0.67 were reported (Gregorich, Helmreich & Wilhelm, 1990). Nevertheless, it is suggested that there is a need to further restrict the focus of the items in each factor with the aim of improving the reliability of the factors.

Discussion: Groups comparison. For both the 'communication' and 'command responsibility' factors, naval aviators across communities and ranks displayed attitudes that were very favorable to these concepts. These findings are consistent with the reaction data reported in

Figure 1 during the 2007 to 2008 time frame. Naval aviators recognize the importance of direct and explicit communication to mission effectiveness, and the need for assertiveness, even with more senior personnel. The strong covariance between the 'communication' and 'stress of others' factors in the CFA model may reflect that aviators can only find out about the stress of others through communication. This finding is very encouraging because studies of communication in aviation research have highlighted the need for assertive behavior in more junior team members (e.g. Jentsch & Smith-Jentsch, 2001).

The factor scores for the two factors related to 'my stress' were not as reflective of the teachings of CRM training as the scores on the other three factors. Naval aviators appear to have attitudes that suggested that they are just as effective in an emergency as in normal operations, even though they recognize this is not the case in others. This finding is not surprising when viewed in the context of the small literature on stress coping in military aviation. After reviewing ten papers on stress in military aviation, Campbell and O'Connor (2010) concluded that military aviators appear to use an avoidance stress coping style. When faced with a stressful situation, military aviators will tend to use strategies such as denial, distraction, repression, and suppression to deal with stress. These are not coping strategies that are trained, rather they represent the coping mechanisms that are reported in the literature as being utilized by military aviators. Therefore, despite responses generally reflective of a positive attitude to CRM training, it is suggested that training should place more emphasis on the effects of stress on performance.

STUDY 4: AN ASSESSMENT OF THE IMPACT OF CRM ON MISHAP RATES

Organization: Introduction

The ultimate aim of any CRM training program is to produce tangible evidence of an effect on the organization such as a decreased mishap rate. There have been a number of studies that have examined the effect of CRM on the mishap rate in naval aviation, with mixed findings. To illustrate, Wiegmann and Shappell (1999) analyzed the causes of naval aviation mishaps from 1990 to 1996 in the rotary and TACAIR communities. It was found that 56% of the mishaps involved at least one CRM failure, with a higher percentage of CRM related mishaps in the rotary wing community as compared to TACAIR (80.4% and 46.5% respectively). Wiegmann and Shappell (1999) compared their findings to an earlier evaluation of naval aviation mishaps. Yacavone (1993) found that 45% of major aviation mishaps from 1986 to 1990 were attributed to a failure in CRM. Wiegmann and Shappell (1999) attributed the lack of change in the aircrew error rate to the failure of naval aviation to specifically tailor the CRM program to needs of the different aviation communities.

The aim of the current study was to analyze a more recent sample of naval aviation mishaps. Big wing mishaps were not included in the analysis as these are very rare in naval aviation, so do not serve as a useful metric of performance.

Organization: Method

Procedure. Data were obtained from the Naval Safety center on all class ‘A’ mishaps for rotary and TACAIR mishaps from fiscal year 1997 until 2007 (fiscal years begin in October of each year). During this time period a class ‘A’ aviation mishap was defined as a mishap in which the total cost of damage is greater than a million dollars, and/or the aircraft is destroyed, and/or fatal and/or permanent total disability (a non-fatal injury that permanently incapacitates a person so that he or she cannot follow any gainful occupation; Chief of Naval Operations, 2001). In many aviation mishaps the aircraft is destroyed. Data were obtained on both the overall mishap rate, and the rate for mishaps in which there was at least one causal factor attributed to a failure of CRM.

Sample. Data were obtained on a 238 class ‘A’ mishaps (58.8% TACAIR and 41.2% rotary wing mishaps).

Analysis. The Poisson distribution is frequently used to model accident data (Nicholson & Wong, 1993). The analysis in this section, then, starts by assuming that mishaps appear in each year as if they were generated by a Poisson process. If this assumption is correct, and if the underlying mishap rate (and numbers of flights and flight hours) were constant from year to year, the observed yearly numbers of mishaps would be similar in magnitude, with some variability attributable to the behavior of the Poisson process. Conversely, observing yearly numbers of mishaps that were very much more spread out than would be expected under the Poisson would constitute evidence of year-to-year variability in the mishap rate.

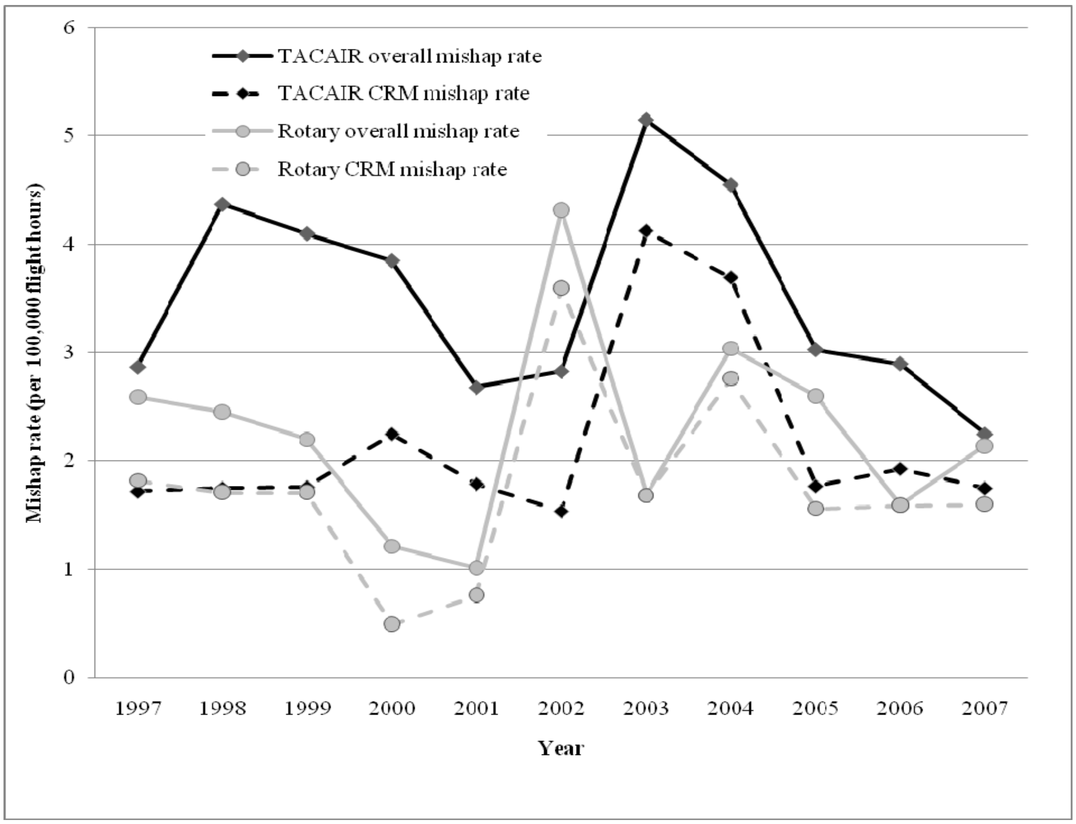
The year-to-year spread of the numbers of mishaps is measured using the standard deviation. The null hypothesis is that the underlying Poisson process has a constant rate from year to year, so that the expected value of the standard deviation is small. The distribution of the standard deviation of the mishap numbers when the null hypothesis is true is made more complicated by the fact that the number of flight hours is not constant from year to year, but it can be easily simulated. The overall mishap rate (that is, total events divided by total flight hours) was computed for each data set (this analysis was performed separately twice, one using all mishaps and once using only CRM-related ones). Under the null hypothesis of homogeneity, it is reasonable to combine all the years’ data into a single overall rate. That rate was then used together with the observed number of flight hours per year, to compute the number of mishaps per year that would be expected under the null hypothesis.

Then Poisson random numbers, one per year, each with the expected value computed above, were generated, and the standard deviation of those random numbers was computed. This process was repeated 10,000 times and the resulting 10,000 standard deviations provided the mull distribution against which the observed standard deviation could be compared. The hypothesis of homogeneity is rejected if the observed standard deviation is larger than all but 5% of the simulated standard deviations.

Organization: Results

A total of 68.9% of the 238 mishaps were determined to have involved at least one CRM failure (62.9% for TACAIR, and 77.6% for rotary wing). Figure 3 outlines the overall and CRM mishap rates for the TACAIR and rotary wing communities.

Figure 3. Overall, and CRM related, mishap rate for TACAIR and rotary aircraft.



The hypothesis of homogeneity was not rejected for the overall, or CRM-related, mishap rates for either aviation community. The *p*-values from the simulation for TACAIR were 0.63 for the overall mishap rate and 0.22 for the CRM-related mishap rate. For rotary-wing aircraft, the *p*-values were 0.11 for the overall mishap rate, and 0.09 for the CRM-related one.

Organization: Discussion

A significant difference in the overall, or CRM related, mishap rate was not found across the 11 years that were examined. Therefore, the hypothesis of homogeneity was not rejected. Another finding is that the overall CRM related mishap rate of 68.9% is higher than the 56.3 % reported by Wiegmann and Shappell (1999) in their examination of naval aviation mishaps from 1990 to 1996. Therefore, do we conclude that, despite the belief of naval aviators that CRM training is useful, the fact that aviators are knowledgeable of, and displayed positive attitudes towards, the concepts addressed in the training that the impact of CRM on the mishap rates has actually decreased?

The main issue is that the mishap rate is clearly affected by many factors other than CRM training. To illustrate, even though combat related mishaps are not included in the analysis, the wars in Afghanistan and Iraq certainly influenced the type of flight training that was being carried out, particularly during the early 2000s. Moreover, there have also be changes in the aircraft themselves overtime, with the introduction of advanced next generation aircraft, and aviators being retrained to fly these aircraft.

CONCLUSION

The lack of a control group and the absence of longitudinal data on the attitudes towards, and knowledge of, the concepts addressed in CRM training are limitations of this study. A further weakness is the lack of an assessment of behavior.

Without an assessment of behavior it is not possible to ascertain whether the skills taught on the CRM program have transferred to the aircraft. It may be that the training is very effective in providing naval aviators with background knowledge on CRM. However, unless aviators are provided with clear examples of ‘good’ and ‘poor’ CRM skills, and provided the opportunity to practice them, these skills may not be used in the aircraft.

Five studies examining the CRM behaviors of naval aviators have been reported in the literature. All of these studies reported an increase in the CRM behaviors of naval aviators immediately after CRM training (e.g. Salas et al, 1999). To encourage more detailed evaluation of specific CRM behaviors, it is suggested that U.S. naval aviation adopt a strategy similar to that of European civil aviation. In Europe, the civil aviation regulator provided an exemplar behavioral marker system called NOTECHS (NOn-TECHnical Skills) to guide operators on the CRM behaviors that should be assessed, and provides guidance for how to do so. NOTECHS has been shown to have good levels of acceptance, reliability and validity (see Flin et al, 2008, for more details). Operators are not mandated to use NOTECHS, rather it can be adapted for the needs of a particular company. Providing naval aviation communities with a framework for evaluating behavior would allow an assessment to be made as to whether aviators are using CRM skills effectively and, more importantly, give them a mechanism for de-briefing aviators on their CRM behaviors.

Despite the limitations, the evaluations reported in this paper allow a number of conclusions to be made about the effectiveness of a mature CRM training program- something that has rarely been reported in the literature. The studies showed that naval aviators believe that

CRM training has a positive effect on safety and performance, something that TACAIR aviators (particularly senior officers) did not agree with as strongly in the early 2000s as compared to the late 2000s. Naval aviators who fly different aircraft and of different ranks were generally knowledgeable of, and display positive attitudes towards, the concepts addressed in CRM training. These findings are important given that there is evidence to suggest that in the 1990s fighter pilots (at least in the U.S. Air Force, Karp et al, 1999), and a proportion of commercial pilots were resistant to the concepts addressed in CRM training (Helmreich & Wilhelm, 1991; Irwin, 1991). Finally, the lack of a significant effect of CRM on the mishap rate is also noteworthy. The mishap rate tends to be where senior leadership are most focused in looking for evidence of the effectiveness of safety programs. This focus has implications for the money and resources provided to safety programs. However, the final study in this paper shows that, even with considerable resources, it may not be possible to show a significant effect of a safety program such as CRM on the mishap rate.

As the next generation of highly automated aircraft becomes part of naval aviation, there is a need to ensure that CRM training evolves to meet the challenges of flying these new machines. It is only through the evaluation of reactions, knowledge, attitudes, behaviors, and mishap rates that a complete understanding of the usefulness of CRM training can be achieved. The availability of this information will allow senior leadership to make informed decisions to ensure that naval aviators receive effective CRM training.

ACKNOWLEDGEMENTS

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.

REFERENCES

Alkov, R. A., & Gaynor, J. A. (1991). Attitude changes in Navy/Marine flight instructors following aircrew coordination training course. *The International Journal of Aviation Psychology, 1(3)*, 245-253.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Byrne, B.M. (2006). *Structural Equation Modeling with EQS.* Mahwah, NJ: Lawrence Erlbaum.

Campbell, J. & O'Connor, P. (2010) Coping with stress in military aviation: A review of the research. In: P. O'Connor & J. Cohn (Eds.). *Human Performance Enhancements in High-Risk Environments: Insights Developments, and Future Directions from Military Research* (pp.169-188). Santa Barbara, CA: ABC- Clio.

Chief of Naval Operations (2001). *Crew Resource Management Program, OPNAVINST 1542.7C.* Washington DC: Author.

DeVellis, R. F. (1991). *Scale Development: Theory and Applications*. London: SAGE Publications.

Federal Aviation Administration (2004). *Advisory Circular No 120-51E: Crew Resource Management Training*. Washington, DC: Author.

Flin, R., O'Connor, P., & Crichton, M. (2008). *Safety at the Sharp End: A Guide to Nontechnical skills.* Aldershot, UK: Ashgate.

Gregorich, S., Helmreich, R., & Wilhelm, J. (1990). The structure of cockpit management attitudes. *Journal of Applied Psychology, 75(6)*, 682-690.

Grice, J. W. & Harris, R. J. (1998). A comparison of regression and loading weights for the computation of factor scores. *Multivariate Behavioral Research, 33*(2), 221-247.

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology,47,* 307-338.

Hair, J.F. Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. (2006). *Multivariate Data Analysis*. Upper Saddle River, NJ: Pearson Educational Inc.

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *International Journal of Aviation Psychology, 9*, 19-32.

Helmreich, R. L., & Wilhelm, J. A. (1991). Outcomes of Crew Resource Management training. *International Journal of Aviation Psychology, 1*, 287-300.

Holt, R. W., Boehm-Davis, D., & Beaubien, J.M. (2001). Evaluating resource management training. In E. Salas & C. A. Bowers & E. Edens (Eds.), *Improving teamwork in organizations: Applications of resource management training* (pp. 165-190). Mahwah, NJ: Lawrence Erlbaum Associates.

Irwin, C. M. (1991). *The impact of initial and recurrent cockpit resource management training on attitudes.* In R. Jensen (ed.) *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 344-349). OH: Ohio State University.

Jentsch, F. & Smith-Jentsch, K. (2001) Assertiveness and team performance: More than “Just Say No”. In E. Salas, C. Bowers & E.Edens (Eds.) *Improving Teamwork in Organisations.* (pp. 73-92).Mahwah, NJ: Lawrence Erlbaum.

Karp, R.M., Condit, D. & Nullmeyer, R. (1999). Cockpit/Crew Resource Management for Single-Seat Fighter Pilots. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference [CD-ROM]. Arlington VA: National Training Systems Association.*

Kirkpatrick, D. L. (1996). Evaluation. In R. L. Craig, *The ASTD Training and Development Handbook* (pp. 294-312). New York, NY: McGraw-Hill.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig & L. R. Bittel (Eds.), *Training and development handbook* (pp. 18.1-18.27). New York: McGraw Hill.

Nicholson, A., & Wong, Y.D. (1993). Are accidents Poisson distributed? A statistical test. *Accident Analysis and Prevention*, *26(3)*, 407–408.

Nullmeyer, R., Spiker, V.A., Wilson, D., & Deen, G. (2003). Key crew resource management behaviors underlying C-130 aircrew performance. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference* [CD-ROM]. Arlignton VA: National Training Systems Association.

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

O'Connor, P. (2007). The nontechnical causes of diving accidents: Can U.S. Navy divers learn from other industries? *Journal of the Undersea and Hyperbaric Medical Society*. *34(1),* 51-59.

O'Connor, P., Campbell, J., Newon, J., Melton, J., Salas, E., & Wilson, K. (2008). Crew resource management training effectiveness: A meta-analysis and some critical needs. *International Journal of Aviation Psychology, 18(4),* 353-368.

O'Connor, P. & Jones, D. (2009). *The crew resource management attitudes of U.S. Naval aviators*. Paper presented at the Human Factors and Ergonomics Society conference. San Antonio, TX.

O'Connor, P., Flin, R. & Fletcher, G. (2002). Methods used to evaluate the effectiveness of CRM training: A literature review. *Journal of Human Factors and Aerospace Safety, 2,* 217-234.

O'Connor, P., Hahn, R., & Nullemeyer, R. (2010). The military perspective. In: R.L. Helmreich, B.G. Kanki, & J. Anca (Eds.). *Crew Resource Management*. (pp. 445-468). San Diego. CA: Elsevier.

O'Connor, P., Hahn, R. & Salas, E. (2010). The U.S. Navy's Crew Resource Management program: the past, present, and recommendations for the future. In: P. O'Connor & J. Cohn (Eds.).*Human Performance Enhancements in High-Risk Environments: Insights Developments, and Future Directions from Military Research* (pp. 90-105). Santa Barbara, CA: ABC- Clio.

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.

Oser, R. L., Salas, E., Merket, D. C., & Bowers, C. A. (2001). Applying resource management training in naval aviation: A methodology and lessons learned. In E. Salas & C. A. Bowers & E. Edens (Eds.), *Improving teamwork in organizations: Applications of resource management training* (pp. 283-301). Mahwah, NJ: Lawrence Erlbaum Associates.

Prince, C., & Salas, E. (1993). Training and research for teamwork in the military aircrew. In E. Wiener, B. Kanki & R. Helmreich (Eds.), *Cockpit Resource Management* (pp. 337-366). San Diego: Academic Press.

Salas, E., Burke, C. S., Bowers, C. A., & Wilson, K. A. (2001). Team training in the skies: Does crew resource management (CRM) training work? *Human Factors*, *41*, 161-172.

Salas, E., Wilson, K.A, Burke, C.S., & Wightman, D.C. (2006). Does CRM training work? An update, extension and some critical needs. *Human Factors, 14*, 392-412.

Salas, E., Fowlkes, J. E., Stout, R. J., Milanovich, D. M., & Prince, C. (1999). Does CRM training improve teamwork skills in the cockpit?: Two evaluation studies. *Human Factors*, *41*, 326-343.

Stout, R. J., Salas, E., & Kraiger, K. (1996). The role of trainee knowledge structures in aviation psychology. *The International Journal of Aviation Psychology*, *7*, 235-250.

Tomas, J. M., & Oliver, A. M. (1999). Rosenberg's self-esteem scale: two factors or method effects. *Structural Equation Modeling*, *6*(1), 84–98.

Venables, W.N. & Ripley, B.D. (2003). Modern Applied Statistics with S-PLUS. New York, NY: Springer Science.

Wiegmann, D., & Shappell, S. (1999). Human error and crew resource management failures in naval aviation mishaps: A review of U.S. Naval Safety Center data, 1990-96. *Aviation, Space, Environmental Medicine, 70 (12)*, 1147-1151.

Yacavone, D. (1993). Mishap trends and cause factors in naval aviation: a review of Naval Safety Center data, 1990-1996. *Aviation, Space, Environmental Medicine,64* (5), 392-395.