Team Training in the Skies:

Does Crew Resource Management (CRM) Training Work?

_Eduardo Salas_
_C. Shawn Burke_
_Clint A. Bowers_
_Katherine A. Wilson_

University of Central Florida
Orlando, FL, USA

Key Words: crew resource management, teamwork, aviation, team training, training evaluation, multi-level evaluation, safety

Shortened Title: Team Training in the Skies

The aviation community has invested great amounts of money and effort into crew resource management (CRM) training. Using Kirkpatrick’s (1976) framework for evaluating training, we reviewed 58 published accounts of CRM training to determine its effectiveness within aviation. Results indicated that CRM training generally produced positive reactions, enhanced learning, and desired behavioral changes. However, we cannot ascertain whether CRM has an effect on an organization’s bottom line (i.e., safety). We discuss the state of the literature with regard to evaluation of CRM training programs and, as a result, call for the need to conduct systematic, multi-level evaluation efforts that will show the true effectiveness of CRM training.

Address correspondence to: Dr. Eduardo Salas
Department of Psychology
University of Central Florida
P.O. Box 161390
Orlando, FL 32816-1350
(407) 823-2552 (w); 823-5862 (f) x
esalas@pegasus.cc.ucf.edu
Team Training In The Skies:  
Does Crew Resource Management (CRM) Training Work?

It is well acknowledged that 60-80% of accidents and mishaps occurring in aviation have been attributed to human error (Freeman & Simmon, 1991). A large part of these are due to failures in coordination among cockpit crews. For example, poor pilot performance and faulty crew resource management (CRM) have been cited as contributing factors in numerous accidents and incidents reported by major airlines during the time period covering 1983-1985 (U.S. GAO, 1997). In addition, CRM deficiencies (e.g., lack of coordination among cockpit crews, captain’s failure to assign tasks to other members, and a lack of effective crew supervision) were a contributing cause in approximately ½ of the above reported accidents that involved one or more fatalities (U.S. GAO, 1997). Other reviews have found similar factors at work within cited accident reports (see Leedom & Simon, 1995; Chidester, Helmreich, Gregorich, & Geis, 1991; Gregorich, Helmreich, & Wilhelm, 1990).

Within the aviation environment, teamwork deficiencies are not only embarrassing and highly publicized, but can lead to tragic consequences. For example, Eastern Airlines Flight 401 crashed in the Florida Everglades in December 1972, because the crew permitted their fully operational Lockheed L-1011 to fly into the ground. What the crew failed to realize was that the altitude hold feature of the autopilot had been accidentally disconnected (as cited in Kayten, 1993). Results of the investigation revealed that the entire three-person crew was pre-occupied with a landing gear light that had failed to illuminate at the time of the accident. Many other aviation accidents resulting in disastrous consequences have also been attributed to faulty CRM skills (Allegheny Airlines, 1971; 1978; Mohawk Airlines, 1972; United Airlines, 1978; and others as cited in Kayten, 1993; U.S. GAO, 1997).

In an effort to manage some of these problems with teamwork and the resulting safety issues, the aviation industry introduced the concept of CRM (Wiener, Kanki, & Helmreich, 1993; Salas, Bowers, & Edens, 2001). CRM was introduced as a way to train aircrews to use all available resources—equipment, people, and information—by communicating and coordinating as a team. At this point in time, CRM has been used within the aviation industry for over 20 years, and has undergone several evolutions with varying foci (Helmreich, Merritt, & Wilhelm, 1999; Helmreich & Foushee, 1993; Maurino, 1999). Specifically, during the first evolution, the emphasis was on changing individual styles and correcting deficiencies in individual behavior, with a heavy focus on psychological testing. The second evolution represented a focus on cockpit group dynamics, was more modular, and dealt more with specific aviation concepts related to flight operations. With the third evolution came a broadening of scope. Specifically, training began to recognize the characteristics of aviation systems in which crews must function, as well as expanding to areas outside the cockpit (e.g., cabin crews, maintenance personnel). With the fourth generation came integration and proceduralization. Specifically, under the Advanced Qualification Program (AQP), carriers were allowed to tailor training to fit the needs of their specific organization, they were required to provide both CRM and line-orientated flight training (LOFT) to all crews, and CRM training was integrated with technical training. The fifth and latest evolution represents an awareness that human error is inevitable and can provide a great
deal of information. CRM is now being used as a way to try and manage these errors by focusing on training teamwork skills that will promote: (1) error avoidance, (2) early detection of errors, and (3) minimization of consequences resulting from CRM errors. Programs are beginning to go beyond error management to include a focus on threat recognition and management.

The evolutions that CRM has witnessed have occurred over roughly two decades of use within the aviation community; research during this time has produced several lessons learned (see Salas, Bowers, & Edens, 2001). For example, research has yielded information about how to maximize the design and delivery of CRM training through scenario design (Prince, Oser, Salas, & Woodruff, 1993; Prince & Salas, 1999), scenario feedback (Salas, Rhodenizer, & Bowers, 2000; Prince, Brannick, Prince, & Salas, 1997), and the training of operational personnel as observers and raters (Brannick, Salas, & Prince, 1997). In this vein, Helmreich and Wilhelm (1987) found that the systematic training of raters in CRM concepts made a significant difference in the quality of ratings and scale use (e.g., raters using entire scale) as compared to raters who were trained less systematically (as cited in Helmreich, Chidester, Foushee, Gregorich, & Wilhelm, 1990). Evidence has also been provided that suggests that low fidelity simulations can be used to practice/train CRM-related skills (Bowers, Salas, Prince, & Brannick, 1992; Baker, Prince, Shrestha, Oser, & Salas, 1993; Jentsch & Bowers, 1998). Furthermore, we have learned that national culture plays a powerful role in determining the effectiveness of CRM training programs (Maurino, 1994; Merritt & Helmreich, 1995b). Specifically, we have learned that attitudes that define the core concepts of CRM differ dramatically across national borders (e.g., individualism/collectivism, power distance, uncertainty avoidance, and division of roles between sexes, see Hofstede, 1988). As such, initial attempts to apply CRM globally were often unsuccessful due to a failure to recognize the power of national culture (Helmreich, Wilhelm, Klinect, & Merritt, in press). Finally, we know that anecdotal evidence, as well as reactions to CRM training, generally suggests that CRM training can prevent accidents (see Diehl, 1991; Kayten, 1993), and that CRM is being applied in domains outside aviation (see Flin, 1995; Howard, Gaba, Fish, Yang, & Sarnquist, 1992; Merritt & Helmreich, 1995a).

Although positive lessons have been learned, there are areas in need of improvement. To begin with, we know that, despite its long history, there remains a lack of consistency within the aviation industry with regard to definitions of CRM, training content, and methods of delivery (Wilhelm, 1991; Helmreich & Wilhelm, 1987; Salas, Prince, Bowers, Stout, Oser, & Cannon-Bowers, 1999). Second, we know that with the development of the Advanced Qualification Program (AQP) guidelines, each individual airline now has two choices in deciding how they want to implement CRM training: (1) traditional requirements as mandated by Federal Aviation Regulation (FAR) part 121, or (2) by using the AQP guidelines. FAR part 121 states the general operating requirements for domestic, flag, and supplemental operations, and contains the general requirements for CRM training. However, it leaves methods for CRM curriculum design and development as well as for evaluation ambiguous. In an effort to resolve the lack of guidance offered by the FAA regarding CRM training, the AQP guidelines were created. Under the AQP guidelines, airlines are provided with: (1) a process for developing the curriculum in order to integrate traditional CRM programs with technical training, (2) the required level of performance that must be achieved, and (3) guidelines that provide
inspectors with evaluation criteria to determine if the curriculum meets all FAA requirements (U.S. GAO, 1997; Helmreich et al., 1999). Despite the creation of AQP guidelines, airlines are not required to use them in replacement of FAR 121 requirements – therefore, there is still much ambiguity and a lack of consensus when it comes to the design, delivery, and evaluation of CRM training programs.

Although this ambiguity is surrounding the design (i.e., content) and evaluation of these programs, the aviation community continues to invest millions of dollars into CRM training and other communities (e.g., medical, offshore oil, maritime shipping companies) are now starting to “jump on the bandwagon”. Given the increased adoption of CRM as a worthwhile team training approach, there is a need to summarize the current state of knowledge about the effectiveness of CRM training in a systematic manner. That is, we need to assess the effectiveness of the implemented team training systems. Therefore, the purpose of this paper is to use Kirkpatrick’s (1976) typology for training evaluation, as a framework to evaluate the effectiveness of CRM training programs in aviation. Specifically, the review is organized via the type of evidence collected after training (i.e., reaction, learning, behaviors, and/or organizational effectiveness).

THE NEED FOR EVALUATION

Training evaluation has been defined as “the systematic collection of descriptive and judgmental information necessary to make effective training decisions related to the selection, adoption, value, and modification of various instructional activities” (Goldstein, 1993, p. 147). Although it is acknowledged that systematic training evaluation is not an easy task, it is the only way to ensure that training programs are having the desired effect and are a worthwhile investment for the organization. Similarly, Cannon-Bowers et al. (1989) have argued that training evaluation may serve a number of important functions. First and most obvious, program evaluation results can indicate whether the goals and objectives of a program are appropriate to achieve the desired outcome. Second, evaluation can indicate whether the content and methods used in training will result in achievement of the overall program goal. Third, evaluation data can be used to determine how to maximize transfer of training. Fourth, it can serve as feedback at both the individual and team level to suggest areas in need of improvement or revision. Goldstein (1993) offers similar arguments as to the benefits of evaluation.

The most popular framework for guiding training evaluations is Kirkpatrick’s (1976) typology. Kirkpatrick argued for a multi-level approach to training evaluation consisting of four levels of evaluation: (1) reactions, (2) learning, (3) behavior (i.e., extent of performance change), and (4) results (i.e., degree of impact on organizational effectiveness or mission success). Within recent years, this typology has been expanded by several researchers (see Kraiger, Ford, & Salas, 1993; Salas & Cannon-Bowers, 2001). For example, Kraiger et al. (1993) expanded Kirkpatrick’s typology by arguing that learning is multi-dimensional and results in cognitive, affective, and skill-based-learning outcomes. Moreover, Kraiger et al. suggest potential methods that can be used to evaluate each of these outcomes: (1) cognitive (verbal knowledge, knowledge organization, cognitive strategies), (2) affective (attitudinal, motivational), and (3) skill-based (compilation, automaticity). Goldsmith and Kraiger (1997) have built upon this work by describing a method for the structural assessment of an individual learner’s knowledge and skill, which has been successfully used in aviation research efforts (see Kraiger, Salas, & Cannon-Bowers, 1995; Stout, Salas, & Kraiger, 1997). The utilization
of Kirkpatrick’s typology and corresponding revisions, serves several important functions within the training evaluation process. First, it has served to organize the type of information that should be collected in the assessment of training. Second, it has served to argue for the added benefit/importance of collecting more than one level of evaluation information. Although both points perform important functions, the second has been more difficult to put into practice than the first. Specifically, Alliger and Janak (1989) reported that less than 10% of organizations assess training programs at all four levels of evaluation, as argued for by Kirkpatrick.

HEEDING THE CALL: EVALUATION OF CRM TRAINING

Our review resulted in the identification of 58 studies that appeared to evaluate the effectiveness of aviation CRM training programs. We next provide a description of the state of CRM evaluation efforts with respect to each of the levels of evaluation as identified by Kirkpatrick (1976). Specifically, studies that assessed training at only one level will be reviewed first, beginning with those collecting reaction data and ending with those collecting results/organizational effectiveness data. Following this will be a brief review of studies that assessed training at multiple levels, as argued for by Kirkpatrick.

For a summary of individual studies see Table 1, which describes the findings of each of the identified studies in relation to reactions, learning, behavioral, and organizational effectiveness data.

Do Aviators Like CRM?: Reaction Evidence

Reaction evidence is the first level of Kirkpatrick’s (1976) typology and amounts to an assessment of trainee’s feelings toward the training program. Reaction data are assessed post training and examine the degree to which participants perceive that training was worthwhile, relevant, interesting, and/or well conducted. Reaction evidence is perhaps the easiest to collect and usually takes the form of a paper-and-pencil questionnaire where the response format is typically a Likert scale. A small sample of studies have also gathered reaction data through the utilization of questionnaires that ask participants to rank order the perceived usefulness of the CRM components included in training.

After reviewing the available literature, reaction data is a commonly collected type of evaluation data. Specifically, of the studies included in our review, 27 of 58 (46%) involved the collection of reaction data (see Table 1). Of these 27 studies, 9 collected information solely related to participant reactions. Furthermore, the reaction data that was collected in the reviewed studies tended to reflect both affective feelings towards CRM programs, as well as the utility of these programs. Alliger, Tannenbaum, Bennett, and Traver (1997) argue that “liking of training” is the most common form of training assessment, and the results from the current review tend to support this argument in that 12 of the 27 studies (44%) assessed participants’ affective reactions towards training. Overall, the results of these studies suggest that most participants like CRM training. In addition to assessing the overall affective reactions to training, some studies (e.g., Baker et al., 1993; Hornan, Goeters, Maschke, & Schiewe, 1995) also assessed affective reaction to particular components of CRM training. For example, Schiewe (1995) found that units that were based on case studies or used role play were very well liked by participants, while those based mostly on lecture were not rated favorably. Others have found similar findings (see Baker, Bauman, & Zalesny, 1991), suggesting
that perhaps methods that promote interaction among participants are liked better than those that are more passive.

Reactions as to how participants “liked” training, are not the only type of reactions that may be collected from participants. Alliger et al. (1997) also suggest assessing participant reactions to the utility of training. Questions related to the assessment of utility, “…attempt to ascertain the perceived utility value, or usefulness, of training for subsequent job performance” (p. 344). Nine of the 27 studies (33%) assessed the utility of training, while the remaining 7 studies assessed both affective and utility reactions. In terms of utility, the reviewed CRM training programs were seen to be worthwhile, useful, and applicable. Specifically, themes were seen as relevant (Grau & Valot, 1997; Horman et al., 1995) and participants felt that CRM class should be expanded to other fleets/populations (Incalcaterra & Holt, 1999). See Table 1 for further information.

Positive reactions to CRM (affective, utility) were found to exist in single airline-studies, as well as in multi-fleet and multi-airline studies (see Butler, 1993; Helmreich & Wilhelm, 1991). Furthermore, the teaching of teamwork behaviors (e.g., communication, decision-making, leadership [see Alkov, 1991; Alkov & Gaynor, 1991]), use of role-playing exercises (see Baker et al., 1991), and inclusion of cabin crewmembers in training (see Vandermark, 1991) have all contributed to obtaining positive affective and utility reactions in training. Although the aviation community should be applauded for beginning to assess both affective reactions to training, as well as the perceived utility of training, there are a few suggested areas of improvement. First, in the assessment of the perceived utility of training, very few studies were found to actually ask participants how they would apply the newly learned behaviors back on the job (i.e., specific instances as to how/when these newly acquired skills might be beneficial to use/fit in). Second, Goldstein (1993) has argued for guidelines that should be followed in developing assessment of participant reaction, yet some of the reviewed studies seemed to fall short on these. For example, although Goldstein recommends that responses should be able to be tabulated and quantified, most of the studies reviewed, or at least the data presented, were not in a format where it was apparent that the data were able to be quantified. Typical evidence provided were things such as, “most of the participants reported liking the training”, or “a few selected cases reported liking the training”. A second concern is that while many studies reported using a Likert type scale to assess participant reaction there were a lack of studies that assessed the reliability of the scales used. A final concern is that many of the reviewed studies did not mention the specific components of CRM that were trained within the evaluated program. As CRM training is still not uniformly taught, the delineation of the particular skills taught is important in attempting to make sense of presented findings, as well as determining the extent to which findings may generalize. This last comment is not so much a critique of the data collection process as it is the dissemination of results.

Summary. Despite the shortcomings mentioned above, the studies that assessed participant reactions to CRM training provide sufficient rigor and converging evidence to lead to the conclusion that CRM training in aviation settings does produce positive reactions. For the next part, aviators like and perceive that CRM training is worthwhile and useful to the safe conduct of their tasks. Although this is the simplest form of evaluation criteria, it serves an important purpose. Positive reactions to training are
Do Aviators Learn About CRM?: Learning Evidence

In a multi-level evaluation effort, learning evidence is the second level of evaluation, and it refers to “the principles, facts, and skills which were understood and absorbed by participants” (Kirkpatrick, 1976, p. 11). Although evidence at this level includes the learning that occurred during the program, it does not include the actual exhibition of learned behaviors (i.e., skills). Also included in this level is the extent to which training leads to desired attitude changes (i.e. positive attitudes towards CRM training). It is considered here because not only are both learning and attitude changes cognitive events, but both processes mediate performance, and as such, should be evaluated together. The bottom line is that assessment of learning criteria provides evidence as to how successful the training program was in imparting the targeted knowledge, skills, and attitudes, as well as providing the basis for feedback and areas in need of further refinement.

Within the reviewed studies, 52% (30 of 58) collected information related to participant learning, with 11 of these studies solely assessing learning criteria (see Table 1). Within these efforts, the most common type of evidence offered in support of CRM affecting learning was changes in attitudes regarding CRM. Evaluation of attitudes was usually done by collecting information with the Cockpit Management Attitudes Questionnaire (CMAQ; Helmreich, 1984), or a modification of this instrument. The CMAQ is composed of three major scales: (1) communication and coordination (i.e., communication of intent and plans, delegation of tasks, assignment of responsibilities, and monitoring of crewmembers), (2) command responsibility (i.e., leadership), and (3) recognition of stressor effects (i.e., consideration and compensation for stressors). Overall, studies that assessed learning via attitude change seem to indicate that CRM training can produce positive changes in attitudes that are somewhat stable given top management support (see Table 1).

For the most part, CRM programs seem to produce positive examples of participant learning, primarily as indexed by attitude change (see Table 1), however there are a few reported instances of CRM programs achieving a “boomerang effect” (e.g., instances of negative attitude change; Helmreich, 1991). One study found that personality type influenced whether participants had positive or negative attitude change (i.e., boomerang) as a result of CRM training (Chidester et al., 1991). The remaining studies reporting this effect, however, did not report the possible cause of the negative attitudes (see Irwin, 1991). For example, was this boomerang effect due to: (1) something within the training itself, (2) crews having very high, positive attitudes prior to training (Chidester et al., 1991), or (3) personality or cultural aspects which may have played a role (Helmreich, 1991; Chidester et al., 1991). Despite evidence of a “boomerang effect” with some participants, the preponderance of evidence seems to suggest that the majority of participants attending CRM training do learn in the sense that there is typically a positive change in targeted attitudes.
Although the assessment of learning via attitude change is the most popular form of assessing learning, a few studies used other methods (see Hayward & Alston, 1991; Incalceterra & Holt, 1999; Salas, Fowlkes et al., 1999). For example, Hayward and Alston (1991) reported that as a result of CRM workshops there was an increased awareness of: human factors, crew performance, and potential stressors, as well as methods by which to handle these stressors. Assessing another form of learning (i.e., knowledge acquisition), Salas, Fowlkes et al. (1999) found that as compared to teams not trained in CRM, CRM trained groups exhibited higher levels of knowledge regarding CRM principles. Finally, several studies by Salas and colleagues (see also Stout, Salas, & Kraiger, 1997) have shown the positive effects of CRM training by assessing learning via a change in participant knowledge structures (i.e., mental models).

Although the overall picture seems to suggest that CRM training does have a positive impact on participant learning, another important factor to consider is the stability of these learning changes over time. Although there have been many cited studies (see Table 1) that show CRM affecting initial learning/change, there have been fewer studies that have assessed the long term stability of these changes. The few studies that have examined whether the changes produced by training programs remain stable over time have found varying results. For example, although some have indicated that positive attitudes are stable anywhere from two (Incalceterra & Holt, 1999) to five years out (Byrnes & Black, 1993), others have reported that initial attitude change produced by CRM programs declines over time, regressing towards pre-CRM levels (Irwin, 1991; Helmreich, 1991; Helmreich et al., 1999; Gregorich, 1993). Empirical results have suggested that one factor contributing to whether attitudes decline over time is whether management reinforces and supports the knowledge, skills, and attitudes learned in CRM programs or only provides “lip” service to material learned in CRM programs (Gregorich, 1993; Helmreich, 1991; Helmreich et al., 1999). Recent work by Helmreich and colleagues has begun to examine the impact of organizational culture, including the safety culture within an organization, on the initial learning of targeted CRM knowledges, skills, and attitudes, as well as the stability of changes over time (see Helmreich et al., in press).

Summary. Similar to the assessment of reaction data, the overall picture provided by learning criteria seems to suggest that CRM training is effective in producing changes in aviator knowledge and attitudes. This conclusion is further supported in that although the predominant form of collecting learning data is through the assessment of attitude change, several studies assessed other forms of learning as well (e.g., knowledge structures, paper-and-pencil tests), and found evidence of learning. The positive evidence offered by multiple measures of learning makes a stronger case for the effectiveness of CRM on learning. Although heading in the right direction, evaluation efforts must continue to strive for the inclusion not only of measures examining attitude change, but measures of declarative knowledge, as well as the assessment of knowledge structures and shared mental models (see Cannon-Bowers et al., 1989; Kraiger et al., 1993).

Do Aviators Apply The Learned CRM Behaviors in the Cockpit?: Behavioral Evidence

Behavioral evidence, the third level identified by Kirkpatrick (1976), provides an assessment of whether the lessons/knowledge learned in training transfer(s) to actual behavior on the job or a similar simulated environment. It has also been argued to
indicate: (1) the extent to which trainees learned how to perform the knowledge, skills, and attitudes (KSAs) taught during training, as well as when to apply these skills, and (2) an indication of trainee readiness and overall program effectiveness (Cannon-Bowers et al., 1989).

Of the 58 studies reviewed for this effort, 32 (55%) gathered some type of behavioral data, as defined by Kirkpatrick (1976). Furthermore, 12 of the reviewed studies collected information solely at this level. The reader is referred to Table 1 for a specific breakdown of studies that collected behavioral evidence. Within the reviewed studies, the most common method of assessing behavioral change was through the measurement of CRM-related behaviors while participants performed line-orientated simulation, such as LOFT. More specifically, this type of assessment was evident in 18 of the 32 studies (56%). Less common was behavioral assessment as measured by on-line assessment of behavior (11 of 32), although a small subset of studies did evaluate both behavior in line-orientated simulators, as well as on-line behavior (3 of 32).

The studies that collected behavioral data tended to use a combination of the following tools: behavioral observation forms, behavioral checklists, analysis of crew communication (via real time ratings or post-hoc analysis of video-tapes), and peer or self evaluations/reports. The predominant number of studies reviewed that collected some form of behavioral evidence indicated that CRM training had a positive impact on behavior (see Table 1). Specifically, results tended to indicate that CRM trained crews tended to exhibit: (1) improved performance as measured by behaviors indicative of CRM (e.g., decision making, mission analysis, adaptability, situation awareness, communication, leadership) or (2) improved performance as compared to crews not trained in behaviors indicative of CRM. For example, in a study by Leedom and Simon (1995) results indicated that after receiving CRM training, crews exhibited improved team communication patterns, more efficient management of crew resources, fewer team errors, and improved team coordination.

**Summary.** Behavioral evidence has been argued to be highly valuable in determining the effectiveness of a training program because it provides a means to assess whether training participants can actually translate the knowledge, behaviors, and attitudes learned in training into action. Overall, the behavioral evidence we reviewed suggests rather strongly that CRM training does have an impact on behavior (primarily as evidenced through LOFT or similar evaluations). Aviators do exhibit more teamwork behavior in the cockpit. And presumably these will lead to safer outcomes. Although more evaluations were found at the behavioral level than initially expected, most of these evaluations have been conducted in simulated environments rather than actually “on the job”. Although resources may put restrictions on the ease with which behavioral data are captured back on the actual job, as opposed to simulated conditions, behavior on the actual job would provide even stronger support for the suggested effectiveness of CRM training with regard to behavioral change. However, behavioral data collected during simulated situations is definitely a close surrogate and a welcome start in the right direction.

**Are The Skies Safer?: Results/Evidence of Organizational Impact**

Organizational impact (i.e., increased safety, less errors) is the highest level of evaluation in Kirkpatrick’s (1976) framework. Although this type of evidence is highly valued, very few evaluations are conducted at this level—only 10% (6 studies) of the
reviewed studies collected evaluation data at this level (see Table 1). Due to the difficulty of collecting this type of information (in terms of time, resources, identification of a clear criterion, and low occurrences of accidents and mishaps), evidence of CRM’s training impact on the organization as a whole is not often sought, nor obtained. Specifically, evidence of this type is difficult to collect because it generally requires some type of longitudinal data, as it takes time for the impact of training to appear at the organizational level. In addition, criterion measures are difficult to identify and it is hard to control the various extraneous variables that may influence (e.g., moderate, mediate) the relationship between CRM training and organizational effectiveness.

Of the six studies that collected some form of results measure, two collected information on organizational effectiveness alone, while the others collected additional types of evaluation evidence also (more on this later). Within the reviewed studies that collected data at the organizational level, most of the information tended to come from one of two sources: (1) anecdotal evidence (e.g., accident reports, incident reports) or (2) longitudinal studies. The predominant type of evidence that has been used to illustrate CRM’s impact on aviation safety is anecdotal reports contained in incident or accident investigations conducted by the National Transportation Safety Board (NTSB). For example, Kayten (1993) cites several examples of reports by the NTSB in which good CRM practices were reported to limit the detrimental effects of either human or mechanical error. Although anecdotal reports contained in accident reports are perhaps the most common and easiest evidence of organizational impact to collect, there are some problems with accident data that argues for other types of organizational effectiveness measures. Perhaps the most predominant is the rarity with which accidents happen (see Maurino, 1999; Gregorich & Wilhelm, 1993; Helmreich & Foushee, 1993; Salas, Prince, et al., 1999), as such the investigation of incidents (as opposed to accidents) has been suggested.

The other source of evidence used within the reviewed studies to assess the impact of CRM training on organizational effectiveness, is based on the longitudinal collection of data. For example, Byrnes and Black (1993) evaluated a CRM program implemented at Delta Airlines and found indications of CRM’s impact on organizational effectiveness—quarterly air carrier discrepancy reports were found to significantly decrease after CRM training was implemented. Although assessing organizational impact, the above study was designed with no control group to act as a comparison against potential intervening environmental confounds across the years referred to within the air discrepancy reports. As such, although the evidence provided is positive, it is not conclusive – more controlled studies need to be conducted.

Summary. So, what can we conclude from the reviewed evidence? Unfortunately, not much. Although anecdotal reports indicate that CRM behaviors may contribute to reducing the impact of human and mechanical error within the aviation community, much stronger evidence is needed. And this is easier said than done. There are a number of difficulties for establishing a clear cause and effect between CRM and safety. For example, various factors may intervene between the point at which a CRM program is implemented and the assessment of the program’s impact, making inference somewhat problematic. Clearly, evaluation efforts that systematically track the impact of a particular CRM program on safety provide a stronger argument than anecdotal reports. The bottom line appears to be that although there is evidence of a positive trend regarding
the impact of CRM on safety more and better evaluations need to be conducted. And, in fact, of the evaluations reported, these need to be more detailed—for there were several evaluations that were found in the course of the literature search that, while hinting at the impact of CRM on a particular subset of crews, did not provide enough information to reach any real conclusions. Specifically, the information presented was either ambiguous or very general in nature; as such, those efforts were not included in the current review as it was felt that the results were too speculative.

**Multi-Level Evaluation Efforts**

Reviewing the 58 identified studies via Kirkpatrick’s typology has suggested that when each level of training evaluation criteria is considered independently, there is a fair amount of evidence to suggest that CRM training does have a positive impact on participant reaction, learning, and behavior, although the impact on safety can not yet be determined. However, the above conclusions are, in some sense, based upon weak data, for in order to truly assess whether CRM training is effective, evaluation efforts should focus on collecting information from multiple levels (e.g., reaction, learning, behavior, results). The collection of multi-level assessment data is important in that it provides a cleaner, more complete picture of training efficacy because the evaluator is not looking at one piece of evidence in isolation. In order for training to be truly effective, it must impact participant learning, learning must transfer to behavior, and behavior must transfer to a difference at the organizational level. The assessment of participant reactions is also an important part of the multi-level assessment, for reactions provide an initial check as to whether training is relevant to the knowledge, skills, and abilities needed on the job, as perceived by participants. In addition, reaction data serves as an important piece of information in that “liking” the training will serve to motivate participants in the learning process.

Of the 58 identified studies, 24 (41%) collected information at multiple levels of Kirkpatrick’s (1976) typology. Of these 24, most collected information at only two levels (13 studies), typically the lower levels (e.g., reaction-learning, reaction-behavior-, see Table 1). Although a review of those studies indicated that CRM training generally produced positive results, assessment at only two levels still provides a limited view of the overall effectiveness of the program. A clearer, more complete picture would be provided by either collecting information at more levels or at least the higher levels (i.e., behavior-results).

There was a fair number of studies that evaluated training programs at three of Kirkpatrick’s four levels (10 studies), but only one study was found that evaluated training at all four levels (see Table 1). As the studies that assess three or four levels of Kirkpatrick’s typology provide the clearest picture as to the actual effectiveness of CRM training, a brief overview of the value added by these studies will be described.

The studies that assessed at least three levels of evidence attest to the fact that the various types of evaluation evidence provide different pieces of information, which may lead to a different overall conclusion as to the effectiveness of CRM training. For example, Smith (1994) found that when participant reactions and learning (attitudes) were assessed, a self analysis technique used to develop CRM skills in 10 undergraduate flight students was not highly valued and produced little change in attitudes. However, when behavior was assessed, it was found that self analysis did have an impact in that it helped crews to perform significantly better in three LOFT sessions and moderately
better in three others. Another example of the benefit of conducting multiple level-evaluation are the results of a study conducted by Stout et al (1997). Specifically, these researchers found positive reactions to training, evidence of learning via changed knowledge structures, and evidence of behavioral change in that trained participants performed an average of 8% more desired CRM behaviors than a control group. However, when learning was assessed via attitude change, there was a positive, but non-significant change. Had either of the above studies only collected single-level evaluation data, they might have come to different overall conclusions as to how to improve the implemented training program.

Alkov and colleagues (see Alkov, 1991; Alkov & Gaynor, 1991) provided the only reviewed study that evaluated CRM training at all four levels of Kirkpatrick’s (1976) typology. Specifically, they evaluated a CRM training program targeted at 45 naval aviation training squadrons—helicopters, attack bombers, and multi-placed fighters. Results suggest that squadrons reported the training as useful and they indicated a desire for training to continue. Positive attitude changes regarding CRM training were noted, and squadron commanding officers reported that training was found to contribute to better communication between instructors and students. Finally, some evidence of organizational impact was found in that following CRM training, overall aircrew mishap rate declined in all three communities. However, it should be noted that this study is not meant to serve as an example to others as a way of conducting multi-level evaluations, as it is a preliminary analytical effort and hence many of the above conclusions offered are tentative. We do however commend their efforts in attempting to evaluate CRM training at all four levels.

Summary. It has been said that in the absence of a well-defined, measurable, “ultimate” criterion (that rarely exists in the real world), it is important to assess training at multiple levels for each additional source of data serves to increase confidence in the overall evaluation (Cannon-Bowers et al., 1989). For example, Although reaction data can indicate whether the trainee felt the program was worthwhile, it has little if any relation to whether the participant learned the material. Similarly, just because a trainee learned the knowledge during training does not guarantee that he/she can translate this knowledge into effective behavior, nor does it guarantee that, if applied, the behavior will have an effect on organizational outcomes. Each source of data provides a limited picture of the results.

The studies reviewed above indicate that this advice, concerning the assessment of multiple-levels, is beginning to take hold within the aviation community. However, all the parties involved (aviators, researchers, regulators, the public, the airlines) need to continue to push for assessment at all levels of evaluation. Although the efforts reviewed within this paper suggest that CRM training programs are generally effective in producing some level of change in participants (e.g., reaction, learning, behavior), the lack of multi-level evaluation efforts makes it difficult to answer whether CRM is truly effective. Specifically, of the 58 reviewed studies, the predominant number of studies (34) collected data at only one level, whereas the remainder broke down in the following manner: 13 studies collected data at two levels, and 10 studies collected data at three levels, as argued for by Kirkpatrick (1976) and others (Robertson, Taylor, Stelly, & Wagner, 1995; Cannon-Bowers et al., 1989). In order to be truly effective and worth the investment companies and airlines make into CRM programs, positive reactions must
transfer to learning, learning must transfer to behavior, and finally changes in behavior must translate into reducing aviation related mishaps and accidents. We recognize that achieving four levels of evaluation might be impractical, if not impossible in many situations. But we need to establish stronger links between CRM training and reduction of accidents. At this point in time, there are not enough multi-level evaluations to assess whether or not this link is there.

**DOES CRM TRAINING WORK?**

As reported by Salas et al (1999) the data are encouraging. Although some have previously argued that there is no evidence that CRM is effective (Besco, 1995, 1997, 1998; Simmon, 1997; Komich, 1997), this review concludes that some evidence does exist. And this is important. The picture that has emerged after reviewing the existing evidence within the current framework suggests that CRM training is effective. But as stated earlier, the picture is not as clear as it should be after 20 years. The lack of systematic studies that can clearly show cause and effect, as well as the transfer of learned material to behavior and behavior to results, is a key factor in this unclear picture. Nevertheless, given that CRM training is one of a number of factors that may influence the practice and effectiveness of CRM behaviors, it may be argued that, although imperfect, the current evidence for the effectiveness of CRM training programs is impressive. Specifically, what can be said is that CRM (generally) produces: (1) positive reactions, (2) enhanced learning, primarily as measured through attitude change, although other learning criteria are also used (e.g., knowledge tests, knowledge structures), and (3) desired behavioral change in the cockpit (simulated or real). However, what can not be answered with certainty is whether CRM training has an effect on the bottom line -- in this case aviation safety. At this point, we believe the tools to determine this are there; what we need are the resources and a mandate to make it happen.

**WHERE DO WE GO FROM HERE?**

In terms of evaluation, the current review would seem to suggest two areas where future efforts should be concentrated. First, is that additional evaluation efforts need to begin to assess CRM training at multiple levels, using multiple sources of criteria within each level. Currently, less than half (40%) of the studies published used multiple levels to evaluate training. As noted, analyses of these studies suggest that the use of multi-level evaluations provides a much clearer and diagnostic picture of training efficacy than does single-level evaluations.

Furthermore, we found that reaction data were primarily gathered via questionnaires or verbal reports of how well participants liked the course or thought it was worthwhile. Learning criteria were primarily collected through the assessment of attitude change via the CMAQ. Although there have been several other types of learning criteria argued for (structural knowledge, knowledge tests), few studies collected more than one type of learning criteria within the same evaluation study. In most of these studies, attitude change was the only criterion collected. Similar arguments can be made with regard to the collection of behavioral data and data examining organizational impact. Utilization of multiple methods to assess the same type of criteria (e.g., learning) increases confidence in the results obtained. The importance of longitudinal evaluation efforts that use multiple measures and methodologies have been made (see Helmreich &
Wilhelm, 1987; Helmreich, Wilhelm, Gregorich, & Chidester, 1990) and need to be conducted.

The second area that would strengthen the evaluation effort(s) is to ensure that the evaluation and dissemination of results regarding CRM training provide diagnostic information. Although multi-level evaluation is one piece of providing diagnostic information, there are at least three others. Specifically, many of the evaluation efforts reviewed for the current paper have been written up in such a way as to make it difficult to assess the degree to which tools used to collect evaluation data were theoretically driven or possessed acceptable psychometric properties (e.g., reliability, validity). [This raises an issue that I don't recall seeing anywhere in your earlier discussions. It might be nice to have a foreshadowing of this earlier on in the paper.] Related to this need is Helmreich and Wilhelm’s (1987) plea for more guidance for the aviation community in how to implement and evaluate CRM programs. The ambiguity that exists in regards to the content of CRM training, implementation, and evaluation methods, combined with the lack of explicit descriptions as to the content contained within evaluated CRM programs makes it hard to compare findings across studies or provide general guidance.

Finally, the predominant number of studies that we reviewed provided descriptive data only or generalized anecdotal evidence. For example, in relation to reaction data several researchers report that -- participants generally found training useful. This type of information is very subjective in interpretation. Specifically, there is no type of relational meaning, nor does the reader know how “useful” is defined (i.e., on a scale, are there anchor points, totally open responses). Although descriptive data are obviously better than generalized anecdotal evidence, it still makes it hard to determine the “effect” of CRM training as neither significance nor effect sizes can be determined. Future studies should attempt to report data at higher levels than merely descriptive information so that better conclusions can be drawn.

**Exploration and Expansion**

As a field in general, CRM training and the corresponding evaluation efforts are moving in two general directions. First, as the aviation community continues to invest money into CRM training, other communities are taking note and beginning to implement similar programs (Salas et al., 2001). The emerging extension of CRM training to other domains further drives the need for multi-level evaluation efforts so the question as to the effectiveness of CRM training can be answered once and for all.

A second area that has begun to be investigated (Merrit & Helmreich, 1995b, Helmreich & Merritt, 1998; Maurino, 1999; Chidester et al., 1991; Helmreich, 1997), but must be further examined, is the identification of variables that may moderate or mediate the relationship between CRM training and performance. Understanding how such factors as culture (e.g., national, professional, safety, and organizational), personality, and organizational climate impact the message delivered in CRM training will help in the design, delivery, and evaluation phases of CRM. This in turn would serve to allow the aviation community, as well as other areas of industry, to get the “most bang for their buck”.

**CONCLUDING REMARKS**

After reviewing the 58 identified evaluations of CRM training programs within the aviation community the following can be said. First, CRM training programs seem to produce positive participant reactions, learning, and application of learned behavior via
simulators or on-line/on the job. However, the final word on whether CRM has an impact on safety remains to be seen. Second, although the aviation community should be commended in that multi-level evaluations are becoming more common, evaluation needs to become a systematically accepted cost of business. Third, the review raised a few methodological concerns in that, many times, descriptions of the components of CRM training or the methodology used to develop and test evaluation measures were not very clear within the literature, making it hard to determine the reliability, validity, and transferability of reported results. At this point in time it is unclear as to whether this is a concern with pure methodology or a combination of methodology and procedures used to disseminate findings. Finally, the current review illustrated that although there are still some rough spots in terms of evaluating implemented CRM training programs, the picture is not as bleak as some opponents would make it out to be – trends seem to indicate that CRM training does have an impact on multiple aspects of the individuals and crews completing the program. However, more and better evaluations are needed. And the aviation community should demand it. We believe that time and continued systematic evaluations will reveal its long term impact—improved safety in the skies.

ACKNOWLEDGEMENTS

We would like to thank Eleana Edens, Deborah A. Boehm-Davis, Janis A. Cannon-Bowers, and two anonymous reviewers for their thoughtful comments on earlier drafts. This work was performed under the auspices of the UCF/FAA/NAWCTSD Partnership for Aviation Team Training.
REFERENCES


Proceedings of the Third Australian Aviation Psychology Symposium (pp. 185-192). Ashgate: Avebury Aviation.


Table 1

Summary of CRM Evaluative Efforts

<table>
<thead>
<tr>
<th>Source</th>
<th>Community</th>
<th>CRM training content</th>
<th>Type of Study/Data collection</th>
<th>Reactions</th>
<th>Learning</th>
<th>Behavior</th>
<th>Results/ Organizational Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reaction Studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baker, Bauman, &amp; Zalesny (1991)</td>
<td>41 CH-46 pilots</td>
<td>- Pre-Flight Brief&lt;br&gt;- Assertiveness</td>
<td>Quasi-experimental&lt;br&gt;Self-report survey</td>
<td>- Review of means indicated that Pre-Flight Brief exercise was a worthwhile addition to ACT course and likely to have an impact on next briefing experience (approx 4.4 on 5 pt scale)&lt;br&gt;- Able to cite specific ways they planned to use information gained in both exercises&lt;br&gt;- Open ended questionnaire revealed favorable impressions of both exercises&lt;br&gt;- 75% ranked role play assertiveness exercise as 1st or 2nd choice (n=4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baker, Prince, Shrestha, Ober, &amp; Salas (1993)</td>
<td>112 male military aviators</td>
<td>- No specification of skills taught&lt;br&gt;- Acceptability of tabletop training system to augment CRM training</td>
<td>Quasi-experimental&lt;br&gt;Self-report survey</td>
<td>- 90% of aviators agreed that tabletop system could be used for CRM skills training&lt;br&gt;- Most felt good way of learning&lt;br&gt;- Most agreed system demonstrated importance of CRM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chute &amp; Wiener (1995)</td>
<td>Survey conducted at 2 airlines</td>
<td>- No specification of skills taught</td>
<td>Quasi-experimental&lt;br&gt;Self-report survey</td>
<td>- Good program that should be reinstated (joint)&lt;br&gt;- Good reactions, suggest extend to cabin crew (other)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clark, Nielsen, &amp; Wood (1991)</td>
<td>135 commercial airline pilots</td>
<td>- No specification of exact skills taught (mention)</td>
<td>Quasi-experimental&lt;br&gt;Self-report survey</td>
<td>- Felt CRM enhanced information processing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Methodology</td>
<td>Survey Type</td>
<td>Findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maschke, Goeters, Hormann, &amp; Schiewe (1995)</td>
<td>Cockpit crews; DLR/Lufthansa</td>
<td>Experimental</td>
<td>Self-report survey</td>
<td>- Majority indicated that negative effects of aviation stress outweigh positive effects of CRM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hormann, Goeters, Maschke, &amp; Schiewe (1995)</td>
<td>750 participants in Lufthansa cockpit crews</td>
<td>Experimental</td>
<td>Self-report survey</td>
<td>- 80% gave thought course was useful or extremely useful (as per Likert scale)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schiewe (1995)</td>
<td>724 cockpit members</td>
<td>Quasi-experimental</td>
<td>Self-report survey</td>
<td>- Preliminary data indicate that 90% indicated that the course content was highly relevant - Preliminary data indicate that 84% thought the method of presentation was attractive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vandermark (1991)</td>
<td>America West Flight attendants and cockpit crew (n=appx. 1200)</td>
<td>Quasi-experimental</td>
<td>Self-report survey</td>
<td>- Units that were based on case studies or those that used role play in job related scenarios had very positive reactions - Methods based mostly on lecture was not rated favorably</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young (1995)</td>
<td>42 Junior year flight students at Purdue University</td>
<td>Quasi-experimental</td>
<td>Self-report survey</td>
<td>- Favorable reactions (Likert scale)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alkov &amp; Gaynor (1991)</td>
<td>58 CRM training instructors</td>
<td>Quasi-experimental</td>
<td>Self-report survey/ CMAQ</td>
<td>- Positive shifts in attitudes were detected as a result of instructor training course</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chidester, Helmreich, Gregorich, &amp; Geis (1991)</td>
<td>528 aviators from USAF military airlift command</td>
<td>Quasi-experimental</td>
<td>Self-report</td>
<td>- Training produced both positive and negative attitude change via type of personality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Community</td>
<td>CRM training content</td>
<td>Type of Study/Data collection</td>
<td>Reactions</td>
<td>Learning</td>
<td>Behavior</td>
<td>Results/Organizational Impact</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>----------------------</td>
<td>-----------------------------</td>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Gregorich, Helmreich, &amp; Wilhelm (1990)</td>
<td>National air carrier (606 participants)</td>
<td>- No specification of skills taught</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Positive attitude change (CMAQ) - Reduction in response variation for 2 scales of CMAQ after training</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Grubb, Morey, &amp; Simon (1999)</td>
<td>Air Force flight crews (n=2095) and mission crews (n=564)</td>
<td>- Air Force CRM behaviors (group dynamics, stress awareness, mission planning, risk mgmt behaviors, workload mgmt, comm., situation awareness, human performance behaviors)</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Results were combined for all platforms – indicate significant positive attitude change as per all 8 behaviors</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Helmreich, Merritt, &amp; Wilhelm (1999)</td>
<td>Pilots in several organizations surveyed just after completion of CRM course and several years later</td>
<td>- No specification of skills taught</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Decay in attitudes not immediately apparent, but some decay over time</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Irwin (1991)</td>
<td>Major U.S. air carrier</td>
<td>- No specification of exact content</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Significant positive attitude change (CMAQ) - Few examples of boomerang - Found attitudes decline over time - Recurrent training results in positive attitude change</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Morey, Grubb, &amp; Simon (1997)</td>
<td>188 fighter, 198 transport, and 77 bomber Air Force pilots undergoing CRM training</td>
<td>Air Force CRM training: 8 core CRM behaviors</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Pre-training attitudes varied by pilot group (transport vs. fighter, bomber) - For all 3 groups pilot attitudes toward CRM significantly improved with training</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
- Improvement rate varied across groups with most positive post training attitudes being shown by transport, followed by bomber and fighter pilots

Rollins (1995)  
3 commercial airlines (2 US, 1 Canadian); 1 US military (n=508)  
- No description of specific CRM skills trained  
- Trained lasted 1-2 days  
Quasi-experimental  
Self-report survey/CMAQ  
- Measurable improvement in positive CRM attitude (CMAQ)

Approximately 88 general aviation pilots  
- No specification of skills taught  
Quasi-experimental  
Self-report survey  
- Pilots that had previously completed human factors course were significantly different in their attitudes than those who hadn’t  
- Pilots who had a human factors course were more confident in terms of their ability to cope with emergency situations and exhibited a relatively heightened level of self-awareness

Yamamori & Mito (1993)  
2300 crew members of Japan airline  
- Recognize and understand different interpersonal styles and effects on crew interaction in cockpit (inquiry, advocacy, conflict, problem def, critique, problem solving)  
Quasi-experimental  
Self-report survey  
- Strengthens and crystallizes attitudes toward more effective CRM in cockpit

Arnold & Jackson (1985)  
96 flight crews at United Airlines  
- Recurrent training (topic changes each year)  
- Decision making  
Quasi-experimental  
Rating  
- No statistical analysis presented  
- On a 6 pt scale that evaluates overall CRM performance in LOFT most fall between 3-6 (most 4-5)

17 crews of active duty USAF  
- No specification of CRM skills previously taught  
- Compared fixed and formed crews  
Quasi-experimental  
Observation  
- No significant differences in CRM behaviors between 2 crews  
- Average overall rating using LOS checklist was 3.35 (fixed) and 3.38 (formed)  
- Fixed crews committed more minor errors than formed crews (4.4 vs 2.6), while there were no significant differences in the number of major errors committed via crew formation

Brannick, Prince, Prince, & Salas (1995)  
51 military air crews (Navy)  
- No specification of skills trained  
- Skills measured (assertiveness, decision making/mission analysis, adaptability,  
Quasi-experimental  
Observation  
- On average participants were rated as slightly above average in performing behaviors related to CRM (mean=3.56, 3.67 Scenario A and B)
Clothier (1991) Major domestic airline (3,000 crews; 2000 untrained, 1000 trained) Also look at data across 5 airlines – same results

- Significant improvement seen in LOFT exercises (485 trained outperformed 1625 untrained crews)
- Crews flying on the line exhibited improved performance (significant improvement in 12 of 14 areas)
- Crew interactions earned higher scores (LOS) as they progressed through initial training to recurrent training

<table>
<thead>
<tr>
<th>Source</th>
<th>Community</th>
<th>CRM training content</th>
<th>Type of Study/ Data collection</th>
<th>Reactions</th>
<th>Learning</th>
<th>Behavior</th>
<th>Results/ Organizational Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connolly &amp; Blackwell (1987)</td>
<td>Aeronautical Science students at Embry Riddle University – (16 exp., 13 control)</td>
<td>- Risk assessment - Decision making - Hazardous thought patterns</td>
<td>Experimental Observation</td>
<td>- Checklist scores and flight ratings indicated exp. group performed significantly better on post-test than control - Compared with control group exp. group witness a significantly greater amount of change on both checklist scores and flt ratings after training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helmreich, Wilhelm, Gregorich, &amp; Chidester (1990)</td>
<td>Major airline Over 2000 line flights &amp; LOFT sessions For 859 crews LOFT involved in flight emer.</td>
<td>- No specification of skills taught</td>
<td>Quasi-experimental Observational ratings</td>
<td>- Global rating of overall crew performance using LINE/LOFT worksheet – behavior in both on line and in LOFT indicates positive changes in CRM performance - Wide variation in specific CRM behaviors used between fleets at same airline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helmreich &amp; Foushee (1993); Taggart (1994)</td>
<td>Major airline</td>
<td>- No specification of skills taught</td>
<td>Quasi-experimental Observations</td>
<td>- Significant positive shifts in process behavior during line operations across a 3-year period.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Methods</td>
<td>Findings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>---------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hansberger, Holt, &amp; Boehm-Davis (1999)</td>
<td>19 I/E evaluating 2 above fleets; considering all pilots that these I/Es have evaluated over past 6 months</td>
<td>Quasi-experimental Instructor Ratings after the fact</td>
<td>ACRM pilots assessed higher than non ACRM pilots on workload management, comm., and planning. Expected questions from behavioral observation form to factor into workload, comm., and situation awareness; situation awareness factored into planning instead. I/E has positive evaluations for ACRM training especially in area of communication. ACRM strong in facilitating establishment of bottom lines and back up plans, not so strong in reducing distractions or helping crew to be generally organized and composed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jentsch, Bowers, &amp; Holmes (1995)</td>
<td>20 instrument rated pilots from Embry Riddle</td>
<td>Experimental Observation</td>
<td>Communication ratio decreased for groups receiving training. Crews tended to maintain efficient comm. patterns through all 7 flights. Comm. frequencies did not change significantly after 45 days. Assessed performance differences, none.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nullmeyer &amp; Spiker (under review)</td>
<td>87 students in MC-130P</td>
<td>Quasi-experimental Instructor comments translated into ratings</td>
<td>Based on rating system devised from instructor comments in grade folders found students rated above average in mission preparation and crew coordination. Students rated below average in decision making and comm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Community</td>
<td>CRM training content</td>
<td>Type of Study/Data collection</td>
<td>Reactions</td>
<td>Learning</td>
<td>Behavior</td>
<td>Results/Organizational Impact</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Predmore (1991)</td>
<td>United Flight 232 (comm. analysis)</td>
<td>- No specification of CRM skills taught at United</td>
<td>Post-hoc Case Study</td>
<td>- Efficient use of resources - Distribution of communication across multiple tasks and members - Maximum utilization of 4th crew member - Explicit prioritizing of tasks - Active involvement of Captain through entire process</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diehl (1991)</td>
<td>Summary paper</td>
<td>- No specification of specific skills</td>
<td>Post-hoc</td>
<td>- Reduction in accident rates cited from 4 environments (USAF, USNavy helicopters, USNavy fighter-bomber, Petroleum Helicopter Inc)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kayten (1993)</td>
<td>NTSB reports</td>
<td>- No specification of skills taught within particular incidents</td>
<td>Post-hoc</td>
<td>Several near incidents saved by good CRM practices – cites incidents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alkov (1991)</td>
<td>45 Naval aviation training squadrons (helicopters, attack bombers, multi-placed fighters)</td>
<td>- Pilot judgment - Situation awareness - Decision making - Policy and regulations - Command authority - Workload performance - Use of available resources - Communication skills</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Squadrons report benefit and wish to continue - Contributed to better communication between instructors and students - Overall arcnew mishap rate declined in all three communities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Byrnes &amp; Black (1993)</td>
<td>Delta airlines</td>
<td>- Crew dynamics (leadership/followership) - Communication - Decision making - Stress management</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Positive attitude change as measured by CMAQ - Stability of change over 5 year period - Flight attendants reported that cockpit crew treated them with more respect after CRM - Flight attendants reported that cockpit crew made them feel more a part of crew and were included in more pre-flight briefings - Quarterly air carrier discrepancy reports significantly decreased</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geis (1987)</td>
<td>838 US Army pilots – 163 completed both pre and post CMAQ (results based on these)</td>
<td>- Decision making - Communication - Resolving conflict - Assertiveness - Feedback/criticism - Pilot attitudes</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Paired t-tests indicated significant positive attitude change - At an item level most of the individual items of the CMAQ witnessed positive - Effective in helping them to reduce potential for mishaps; consciously apply techniques taught - 135 felt had become safer pilots - Some were able to comment on specific instances where had used</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Methodology</td>
<td>Findings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halliday, Biegański, Inzana (1987)</td>
<td>Quasi-experimental survey, Peersurvey</td>
<td>95% of trainees see themes as relevant, 50% said they frequently changed their cockpit behavior with other crew members, 65% report that the way they analyze situations or their own behavior has changed, 65% reported a difference in the ways that crews operate after CRM training, 80% self reported changes during preparation phase, 35% reported changes during debriefing phase, 10% report a large impact in terms of changing life at squadron level (51% felt some occasional change)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hayward &amp; Alston (1991)</td>
<td>Quasi-experimental survey</td>
<td>Most participants responded enthusiastically, Argue that attitudes toward CRM are also improving, Of those not trained in CRM who had flown with a person trained in CRM, 75% indicated that CRM trained crew members exhibited recognizable behavioral changes, 80% of untrained individuals felt that they had observed better coordination and flightdeck atmosphere from crew members who had undergone training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helmreich (1991)</td>
<td>Quasi-experimental survey</td>
<td>8,000 surveys from 3 airlines indicate that LOFT is valued by crews as a training technique, considerable variability in quality of scenarios, Without reinforcement flightdeck management attitude’s began to regress to pre-CRM levels, Cite normative cultural differences in attitudes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Community</td>
<td>CRM training content</td>
<td>Type of Study/Data collection</td>
<td>Reactions</td>
<td>Learning</td>
<td>Behavior</td>
<td>Results/Organizational Impact</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------</td>
<td>----------------------------------------------------------</td>
<td>--------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Helmreich &amp; Wilhelm (1991)</td>
<td>2 major airlines; independently developed CRM courses</td>
<td>- No specification of exact content</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Overall pattern of evaluation was extremely positive</td>
<td>- Significant positive change on all 3 scales of CMAQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ikomi, Boehm-Davis, Holt, &amp; Incalcaterra (1999);</td>
<td>50 crews eastern US regional airline; 2 fleets (experimental/control fleet)</td>
<td>- Communication - Situation assessment - Planning/Decision Making</td>
<td>Quasi-experimental Self-report survey</td>
<td>- 93% of pilots voted to expand use of ACRM to other fleets</td>
<td>- Pilots trained in ACRM showed positive attitudes toward CRM (CMAQ)</td>
<td>- 91% felt it had improved their flight performance</td>
<td>- Trained fleet performed significantly better overall (observer summary evaluation – jump seat observations)</td>
</tr>
<tr>
<td>Incalcaterra &amp; Holt (1999)</td>
<td>600 active pilots in above airline (184 trained, 84 non ACRM group)</td>
<td>- Communication - Situation assessment - Planning/Decision Making</td>
<td>Quasi-experimental Self-report survey</td>
<td>- 93% of pilots voted to expand use of ACRM to other fleets</td>
<td>- Pilots trained in ACRM showed positive attitudes toward CRM (CMAQ)</td>
<td>- 91% felt it had improved their flight performance</td>
<td>- Trained fleet performed significantly better overall (observer summary evaluation – jump seat observations)</td>
</tr>
<tr>
<td>Holt, Boehm-Davis, &amp; Hansberger (1999)</td>
<td>All pilots in ACRM fleet and control fleet</td>
<td>- Communication - Situation assessment - Planning/Decision Making</td>
<td>Quasi-experimental Self-report survey</td>
<td>- 93% of pilots voted to expand use of ACRM to other fleets</td>
<td>- Pilots trained in ACRM showed positive attitudes toward CRM (CMAQ)</td>
<td>- 91% felt it had improved their flight performance</td>
<td>- Trained fleet performed significantly better overall (observer summary evaluation – jump seat observations)</td>
</tr>
<tr>
<td>Jackson (1983)</td>
<td>4059 participants at United Airlines</td>
<td>- Inquiry - Advocacy - Conflict resolution - Critique - Decision making</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Descriptive stats only - Participants tend to find experience quite rewarding (avg. response 6.8 on a 9 pt scale) - Over 95% of trained crew members felt CRM training has value to crew and can be applied to operation of their aircraft - LOFT program seems quite successful overall; most participants rate LOFT experience as very good</td>
<td>- 41% shift in self-perceived style of interaction – crew members more aware of own interaction style</td>
<td>- Argues that improvement has been noted in nearly all of the targeted areas during each of four quarters of the first year – sounds like self-reported improvement. - Steady decline in the number of proficiency check failures since advent of LOFT at United</td>
<td>- Trained fleet performed significantly better overall (observer summary evaluation – jump seat observations)</td>
</tr>
<tr>
<td>Lassiter, Vaughn, Smaltz,</td>
<td>Undergraduates from University of Central</td>
<td>- Communication - 3 conditions (control, Experimental)</td>
<td>- Attitude change via CMAQ non-significant</td>
<td>- Teams receiving skill based comm. training exhibited significantly</td>
<td></td>
<td></td>
<td>- Trained fleet performed significantly better overall (observer summary evaluation – jump seat observations)</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Methodology</td>
<td>Key Findings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morgan, &amp; Salas (1990)</td>
<td>Florida (n=90) knowledge based training, skill based training</td>
<td>Self-report survey, Observational ratings</td>
<td>better communication skills than other two conditions (behavioral observation rating scale)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 US Army AH-64 aviators (battle-rostered crews)</td>
<td>Quasi-experimental Self-report survey, Observations</td>
<td>- Improved team communication patterns, more efficient mgmt. of crew resources, fewer team errors; team coordination improved (ACE checklist) - Mission performance ratings improved as did mission accomplishment (# completed as well as quality) - Improvement across all 13 dimensions of team coordination - Higher flight proficiency ratings - Higher mission performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mudge (1983)</td>
<td>25 pilots enrolled – some with FAA, corporate pilots, individual pilots, airline executives testing program</td>
<td>Quasi-experimental Self-report survey</td>
<td>- No final conclusions as program is ongoing – present trends - No pilot has rated a study unit at less than good – more flight time=higher ratings - Too early to answer if produces positive behavioral changes, but positive indications - The four pilots that have graduated at this point self report that their actions on the flight deck are changing as a result of program - Also have unsolicited comments from several still in program that their own actions are changing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naef (1995)</td>
<td>Swissair line pilots and instructors</td>
<td>Quasi-experimental Self-report survey</td>
<td>- Majority of pilots felt course presentation was above average (approx. 560 out of 640) - Majority of instructor pilots felt the course related to their needs and was worthwhile - 53% response rate 6 months later indicated that 70% of instructors would invest an off duty day for this type of training - 53% response rate on survey taken 6 months later indicated that 97% of flight crew members reported one or more positive behavior transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salas, Fowlkes, Stout, Milanovich, &amp;</td>
<td>35 pilots &amp; 34 enlisted aircrewmen from Navy transport</td>
<td>Quasi-experimental</td>
<td>- Trained group showed more positive attitudes towards use of CRM - Overall trained teams performed better than untrained teams (TARGETS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
###Prince (1999)
- helicopter squadron
- Mission analysis
- Self-report survey, Multiple choice, Observation
- Significant increases in positive attitudes via overall attitude scale (ACAQ) and CMAQ Communication and Coordination subscale
- Trained group exhibited higher levels of knowledge regarding CRM principles
- Strong endorsement of training
- Positive support for CRM training
- Reported that skills learned would be implemented for more effective pre and debriefs
- Correctly managed 15% more during prebrief and 9% more during higher workload segment

###Smith (1994)
- 10 undergraduate flight students
- Self analysis vs. traditional debrief to develop CRM skills
- Quasi-experimental Self-report survey, Observation
- Self analysis was not valued highly as a training technique
- LOFT reported as being more helpful than self analysis technique
- Positive reaction to active learning technique
- Self analysis helped crews to perform significantly better in 3 LOFT sessions, moderately better in 3

###Stout, Salas, & Fowlkes (1996); Stout, Salas, & Fowlkes (1997)
- 42 student pilots in Navy Advanced Maritime Curriculum
- Communication Assertiveness Situation Awareness
- Experimental Self-report survey, Observation
- Positive reaction to training
- Examples provided of how use new skills on job
- Trained group scored higher on knowledge test
- Positive attitude change found (CMAQ)
- Trained teams performed better (TARGETs) as per both real-time instructor ratings and post-hoc ratings of video

###Stout, Salas, & Kraiger (1997); Fowlkes, Lane, Salas, Oser (1994); Fowlkes, Lane, Salas, Oser, & Prince (1992)
- 22 aviators, helicopter community
- Communication Assertiveness
- Experimental Self-report survey, Observation
- Positive reaction to training (n=12)
- Evidence of learning via changed knowledge structures (n=22)
- Attitude change positive, but not significant (n=12)
- Knowledge test did not show learning effects (suggest restriction of range)
- Trained participants performed an average of 8% more desired behaviors than control (TARGETS, n=12)

###Wilhelm (1991)
- Subset of this also presented in
- 8300 crew members from 4 airlines
- Summary article results obtained NASA/UT/LOFT survey
- No specification of skills
- Crew members value LOFT as a training technique
- Crews typically think they do a much better than average job during LOFT (self evaluation)
- Found low to moderate agreement
<table>
<thead>
<tr>
<th>Butler (1991; 1993)</th>
<th>across airlines survey, Instructor evaluation (1 organization)</th>
<th>in performance ratings between self evaluation and instructor evaluation (one organization)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>- When asked about the exhibition of specific CRM behaviors in LOFT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>large differences found across organizations in CRM behavior</td>
</tr>
<tr>
<td></td>
<td></td>
<td>patterns, although self reports all tended to be high</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Significant, but small relationship between instructor ratings</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and self ratings of specific CRM behaviors</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Over a 2 yr period at one airline CRM behaviors decreased as did</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ratings of scenario and instructor quality in LOFT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Over a 3 year period at one airline scenario, instructor, and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LOFT delivery remained constant in perceived quality and self</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reports of CRM behavior steadily increased; overall this airline</td>
</tr>
<tr>
<td></td>
<td></td>
<td>received lower ratings than others’ on several LOFT scales</td>
</tr>
</tbody>
</table>
Biographies

**Eduardo Salas**
Affiliation:
Institute for Simulation and Training
University of Central Florida
Orlando, FL 32826
and
University of Central Florida
Psychology Department
Orlando, FL 32816
Degree/Institution:
Ph.D. Industrial/Organizational Psychology, 1984
Old Dominion University, Norfolk, VA

**C. Shawn Burke**
Affiliation:
Institute for Simulation and Training
University of Central Florida
Orlando, FL 32826
Degree/Institution:
Ph.D. Industrial/Organizational Psychology, 2000
George Mason University, Fairfax, VA

**Clint A. Bowers**
Affiliation:
Team Performance Laboratory
University of Central Florida
Orlando, FL 32826
and
University of Central Florida
Psychology Department
Orlando, FL 32816
Degree/Institution:
Ph.D. Clinical and Community Psychology, 1987
University South Florida, Tampa, FL

**Katherine A. Wilson**
Affiliation:
Institute for Simulation and Training
University of Central Florida
Orlando, FL 32826
Degree/Institution:
Graduate Student
Human Factors and Applied Experimental Doctoral Program
University of Central Florida, Orlando, FL