Does Crew Resource Management Training Work? An Update, an Extension, and Some Critical Needs

Eduardo Salas, Katherine A. Wilson, and **C. Shawn Burke,** University of Central Florida, Orlando, Florida, and **Dennis C. Wightman,** Army Research Institute, Fort Rucker, Alabama

Objective: This review provides the state of crew resource management (CRM) training evaluations since the E. Salas, C. S. Burke, C. A. Bowers, and K. A. Wilson (2001) review and extends it to areas beyond aviation cockpits. Some critical evaluation needs in CRM training are also covered. **Background:** Because of the purported success of CRM training in aviation, other high-consequence domains have begun to implement CRM training for their workforces. However, the true impact of CRM training in aviation and these other domains has yet to be determined. **Method:** Using D. L. Kirkpatrick's (1976) framework for evaluating training (i.e., reactions, learning, behavior, and organizational impact), we reviewed 28 published accounts of CRM training to determine its effectiveness within aviation, medicine, offshore oil production and maintenance, shipping/maritime, and nuclear power domains. **Results:** Findings indicate that CRM training generally produced positive reactions from trainees; however, the impact of training on learning and behavioral changes suggest mixed results across and within domains. Furthermore, and as was found by Salas, Burke, et al. in 2001, we cannot ascertain whether CRM has had an impact on the organization's bottom line (i.e., safety). **Conclusion:** Based on the results, there are several critical needs that the CRM training community must address before CRM training can have the desired impact on safety: a mandate, access to data, and resources. **Application:** As CRM training expands to organizations beyond aviation, it is critical that its impact be understood such that it can be improved and achieve the intended results.

INTRODUCTION

Commercial and military aviation have been utilizing crew resource management (CRM) training for more than 2 decades. CRM is an instructional strategy that trains crews to effectively use all of their available resources (i.e., people, equipment, and information; Helmreich, Merritt, & Wilhelm, 1999). CRM training has been defined as a set of "instructional strategies designed to improve teamwork in the cockpit by applying well-tested training tools (e.g., performance measures, exercises, feedback mechanisms) and appropriate training methods (e.g., simulators, lectures, videos) targeted at specific content (i.e., teamwork knowledge, skills, and attitudes)" (Salas, Prince, et al., 1999, p. 163). So, CRM training can take many shapes and forms, but it can be conceptualized as a team training strategy focused on improving crew coordination and performance.

Despite its long history and touted success within aviation, some have begun to argue that the true impact of CRM training on aviation safety is still not understood (Salas, Burke, Bowers, & Wilson, 2001). Salas, Burke, et al. (2001) conducted a comprehensive review of studies published between 1983 and 1999 that evaluated the effectiveness of CRM training for flight crews within the commercial and military aviation domains. The review identified 58 studies suggesting that the data are encouraging, but the picture is not as clear as it should be after 20 years of implementation. CRM training was found to generally produce (a) positive reactions (i.e., affective and

Address correspondence to Eduardo Salas, University of Central Florida, Institute for Simulation and Training, 3100 Technology Parkway, Orlando, FL 32826; esalas@ist.ucf.edu. *Human Factors*, Vol. 48, No. 2, Summer 2006, pp. 392–412.

utility based), (b) enhanced learning (primarily measured through attitude change), and (c) desired behavior change in the cockpit. However, because of a lack of studies that have systematically (and directly) tested the effects of CRM training, its impact on safety was unable to be determined. Recently, other researchers drawing from the same studies reported similar results (see Edkins, 2002; O'Connor, Flin, Fletcher, & Hemsley, 2002, 2003).

CRM training became mandatory for military flight crews in the early 1990s, and although a majority of commercial airlines were voluntarily implementing CRM training, it did not become mandatory for commercial flight crews until 1998. As noted, the conclusion remains that the impact of CRM training is not understood as well as it should be after 20 years. In fact, recent work indicates that the number of accidents involving a breakdown of CRM as a causal factor has remained fairly consistent over time despite the presence of CRM training. Specifically, Wiegmann and Shappell (2000) examined a breakdown of the factors contributing to U.S. Navy and Marine Corps accidents and found that in approximately 60% of the accidents between 1991 and 2000, a CRM failure in the cockpit was a factor. Similarly, an examination of aviation accidents during 1991 and 2000 indicated that approximately 41% of Part 121 (i.e., aircraft conducting domestic commercial or flag operations) and 23% of Part 135 (i.e., aircraft conducting commuter or on-demand operations) accidents involve a breakdown in CRM by the flight crew (Wiegmann & Shappell, 2001).

The full impact of CRM training on safety cannot yet be ascertained. Despite this, it is interesting that other domains have begun to implement CRM training. For example, the medical community has implemented CRM training in the operating room, emergency room, and intensive care units. Other communities, such as nuclear power and offshore oil production, have also implemented CRM training for employees. Therefore, the purpose of our paper is threefold. First, we provide an update to the review by Salas, Burke, et al. (2001). We do this by analyzing the state of aviation CRM training evaluations that looked at the impact of CRM training on flight crews since 1999. Second, we extend the Salas, Burke, et al. (2001) review to include CRM training efforts in aviation areas beyond the cockpit (i.e., maintenance, air traffic control) and in domains outside aviation (e.g., medical, offshore oil production). Finally, we conclude with several critical CRM training evaluation needs.

AN UPDATE AND EXTENSION

A comprehensive review was conducted to uncover studies published within the aviation domain since the Salas, Burke, et al. (2001) review as well as studies outside the aviation domain that have evaluated CRM training. In an effort to identify these studies, numerous databases were searched (e.g., EBSCOhost, Medline, PsychInfo, Science Direct) using terms such as *resource management, crew resource management training, crisis resource management,* and *aircrew coordination training*. In addition, the references of each identified article were examined to identify any research that was not uncovered as a part of our search efforts. To be included within the current critical review, the paper must have presented the findings of a study evaluating the impact of CRM training on trainees' reactions, learning, or behaviors and/or its impact on the organization (see Table 1).

Although the names of the CRM training programs in domains outside of aviation sometimes differed (e.g., CRM training in health care is commonly referred to as *crisis resource management training*), we included those studies that we determined have the same underlying meaning (i.e., trained and evaluated CRM skills). Our extended effort found evaluation evidence in the areas of aviation maintenance, medicine (especially anesthesiology), and offshore oil production and maintenance. We are aware that other domains, such as fire service and railroad transportation, have also been implementing CRM training; however, no published empirical evidence regarding the evaluation of these programs was found.

Similar to the Salas, Burke, et al. (2001) effort, the current critical review employs the extended Kirkpatrick (1976) typology (see Kraiger, Ford, & Salas, 1993) as the evaluation typology guiding the analysis. Within this typology, the lowest level, reactions, examines trainees' affective feelings toward the program (i.e., did they like it?) and the utility of the program (i.e., is it viewed as worthwhile?). Learning is the second level of the typology and relates to the knowledge (i.e.,

text continues on page 401

TABLE 1: Summary of CRM Evaluative Efforts Where Training Was Provided

Source	Community	CRM Training Content	Type of Study/ Data Collection	Findings
Brun et al., 2000	Royal Norwegian Naval Academy (24 cadets)	Coping in critical situations Communication and decision making Teamwork	Experimental	*Reactions:* In general, reactions to training were positive. *Learning:* No change in shared mental model scores. No difference in shared mental model scores between trained and untrained groups.
Byrdorf, 1998, as cited in O'Connor et al., 2003	Danish shipping/ maritime (AP Moeller and MAERSK Company)	Resource management Assertiveness Communication Teamwork Stress coping	Quasi-experimental	*Organizational impact/results:* Reduction in nautical and machinery casualties since training's inception 4 years earlier. Insurance premiums decreased by 15%.
Elliott-Mabey, 1999, as cited in O'Connor et al., 2003	UK Royal Air Force (3212 air crew)	No specification of skills taught	Quasi-experimental Self-report survey (Aircrew Attitude Questionnaire)	*Learning:* Significant positive change in attitudes following initial training.
Ellis & Hughes, 1999	Australasian medical community (7 emergency medicine, 1 surgical, 2 anesthetic, 2 paramedics)	Simulations used to develop crisis management skills Training more focused on use of simulation	Quasi-experimental Not clear if given a survey or interviewed Simulator observations	*Reactions:* Trainees responded positively to the course. *Behavior:* Errors were observed during the scenarios (e.g., prolonged period of poor oxygenation before intubation; failure to consider reactions to drugs).
Fonne & Fredriksen, 1995	Norwegian maritime (432 navigators)	No specification of skills taught	Quasi-experimental Self-report survey	*Learning:* Attitude changes for trainees were overall very positive (e.g., more likely to inform other crew members of personal stress, more likely to reduce speed in adverse weather conditions). Attitude changes were stable or improved over time.
Gaba et al., 1998, 2001	U.S. medical community (anesthesiology: 37 residents, 31 faculty, 4 nurses)	No specification of skills taught Behavioral markers included communication, leadership/ followership, distribution of workload, and overall CRM performance	Quasi-experimental Videotape observations (Modified Line/LOS checklist)	*Behavior:* Individual and team performance varied. For almost all cases, team ratings were slightly better than those of individuals. Depending on the scenario, 21%–35% of primary anesthesiologists and 14%–28% of anesthesia crews scored below minimally acceptable for CRM performance.
Goeters, 2002	Eastern European airline (6 crews)	Management Communication Teamwork Decision making	Quasi-experimental Simulator observations	*Behavior:* Significant changes found in team building, providing and maintaining standards, awareness of systems, problem definition and diagnosis, option generation, situation awareness, decision making, and risk assessment and optimum selection.

Continued on next page

TABLE 1 (continued)

Source	Community	CRM Training Content	Type of Study/ Data Collection	Findings
Grubb & Morey, 2003; Grubb et al., 2002	U.S. Army (35 unit leaders, instructor pilots, and aircrews)	No specification of skills taught	Quasi-experimental Performance evaluation checklist Scenario worksheets Course critiques	*Behavior:* Results show increases in team relationships, maintaining workload levels, mission information exchange, cross-monitoring performance, managing mission-threatening error, and mission segments completed. *Organizational impact/results:* There was a decrease in the number of aircraft crashes.
Halamek et al., 2000	U.S. medical community (delivery room: 9 nurses and 29 doctors)	Provided exemplars of crisis management Training more focused on use of simulation	Quasi-experimental Self-report survey	*Reactions:* 97% agreed that the scenarios recreated real-life situations; 97% agreed that the debriefings focused on crisis management skills; 100% agreed that the debriefings allowed them to self-critique.
Harrington & Kello, 1992 (as cited in O'Connor et al., 2003)	U.S. nuclear power community (170 control room personnel)	No specification of skills taught	Quasi-experimental Self-report survey (Control Room Operators Attitude Questionnaire, modified CMAQ)	*Learning:* Attitude changes following training were positive. *Behavior:* Increased recognition of stressor effects, communication and coordination, and command responsibility.
Holzman et al., 1995	U.S. medical community (68 anesthesiologists and 4 nurse-anesthetists)	Decision making Human error Human performance	Quasi-experimental Self-report survey	*Reactions:* Course was valued highly and seen to have a potential benefit to anesthesiologists. Residents felt they "learned a lot" more than did attendings. Video debriefing was seen as a positive experience; lecture was not as well received.
Howard et al., 1992	U.S. medical community (38 anesthesiology residents and faculty)	Planning Recognition of event Leadership Communication Distribution of workload Attention allocation More... *	Quasi-experimental Self-report survey Knowledge test Videotape observations	*Reactions:* Trainees enjoyed the training. *Learning:* Residents' scores significantly improved after training; expert anesthesiologists' scores did not. *Behavior:* Trainees didn't coordinate tasks; poor communication; poor delegation of tasks by leader; leadership was assumed by less experienced person.
i Gardi et al., 2001	Danish medical community (32 nurse-anesthesiologist teams)	Leadership Communication Other human performance issues	Quasi-experimental Videotape observations	*Behavior:* A majority of teams asked for assistance during the scenarios. All teams were able to diagnose the emergency. Less than half hyperventilated the patient as they had intended to do so. Difficulties were attributed to poor resource management and not knowledge.

*The training content for this study was very extensive; only a subset of the competencies trained can be included because of space limitations.

Continued on next page

TABLE 1 (continued)

Source	Community	CRM Training Content	Type of Study/ Data Collection	Findings
Jacobsen et al., 2001	Scandinavian medical community (42 anesthetists)	Awareness Leadership Communication Delegation Cooperation Declaration Reevaluation Allocation Help hands Help competence Start of initial treatment	Quasi-experimental Videotape observations	*Behavior:* Trainees had difficulty diagnosing and treating problem. Trainees showed high awareness. No trainees took the lead or delegated tasks. Lots of communication.
Katz, 2003	U.S. Army (16 rotary wing aircrews)	No specification of skills taught	Quasi-experimental Simulator observations	*Behavior:* Positive changes occurred in the crews' ability to establish and maintain team relationships and workload levels, exchange mission information, and cross-monitor performance, and in all ACT behaviors. Following training, crews' mission-threatening error rating improved, the number of crews receiving S or S+ ratings improved, and the number of crashes declined.
Kurrek & Fish, 1996	Canadian medical community (35 anesthetists)	No specification of skills taught in workshop	Quasi-experimental Self-report survey	*Reactions:* Overall, participants enjoyed the workshop and were positive about the implications of the workshop to anesthetists. Respondents reported that the new ACRM knowledge would be helpful to them. Many believed the course would help them be safer and should be offered regularly.
Lainos & Nikolaidis, 2003	Greek airline (cockpit and cabin crew members; no specification of number of participants	No specification of skills taught	Quasi-experimental Self-report survey	*Reactions:* 78% of trainees agreed that the course content is appropriate to flight crew job needs; 71% reported approval of the quality of training methods and procedures; 82% approved the quality of the trainer. *Learning:* 86.6% of cockpit crew view their role as authoritarian. Only 10% of captains and 18% of FO see selves as member of team/flight crew. Only 20% of cockpit crew and 23% of cabin crew reported an intention to cooperate together during duty time.

Continued on next page

TABLE 1 (continued)

Source	Community	CRM Training Content	Type of Study/ Data Collection	Findings
Morey et al., 2002	U.S. medical community (emergency departments at 9 teaching and community hospitals: 684 physicians, nurses, technicians)	48 teamwork behaviors, but no specification of skills taught	Quasi-experimental Self-report survey On-the job observation	*Learning:* Trainees' attitudes toward teamwork significantly improved. *Behavior:* Teamwork improved significantly. Clinical error rate was reduced significantly; the observed error rate did not change.
Nullmeyer et al., 2003; Spiker, Wilson, & Deen, 2003	U.S. Air Force (20 C-130 aircrews)	Situation awareness Crew coordination Communication Risk management/ decision making Task management Mission evaluation	Quasi-experimental Simulator observations (C-130 CRM Process Worksheet)	*Behavior:* The better performing crews demonstrated CRM behaviors (e.g., open communication, feedback, planning) more regularly than did other crews. CRM quality was correlated with mission performance. Crew CRM and mission performance were lower during mission execution than during planning.
O'Connor & Flin, 2003; Flin, O'Connor, Mearns, & Gordon, 1999	UK offshore oil community (3 North Sea production platforms: 77 participants)	Situation awareness Decision making Communication Teamwork Supervision/ leadership Personal resources	Quasi-experimental Self-report survey	*Reactions:* Positive reactions shown toward training. Training may be beneficial to offshore oil and gas production industry. Wanted more emphasis on CRM skills rather than theory behind them. *Learning:* Knowledge regarding human factors did not increase following training. More positive attitudes were found toward decision making and personal limitations but not toward CRM concepts.
O'Donnell et al., 1998	U.S. medical community (34 anesthesia nurses)	No specification of skills taught	Quasi-experimental Self-report survey	*Reactions:* Results suggested that participants' experiences were positive and that the biggest strength was the realism of the scenarios.
Small et al., 1999	U.S. medical community (15 MedFlight nurses and paramedics)	No specification of skills taught	Quasi-experimental	*Reactions:* Participants found course to be enjoyable and intense and that it suspended disbelief easily.

Continued on next page

TABLE 1 (continued)

Source	Community	CRM Training Content	Type of Study/ Data Collection	Findings
Spiker, Tourville, Bragger, Dowdy, & Nullmeyer, 1999	U.S. Air Force (16 C-5 aircrews)	Situation awareness Crew coordination/flight integrity Communication Risk management/ decision making Task management Mission evaluation	Quasi-experimental C-5 CRM Process Worksheet to rate CRM behaviors	*Behavior:* The best crews prioritized events, accomplished planning early, had a strong functional leader who emerged early in the process, and continually and directly exchanged information between one pilot and one FE. For all crews, a majority of discussions focused on performance rather than issues related to the process of CRM. Successful crews were more critical of themselves and identified areas in need of improvement during discussions. Successful crews maintained an atmosphere for open information exchange. Overall CRM process ratings were substantial and significant. A strong, positive relationship was found between CRM and performance.
Taylor (1998, 2000)	U.S. aviation maintenance (2 airlines: Company A, 6200+ mechanics; Company D, 2400+ maintenance employees and managers) Training conducted 1998–1999	No specification of skills taught across airlines	Quasi-experimental Self-report survey (MRM/ Technical Operations Questionnaire) Longitudinal data (before, during, and after training and +2 and +6 months posttraining)	*Reactions:* Enthusiasm toward MRM training was high regarding its effect on safety and teamwork and usefulness of training at both airlines. Enthusiasm diminished over time at Company A but remained relatively stable at Company D. In Company A, frustrations toward MRM training were seen after training because of management's unwillingness to open further communication channels for safety and pace of change in coworkers. *Learning:* Attitudes toward shared decision making, communication, coordination, and stress management increased in Companies A and B. Attitudes toward assertiveness improved at Company D while diminished at Company A. Company A participants' expectations to change were below normal, however; of those with intentions to use what they learned, only half stated they would actively (vs. passively) engage with others to improve safety. Intentions to change at Company D were at or above normal, and only half stated they would actively engage with others to improve safety.

Continued on next page

TABLE 1 (continued)

Source	Community	CRM Training Content	Type of Study/ Data Collection	Findings
				Behavior: Approximately 60% of mechanics at Company A said they had made changes on the job. There was an increase in the number of employees at Company D who reported that "training changes my behavior" between Phases 1 and 2 of training. Three times as many employees at Companies A and D who reported interacting with others said they used passive vs. active tactics during those interactions regarding safety issues. *Organizational impact/results:* In both airlines, aircraft ground damage was reduced. In Company A, lost-time injuries improved. In Company D, written communication did not improve following training.
Taylor & Thomas, 2003	U.S. aviation maintenance (subject site: 250+ repair mechanics, quality inspectors, immediate supervisors, and middle managers from one airline; compared results with maintenance facility at same airline [800+] and 2 other airlines [1800+ and 150+]). Training conducted during 2000	Written communication Written turnover No other specification of skills taught	Quasi-experimental Self-report survey (Maintenance Resource Management/Technical Operations Questionnaire) Interviews Nonroutine work cards Turnover data	*Learning:* When asked what is remembered best about training following Phase 1, respondents answered "improve turnovers," "write more clearly," and "communication." No mention of turnover following Phase 2 or at other airlines. Respondents at subject site expressed intentions to write more clearly and improve turnover more frequently following Phase 1 than Phase 2. *Behavior:* Increase in turnover completeness (length) between Phases 1 and 2 of training; however, there is a decline following the completion of training. Legibility increased throughout training but returned to baseline after. Written quality improved but diminished over time. When asked what changes have been made on the job 2 months after training, respondents from the subject site and facility at same airline showed slight application of what was learned on the job; however, the percentages were very small. There was no clear effect that MRM training eliminated prescriptive turnovers. *Organizational impact/results:* Paperwork errors showed no subsequent or long-term effects and increased when inexperienced personnel were hired.

Continued on next page

TABLE 1 (continued)

Source	Community	CRM Training Content	Type of Study/ Data Collection	Findings
Taylor et al., 1993; Robertson & Taylor, 1995	U.S. aviation maintenance (2200+ from management level at one airline) Training conducted during 1991	Teamwork Decision making Assertiveness Stress management Interpersonal skills	Quasi-experimental Longitudinal (before and after training and +2, +6, and +12 months posttraining) Self-report survey (CRM/Technical Operations Questionnaire)	*Reactions:* Managers believed their behaviors would moderately to largely change. Majority (90%+) rate training as "very useful" or "extremely useful." Majority (96%+) enjoyed training and would rate it as one of the best training courses at airline. 40%+ of managers desire more training and think it is needed. *Learning:* Positive attitude changes occurred after training toward command responsibility, communication, and coordination, and recognizing stressor effects. No changes were found 2 months later. No difference found for willingness to voice disagreement (i.e., assertiveness). A significant increase was found 2 months later. *Behavior:* Better listening by participants indicated a slight decrease over 12 months following training. Awareness of others remained relatively stable over time. Ability to deal with others shifted from passive to active, indicating that managers are applying learned skills on the job over 12 months following training. *Organizational impact/results:* Ground damage incidents decreased during the posttraining year. Dependability increased during the posttraining year. Rate of lost time attributable to injuries has decreased since training.
Thompson et al., 1999	U.S. Air Force (16 MH-53J aircrews)	Evaluated relationship between CRM and performance during simulator refresher training	Quasi-experimental Simulator observations (CRM worksheet)	*Behavior:* Highest rated crews demonstrated good mission evaluation, task management, situation awareness, crew coordination, communication,risk management, and tactics employment. Overall CRM and performance showed positive and significant correlation.
Woldring & Isaac, 1999, as cited in O'Connor et al., 2003	European air traffic control (126 participants from 7 countries)	Teamwork Team roles Communication Situation awareness Decision making Stress	Quasi-experimental Self-report survey (modified FMAQ)	*Reactions:* Positive reactions were generally reported. *Learning:* A significant positive change in attitudes was found.

Note. LOS = line-oriented simulation; ACT = aircrew coordination training; FO= first officer; FE = flight engineer; MRM = maintenance resource management.

principles, facts, and skills) understood and acquired by trainees. In addition, learning is measured in terms of a desired change in trainees' attitudes toward CRM. The next level, behaviors, assesses whether the knowledge learned during the training is transferred into actual behaviors on the job or in a simulation. Finally, the highest level of Kirkpatrick's (1976) typology is the impact of training on the organization. Evaluating training at this level determines whether the training had an impact on the goals of the organization (e.g., improved safety, error reduction, increased productivity). Each of these levels can be incorporated into a training evaluation program by itself or in combination with one or more of the other levels. A recent study by O'Connor et al. (2002) investigated the typical methods used to evaluate CRM training's effectiveness in the UK's aviation industry. Their research found that 69% evaluated training in terms of reactions, 21% measured learning in terms of attitudes, 36% measured learning in terms of knowledge tests, 53% measured behaviors, and 33% evaluated the training's impact on the organization. These numbers are relatively consistent with what is published in the literature, as we will present next.

FINDINGS

Based on the literature available that met our search criteria, we identified 28 new studies that evaluated the effectiveness of a CRM training program administered as a part of the research (see Table 1). Table 1 provides a description of each study as well as the findings. The 28 identified studies broke down by the following domains: commercial aviation (2 studies), military aviation (7 studies), medicine (11 studies), offshore oil production and maintenance (1 study), aviation maintenance (3 studies), shipping/maritime navigators (2 studies), nuclear (1 study), and air traffic control (1 study). Using the same framework used in the previous review, we followed Kirkpatrick's (1976) training evaluation typology for organizing the results of the analyzed studies. We should note that although all of these papers stated that a CRM training program had been implemented, many of them did not specifically state the CRM skills that were taught. As such, this makes it difficult to truly know why a particular CRM training program was a success or why it was not.

Do Trainees Like CRM Training? Reactions Evidence

Of the 28 studies available in the literature, 13 (46%) evaluated CRM training at the reactions level (i.e., attitudes), or the lowest level of Kirkpatrick's (1976) typology. Collection of data at this level seeks to determine if trainees liked the training and if they found it useful. Of these 13 studies, 5 collected data solely at the reactions level. A majority of these data was collected using self-report surveys, including the Cockpit Management Attitudes Questionnaire (CMAQ; Helmreich, 1984) or modified versions for the respective communities (e.g., operating room management attitudes questionnaire). Similar to the Salas, Burke, et al. (2001) review, all of these studies found positive reactions toward the training from participants. For example, trainees reported that training was enjoyable (O'Donnell, Fletcher, Dixon, & Palmer, 1998), that it suspended disbelief (Small et al., 1999) and that they believed it to be useful to them (Kurrek & Fish, 1996). This information suggests that training was liked and will be useful to participants on the job, but it is minimally informative as to whether trainees learned from the training, whether they applied what they learned, or whether the training improved organizational outcomes. These findings will be addressed in the next few sections.

Do Trainees Learn From CRM Training? Learning Evidence

Of the 28 studies, 12 (43%) were found to have evaluated CRM training at the learning level. Furthermore, 10 of those 12 studies were a part of a multilevel evaluation study. Data collected at this level include not only whether trainees learned the knowledge, skills, and attitudes (KSAs) taught but also the extent to which the training led to a desired change in trainees' attitudes (i.e., positive change toward CRM). Learning evidence was found in eight communities (i.e., medical, aviation maintenance, military and commercial aviation, nuclear, air traffic control, maritime, and offshore oil production), and the results were not consistent. The 2 studies that collected data solely at the learning level found that changes in attitudes toward CRM concepts improved. For example, Fonne and Fredriksen (1995) found that changes in trainees' attitudes were overall very positive

and that they were now more likely to change their behaviors, such as informing other crew members of personal stress and reducing aircraft speed in adverse weather conditions. Whereas the 2 studies that examined only learning evidence indicated a positive change in attitudes, studies that conducted multilevel evaluations (including learning evidence) indicated more mixed results. Specifics on these multilevel studies and others will be discussed in a later section.

Do Trainees Apply the Learned CRM Behaviors? Behavioral Evidence

Behavioral evidence was collected in 16 of the 28 studies obtained (57%). Of these, 8 collected data solely at the behavioral level of Kirkpatrick's (1976) typology. Behavioral evidence provides an assessment of whether the KSAs learned in training transfer to actual behaviors on the job or in a simulated environment. Behavioral data were collected for nearly all of the 16 studies through simulation observations. Unlike the consistently positive results found by those studies that looked solely at reactions or learning, the results were not as conclusive for the studies that examined only behaviors. Half of those 8 studies suggest mixed results – both positive and negative transfer of behaviors were found. For example, Gaba, Howard, Fish, Smith, and Sowb (2001) and Gaba et al. (1998) found that although teams on average performed better than individuals, approximately one quarter of trained teams scored below the minimally acceptable level. Additionally, in 1 training program, some skills were observed to transfer (e.g., communication) whereas others were not (e.g., leadership; Jacobsen et al., 2001). The authors argued that the increased communications were attributable to a lack of leadership, requiring the team to communicate more in order to solve the problem. Similarly, research by Nullmeyer, Spiker, Wilson, and Deen (2003) suggests that CRM training leads to positive behaviors during some tasks (i.e., planning) and less-than-positive behaviors in others (i.e., mission execution). Although mixed results were found for some studies, 4 also found positive results (e.g., Goeters, 2002; Thompson, Tourville, Spiker, & Nullmeyer, 1999). More research is needed to determine why this may be the case. Potential reasons for these inconsistencies are examined later, when we take a closer look at CRM training in the medical community.

Are the Organizations Safer? Results/Organizational Impact Evidence

Similar to findings of the Salas, Burke, et al. (2001) review, the number of evaluations conducted at this level was small. Specifically, evidence of the impact of CRM training on the organization was examined in 5 of the 28 studies (18%) analyzed (see Table 1). Data collected at this level (considered the highest level of evaluation) provide evidence of the impact of training on the safety of the organization. Only 1 of the 5 studies examined organizational results in isolation (i.e., Byrdorf, 1998, as cited in O'Connor et al., 2003). The remaining studies looked at organizational results in combination with other levels of Kirkpatrick's (1976) typology. The results of these and other multilevel evaluations will be discussed in the next section. The 1 study that investigated solely organizational results found that following CRM training, there was a reduction in the number of nautical and machinery casualties and insurance premiums. This study (and others discussed later) indicates a positive effect of CRM training; however, given the low occurrence of accidents in many of these domains and the difficulty in collecting these data (e.g., time, resources), we caution researchers when concluding that CRM training alone led to these changes. Potential effects of extraneous variables may have also had an influence.

Multilevel Evaluation Evidence

Surprisingly, 12 of the 28 studies (43%) evaluated training using multiple levels of Kirkpatrick's (1976) typology. This is similar to the findings by Salas, Burke, et al. (2001) in which 41% of the studies reviewed examined training at multiple levels. The most common multilevel study looked at the reactions and learning of trainees (4 studies) – the two lowest levels of Kirkpatrick's (1976) typology. The remaining studies evaluated training at various levels – specifically, reactions and behavior (Ellis & Hughes, 1999); learning and behavior (Harrington & Kello, 1992, as cited in O'Connor et al., 2003; Morey et al., 2002); reactions, learning, and behavior (Howard, Gaba, Fish, Yang, & Sarnquist, 1992); behavior and organizational impact (Grubb & Morey, 2003);

learning, behavior, and organizational impact (Taylor & Thomas, 2003), and reactions, learning, behavior, and organizational impact (Taylor, 2000; Taylor, Robertson, Peck, & Stelly, 1993). The dispersion of multilevel studies across domains was fairly consistent: medical community (3 studies), commercial aviation (1 study), aviation maintenance (3 studies), military aviation (2 studies), air traffic control (1 study), nuclear (1 study), and offshore oil production (1 study). Somewhat disappointingly, the aviation community (commercial and military) decreased the percentage of multilevel studies it has conducted since the Salas, Burke, et al. (2001) review (33% vs. 41%).

When taking a closer look at the results of the multilevel evaluations published in the literature, we found inconsistent results indicating only partial support for the effects of CRM training. Of the 12 studies, 4 found positive results at all levels evaluated, whereas the remaining 8 found mixed results. What we mean by this is that within the same study, both positive and negative or neutral results were found (e.g., positive reactions but negative behavior transfer). We discuss the results of several of these studies next.

As previously stated, four studies found positive outcomes across all levels evaluated. For example, Grubb, Crossland, and Katz (2002) and Grubb and Morey (2003) found that there was an increase in CRM-related behaviors (e.g., cross-monitoring, information exchange) and also a decrease in the number of aircraft accidents. Harrington and Kello (1992, as cited in O'Connor et al., 2003) examined learning and behavioral changes of nuclear control room personnel. They found that attitudes toward CRM were more positive following training and that trainees showed an increased recognition of CRM concepts (e.g., communication, coordination) on the job. We should mention, however, that the information contained within much of the research discussed here did not provide detailed information pertaining to the robustness of the methodologies used (e.g., what was trained, how it was trained, number of participants). Therefore, we must "take a leap of faith" in suggesting that there is a connection among CRM training, learning, and performance.

In all but one of the remaining multilevel evaluations, reactions to training were positive, indicating that trainees liked training. Despite these favorable reactions, results of learning or behavioral changes were not as clear cut. For example, some research found that although trainees reported positive attitudes towards the training, the behaviors did not transfer to a simulated environment (Howard et al., 1992) or errors were observed during simulations (Ellis & Hughes, 1999), suggesting that the training was not as successful as anticipated. In another study, reactions were positive but there was no indication of learning, and trained groups received scores similar to those of nontrained groups (i.e., neutral results). The two studies that evaluated training at all four levels found that trainees enjoyed the training (reactions), attitudes toward CRM concepts were more positive (learning), learned behaviors were transferred to the work environment (behavior), and ground damage incidents were reduced (organizational impact; Taylor, 2000; Taylor et al., 1993). However, even these studies showed that the positive attitudes toward CRM declined when not supported by management following training (Taylor, 2000) and that behaviors regressed toward pretraining levels (Taylor et al., 1993). As a final example of these mixed results, Taylor and Thomas (2003) found that even though trainees learned from training and transferred learned behaviors to the job (e.g., improved written communications), positive behaviors shifted toward pretraining levels in the months following training. This indicates that training did not have a lasting effect on trainees.

The studies discussed here reinforce the need to conduct multilevel evaluations so that a complete picture can be obtained of CRM training's effectiveness and to provide support for recommendations for recurrent CRM training. These findings reinforce the notion that positive attitudes do not necessarily lead to learning and learning does not necessarily result in a behavioral change (Alliger, Tannenbaum, Bennett, & Traver, 1997). The inconsistencies of results found in the studies discussed in this section suggest the value of evaluating training at multiple levels to truly understand whether or not CRM training was effective.

Post Hoc Studies

Our review also uncovered six post hoc studies that examined the effectiveness of CRM training. These studies did not actually set out to evaluate CRM training; rather, the researchers examined

the impact of a previously administered CRM training program on participants' reactions, learning, and performance (none of these studies looked at the organizational impact; see Table 2). Overall, the findings of these studies were similar to those discussed previously: Reactions were positive (e.g., Small, 1998) and performance improved (e.g., Fisher, Phillips, & Mather, 2000). Another study examined how culture affects CRM training (Davis, Bryant, Liu, Tedrow, & Say, 2003). These researchers compared homogeneous and heterogeneous cultural teams and found that American teams performed better than Chinese teams (e.g., more decisions, fewer communication errors) and that both American and Chinese homogeneous teams communicated more frequently than did heterogeneous teams. In a final study, Karp, Condit, and Nullmeyer (1999, 2000) found some less-than-positive results. They examined trainees' reactions and learning to CRM training for single-seat pilots, who must interact with team members outside the cockpit. Their results indicate that attitudes toward CRM were negative and that trainees found it difficult to see the relevance of CRM training to their operations. The results of the latter two studies emphasize the importance of adapting the training to the audience, whether multicultural or distributed.

Summary of Evidence

The Salas, Burke, et al. review (2001) suggested that CRM training in the aviation community led to positive attitudes, learning, behavioral changes on the job, and (potentially) safety in organizations, although the results were not quite that clear. When an examination of the impact of CRM training in all communities is made (as in the current effort), the results are even less clear. Although we did not uncover any study that suggests CRM training does not work, approximately half of the studies indicated mixed results, leading us to question its effectiveness (see next section). In addition, the amount of detailed information regarding the training (e.g., skills trained, length of training) is often limited within the published literature. This makes it difficult to assess the quality and suitability of the training program.

Overall, it appears that CRM training has had a positive impact on attitudes in numerous domains, as indicated by the 13 studies that evaluated trainees' reactions. As to whether learning took place during training, those studies that examined trainees' attitude changes suggest that learning did occur for the most part. Six of eight studies indicated positive changes in attitudes towards CRM. However, studies that looked at actual learning of trained competencies indicated some inconsistencies. Three of five studies examining trainees' knowledge following training showed mixed results (all found positive as well as negative or neutral results). For the remaining two studies, one indicated neutral outcomes, whereas the other found negative outcomes. Likewise, behavioral evidence across communities suggests inconsistencies as well. Of those studies examining solely behavioral evidence, half indicated positive results and the remaining half indicated mixed results. When behavioral evidence was collected in conjunction with additional levels, the data suggest that three of eight studies found either negative or mixed results. The remaining five studies found positive changes in behaviors. Finally, three of the five studies that evaluated training at the organizational level did suggest that CRM training had a positive impact on safety in terms of reduced errors or incidents. Two studies indicated neutral (no change) or mixed results following training. Given the difficulty of evaluating training at this level (e.g., criterion measures are difficult to identify and it is hard to control the influence of extraneous variables), it cannot be said with certainty that CRM training led to the positive results. See Table 3 for a breakdown of which studies found positive, negative, or mixed results.

The variability of much of these results leads us to argue (again) for the need to evaluate CRM training at multiple levels. This is necessary because data from just one level provides a rather limited picture of the true impact of CRM training on trainees and the organization. More specifically, although reaction data is worthwhile in that it provides a picture of whether trainees liked the training and found it useful, it does not indicate whether the trainees learned the competencies trained. Furthermore, we would argue that just because trainees learned the competencies or had a positive change in attitudes, it does not signify a behavioral change on the job. Finally, a change in trainees' behaviors will not necessarily lead to a change in the safety of the organization. By collecting evidence at multiple levels, researchers can begin to get a clearer picture as to where CRM

text continues on page 408

TABLE 2: Summary of Post Hoc CRM Training Analyses

Source	Community	Type of Study/ Data Collection	Findings
Davis et al., 2003	Undergraduate and graduate students (196 Chinese and American students)	Post hoc Videotape observations	*Behavior:* Cultural differences were found for several team behaviors: American teams had higher situation awareness, made more decisions, displayed more coordination, and had fewer communication errors. Homogeneous American and Chinese teams communicated more frequently than did heterogeneous teams.
Fisher et al., 2000	U.S. air medical community (108 crew members)	Post hoc Survey Observations	*Behavior:* Increased crew awareness. Increased promotion of team concepts.
Karp et al., 1999, 2000	U.S. Air Force (36 F-16 pilots)	Post hoc Interviews	*Reactions:* Many liked the case studies and interactive computer-based training but did not like the PC-based situational trainer. The value of the CRM training program was not evident and many did not understand the full breadth of the training. Over 85% of respondents reported a neutral, mixed, or less-than-positive experience with CRM training. *Learning:* Majority of pilots feel they are already doing CRM, just with another name, so resist "CRM." Attitudes toward CRM can be seen as negative because many feel they are forced to do CRM. Pilots tended to have a good working knowledge of mission planning/debrief, flight coordination, communication, roles of flight leader, task management, and situational awareness. Most pilots had a broad understanding of risk management/ decision making but were unclear of when to seek assistance.
Orasanu, Murray, Rodvold, & Tyzzer, 1999	U.S. airline (383 captains, first officers, and flight engineers)	Post hoc Self-report surveys (assertiveness questionnaire; intervention basis questionnaire)	*Behavior:* Captains report that they are more likely to intervene in the situations presented. If a situation has procedural/safety concerns, all flight crew reported that they would intervene. Flight crew position makes a difference as to whether or not they would intervene in situations governed by CRM issues.
Povenmire, Rockway, Bunecke, & Patton, 1989	U.S. Air Force (B-52 pilots; no specification of number of participants)	Post hoc Self-report surveys ("Today's Flight," CMAQ, and Flight Crew) Simulator observations (Line/ MOST worksheet)	*Reactions:* Sample was too small to provide any useful information. *Behavior:* A wide range of CRM skills were exhibited. Confirmation and challenge of information led to improved situation awareness (SA). SA appeared to be a major factor in performance. Fewer conflicts resulted in better performance.
Small, 1998	U.S. medical community (50 anesthesia residents)	Post hoc Interviews	*Reactions:* Residents had a positive experience. They saw the relevance of ACRM to clinical practice.

Note. Line/MOST = mission-oriented simulator training; ACRM = advance crew resource management.

TABLE 3: Abbreviated Findings of CRM Evaluative Efforts

Source	Community	Type of Study/ Data Collection	Reactions	Learning	Behavior	Organizational Impact
Brun et al., 2000	Royal Norwegian Naval Academy	Experimental	+	–		
Byrdorf, 1998, as cited in O'Connor et al., 2003	Danish shipping/maritime	Quasi-experimental				+
Davis et al., 2003	Undergraduate and graduate students	Post hoc			+/–	
Elliott-Mabey, 1999, as cited in O'Connor et al., 2003	UK Royal Air Force	Quasi-experimental		+		
Ellis & Hughes, 1999	Australasian medical community	Quasi-experimental	+		–	
Fisher et al., 2000	U.S. air medical community	Post hoc			+	
Fonne & Fredriksen, 1995	Norwegian maritime	Quasi-experimental		+		
Gaba et al., 1998, 2001	U.S. medical community	Quasi-experimental			+/–	
Goeters, 2002	Eastern European airline	Quasi-experimental			+	
Grubb & Morey, 2003; Grubb et al., 2002	U.S Army	Quasi-experimental			+	+
Halamek et al., 2000	U.S. medical community	Quasi experimental	+			
Harrington & Kello, 1992, as cited in O'Connor et al., 2003	U.S. nuclear power community	Quasi-experimental		+	+	
Holzman et al., 1995	U.S. medical community	Quasi-experimental	+/–/~			
Howard et al., 1992	U.S. medical community	Quasi-experimental	+	+/–	–	
i Gardi et al., 2001	Danish medical community	Quasi-experimental			+/–	
Jacobsen et al., 2001	Scandinavian medical community	Quasi-experimental			+/–	
Karp et al., 1999, 2000	U.S. Air Force	Post hoc	+/–	+/–		
Katz, 2003	U.S. Army	Quasi-experimental			+	
Kurrek & Fish, 1996	Canadian medical community	Quasi-experimental	+			
Lainos & Nikolaidis, 2003	Greek airline	Quasi-experimental	+	–		
Morey et al., 2002	U.S. medical community	Quasi-experimental		+	+/–	

Continued on next page

TABLE 3 (continued)

Source	Community	Type of Study/ Data Collection	Reactions	Learning	Behavior	Organizational Impact
Nullmeyer et al., 2003; Spiker et al., 2003	U.S. Air Force	Quasi-experimental			+/–	
O'Connor & Flin, 2003; Flin et al., 1999	UK offshore oil community	Quasi-experimental	+/~	~/–		
O'Donnell et al., 1998	U.S. medical community	Quasi-experimental	+			
Orasanu et al., 1999	U.S. airline	Post hoc			+	
Povenmire et al., 1989	U.S. Air Force	Post hoc	~		+	
Small, 1998	U.S. medical community	Post hoc	+			
Small et al., 1999	U.S. medical community	Quasi-experimental	+			
Spiker et al., 1999	U.S. Air Force	Quasi-experimental			+/–	
Taylor, 1998, 2000	U.S. aviation maintenance	Quasi-experimental	+	+/– (overall inconclusive)	+/–	+/~
Taylor & Thomas, 2003	U.S. aviation maintenance	Quasi-experimental		+/– (overall inconclusive)	+/– (overall inconclusive)	~/–
Taylor et al., 1993; Robertson & Taylor, 1995	U.S aviation maintenance	Quasi-experimental	+	+/~	+/–	+
Thompson et al., 1999	U.S. Air Force	Quasi-experimental			+	
Woldring & Isaac, 1999, as cited in O'Connor et al., 2003	European air traffic control	Quasi-experimental	+	+		

Note. The positive (+), negative (–), and neutral (~) conclusions indicated were determined based on our best judgment of the information contained within the publications that we obtained.

training was a success and where it might have gone wrong so that they can correct it in the future.

CRM TRAINING IN HEALTH CARE

As about half of the new studies examined evaluated CRM training in the health care community, we wanted to further analyze these to determine their effectiveness. Of the 11 studies conducted in the health care community, 8 evaluated training at only one level, namely reactions or behaviors. The remaining 3 studies used a multilevel approach to evaluating training. Overall, the 7 studies that collected reactions data suggest that the training was realistic (e.g., Halamek et al., 2000) and was seen as having a benefit (e.g., Holzman et al., 1995). Only 2 of 11 studies looked at learning evidence and were a part of multilevel evaluations. One study that examined trainees' learning found positive results in that trainees' attitudes toward teamwork improved, as did the demonstration of teamwork behaviors on the job (Morey et al., 2002). The second study showed that whereas residents improved their knowledge, expert anesthesiologists did not; however, this may be attributable to ceiling effects (Howard et al., 1992). Trainees from this study also performed poorly in a simulated environment (e.g., poor communication and coordination). These findings occurred despite the positive reactions to the training reported by trainees.

The evidence reported for the remaining four studies that examined trainees' transfer of the learned behaviors to the job or simulated environment also provides only partial support for the effectiveness of CRM training. For example, i Gardi, Christensen, Jacobsen, Jensen, and Ording (2001) found that although trainees were able to diagnose the emergency in a simulation, less than half carried out their intentions to hyperventilate the patient. In another study, although trainees showed a high awareness and communicated frequently, they had a difficult time diagnosing the problem, and no team member took the lead or delegated tasks (Jacobsen et al., 2001). Others found that transfer of trained behaviors was based on a particular simulated scenario (Gaba et al., 2001). Finally, no studies were found from the medical community that examined the impact of CRM training on the organization.

So what can be said about the findings from the medical community? Overall, it appears that there is only partial support for training's effectiveness. Whereas reactions to training were positive, transfer of the learned behaviors to the job were somewhat less concrete. Because learning evidence was not collected for a majority of the studies, it is difficult to determine whether the trained knowledge and behaviors were even learned. Furthermore, only 3 of 11 studies (27%) examined training at multiple levels. It is admirable that the medical community is implementing CRM training in health care. It is needed. However, the results of their evaluation efforts are not as positive as one would wish. These results are indicative of "newcomers" to the area of crew (or crisis) resource management, as the aviation community also struggled in the beginning. The fact that CRM training in health care is still in its infancy leads us to believe that it can be a success – if designed, implemented, and managed properly. Therefore, we hope that the health care community, along with other newcomers, will rely on what is known about the science of training (e.g., team training, transfer of training) and the lessons learned from the aviation and military communities (see Salas, Bowers, & Edens, 2001; Salas & Cannon-Bowers, 2001).

WHERE TO GO FROM HERE? SOME NEEDS

It appears that after two decades of CRM training research and practice, the industry continues to struggle with how to evaluate and institutionalize it. Since the inception of CRM training, a great deal has been learned about how to design, develop, implement, and evaluate training programs in organizations (see Salas & Cannon-Bowers, 2000, 2001; Tannenbaum & Yukl, 1992). What is perplexing is that despite what is known about the science of training, organizations seem to ignore the available relevant literature (see Salas, Fowlkes, Stout, Milanovich, & Prince, 1999) that could guide and manage their CRM training efforts. That is, the explanation, application, and integration of what has been learned about the science of training needs to make its way to the design, implementation, evaluation, and institutionalization of CRM training programs (of course, there are some exceptions). But why is this often not the case? We deal with that next.

Based on our review of the available CRM training evaluation literature, we believe a number of critical factors appear to be causing the CRM community to struggle. Although we recognize that some organizations (e.g., aviation) are further along than others (e.g., railroad) in this regard, more work still needs to be done. First, the community is lacking standardization for CRM training (e.g., what to train, how to train it). While governing agencies, such as the Federal Aviation Administration (FAA) and Civil Aviation Authority (CAA), provide resources to help organizations develop CRM training programs, recommendations are merely provided. In addition, the various names associated with CRM training (e.g., air crew coordination training, crisis resource management) indicate the lack of consensus among domains as to how to label or define CRM. Furthermore, there is no standardization as to what competencies (i.e., knowledge, skills, and attitudes) are to be trained. Our review of the studies reported here illustrates the variability in what is trained under the term *CRM* – from attention allocation to communication to flight integrity. The CRM community needs to identify what set of core CRM competencies are necessary (allowing for adaptation of unique skills in some domains) for CRM to be a success (some may already have an idea), and those should be the core skills that are trained.

Second, there is a performance measurement problem. The tools used to assess performance in learning simulations must be diagnostic of what the program is training and must be tied to the learning objectives. The tools being used now provide static information but are limited at allowing proper diagnosis of the required dynamic teamwork (CRM) competencies. The diagnosis needs to be rich, detailed, relevant, and robust so that appropriate remediation can be done. Without this the trainees, and therefore the training programs, cannot be improved. Therefore, much better mechanisms are needed to diagnose CRM-related performance.

Third, there are a number of myths and misconceptions that organizations may fall prey to as they design and implement CRM training (see Salas, Wilson, Burke, & Bowers, 2002). For example, it is often assumed that high physical fidelity simulations are better for learning. However, research suggests that high physical fidelity is not always needed, as long as the psychological fidelity is high (i.e., the simulation allows trainees to progress through the simulation using the same cognitive processes as required in the real-world environment; Bowers & Jentsch, 2001). More is not necessarily better. Another myth discussed by Salas and colleagues (Salas et al., 2002) is the misperception that subject matter experts should drive the design of training. Learning is a behavioral and cognitive event. Although subject matter experts can articulate task requirements, learning experts should drive the design of training – a partnership is indeed needed. These and others myths need to be avoided. Finally, CRM training supporters continuously struggle to make a "business case" to the upper echelons of the organization as to the importance of CRM training. So, how can researchers move ahead and show the value of CRM training? We submit that this community needs a mandate, access to data, and resources to determine the efficacy of CRM training.

A Mandate

We need a mandate to compel researchers to conduct sound, systematic, reliable, and robust evaluations. In other words, a directive from the organizations, institutions, or agencies that are designing, delivering, and implementing CRM training is needed. This direction should ensure that the training is evaluated at multiple levels, on a continuous basis, and in a standardized way. Without this mandate, CRM training will continue to be implemented as merely a "check in the box" without necessarily utilizing what is known about the proper design, development, implementation, and evaluation of training programs. Whereas the European Joint Aviation Authority, for example, requires that CRM skills be evaluated by an approved methodology (e.g., NOTECHS; Flin et al., 2003), this is not the case in domains newer to the CRM community (e.g., medical, nuclear power). Therefore, what is needed is an organization in each domain that cares to know whether or not CRM training works and will issue the mandate to find out.

Access to Participants In Situ

Once a mandate is established, access must be granted to collect the data. This requires a change in climate such that, for example, the aviation community opens its cockpits and that the medical

community opens its operating rooms to researchers. Furthermore, these communities must make the trainees (e.g., pilots, doctors, nurses, technicians) and supervisors available such that the short-term and long-term effects of CRM training can be determined. One of the things lacking within the CRM data is a look at the long-term impact of CRM training on the job (e.g., do attitudes, knowledge, and behaviors maintain over time?), not just in a simulated environment. This includes administering surveys to evaluate attitudes (e.g., CMAQ); conducting behavioral observations in the field (e.g., line operational safety audits); and evaluating accident, incident, and error reporting databases (e.g., Aviation Safety Reporting System; UK National Reporting and Learning System in health care) that are internal and external to organizations.

However, opening the cockpit and operating rooms is not enough. It is necessary but not sufficient to conduct a credible and defensible evaluation protocol. Some basics are needed: sufficient power, control groups, and randomization. More important, the right data need to be collected at the right time to allow one to make the right conclusions.

Resources for Execution

The third need of the CRM community is resources (e.g., funding, time) to allow researchers to conduct the necessary training and subsequent evaluations. First, resources are needed to systematically design and deliver the training. Second, resources are needed to standardize training evaluation methodologies and to conduct both immediate and longitudinal evaluations of training programs. Evaluating training programs at multiple levels is time consuming, labor intensive, and costly. Therefore, it is important that resources become available to make this a possibility. Finally, the communities involved need resources to properly train the trainers and the evaluators such that they can properly facilitate the training, provide the necessary feedback to trainees, and make improvements to future training programs.

We are aware of the difficulty in establishing a credible, direct cause-and-effect relationship between CRM training and safety. The low occurrence of mishaps prevents it. Maybe all one can do is rely on proxy measures of safety – the dependent variables – not safety rates. However, we submit that researchers must continue to seek better and robust evaluations of safety, clinical, and performance outcomes.

CONCLUSIONS

Although some have argued that there is no evidence that CRM is effective (Besco, 1995, 1997, 1998; Komich, 1997; Simmon, 1997), this review and others conclude that some evidence does exist. This is important. The picture that has emerged after reviewing the existing evidence within the current framework suggests that CRM training is effective at some levels (e.g., attitudes). As previously stated, however, 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 is a key factor in this unclear picture. Without evaluations at multiple levels, one cannot ascertain whether reactions lead to learning, learning leads to behavioral change, and behavioral change leads to organizational results.

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 the current evidence for the effectiveness of CRM training programs is impressive, albeit imperfect. Specifically, what can be said is that CRM (generally) produces positive reactions, enhanced learning, and desired behavioral change in a simulated or real environment. However, what cannot be answered with certainty is whether CRM training has an effect on the bottom line: safety. At this point, we believe the tools to determine this do exist; what is needed are a mandate, access to facilities and trainees (e.g., pilots, doctors), the resources to make it happen, and, of course, examination of additional measures of safety beyond accident rates. We believe that all of those interested should demand that these needs be met. Without these needs being met, CRM training programs will continue to be implemented and evaluated without ever reaping the true benefits.

ACKNOWLEDGMENTS

This research was partially funded by the U.S. Army Research Institute, Fort Rucker, Alabama, and partially supported by the U.S. Army Research Institute for the Behavioral and Social Sciences

through contract DASW01-03-C-0010. The research and development reported here was also partially supported by the FAA Office of the Chief Scientific and Technical Advisor for Human Factors, AAR-100. Dr. Eleana Edens was the contracting officer's technical representative at the FAA. All opinions stated in this paper are those of the authors and do not necessarily represent the opinion or position of the University of Central Florida, the U.S. Army Research Institute, or the FAA. We would like to thank Nancy Cooke (action editor for the paper) and three anonymous reviewers for their comments on earlier drafts of this paper.

REFERENCES

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., & Traver, H. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50,* 341–358.

Besco, R. O. (1995). The potential contributions and scientific responsibilities of aviation psychologists. In N. Johnston, R. Fuller, & N. McDonald (Eds.), *Aviation psychology: Training and selection: Proceedings of the 21st Conference of the European Association for Aviation Psychology* (Vol. 2, pp. 141–148). Aldershot, UK: Avebury Aviation.

Besco, R. O. (1997). The need for operational validation of human relations-centered CRM training assumptions. In R. S. Jensen & L. A. Rakovan (Eds.), *Proceedings of the 9th International Symposium on Aviation Psychology* (pp. 536–540). Columbus: Ohio State University.

Besco, R. O. (1998). *Crew resource management training: What to teach and how to teach it!* Retrieved September 21, 2000, from http://www.ppi.aero/commentary/crmcritique.htm

Bowers, C. A., & Jentsch, F. (2001). Use of commercial, off-the-shelf, simulations for team research. In E. Salas (Ed.), *Advances in human performance* (Vol. 1, pp. 293–317). Amsterdam: Elsevier Science.

Brun, W., Eid, J., Jihnsen, B. H., Ekornas, B., Laberg, J. C., & Kobbeltvedt, T. (2000). *Shared mental models and task performance: Studying the effects of a crew and bridge resource management training program* (Project Report: 1 2001). Bergen, Norway: Militaer Psykologi og Ledelse.

Davis, D. D., Bryant, J., Liu, Y., Tedrow, L., & Say, R. (2003). National culture, team behavior and error management in U.S. and Chinese simulated aircrews. In R. S. Jensen (Ed.), *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 273–277). Columbus: Ohio State University.

Edkins, G. D. (2002). A review of the benefits of aviation human factors training. *Human Factors and Aerospace Safety, 2,* 201–216.

Ellis, C. & Hughes, G. (1999). Use of human patient simulation to teach emergency medicine trainees advanced airway skills. *Journal of Accident Emergency Medicine, 16,* 395–399.

Fisher, J., Phillips, E., & Mather, J. (2000). Does crew resource management training work? *Air Medical Journal, 19,* 137–139.

Flin, R., Martin, L., Goeters, K., Hormann, H., Amalberti, R., Valot, C., et al. (2003). Development of the NOTECHS (non-technical skills) system for assessing pilot's CRM skills. *Human Factors and Aerospace Safety, 3,* 95–117.

Fonne, V. M., & Fredriksen, O. K. (1995). Resource management and crew training for HSV-navigators. In R. S. Jensen & L. A. Rakovan (Eds.), *Proceedings of the 8th International Symposium on Aviation Psychology* (pp. 585–590). Columbus: Ohio State University.

Gaba, D. M., Howard, S. K., Fish, K. J., Smith, B. E., & Sowb, Y. A. (2001). Simulation-based training in anesthesia crisis resource management (ACRM): A decade of experience. *Simulation and Gaming, 32,* 175–193.

Gaba, D. M., Howard, S. K., Flanagan, B., Smith, B. E., Fish, K. J., & Botney, R. (1998). Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology, 89,* 8–18.

Goeters, K. M. (2002). Evaluation of the effects of CRM training by the assessment of non-technical skills under LOFT. *Human Factors and Aerospace Safety, 2,* 71–86.

Grubb, G., Crossland, N., & Katz, L. (2002). Evaluating and delivering the U.S. Army aircrew coordination training enhancement (ACTE) program. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference* (pp. 1143–1149). Arlington, VA: National Training Systems Association.

Grubb, G., & Morey, J. C. (2003). Enhancement of the U.S. Army aircrew coordination training (ACT) program. In R. S. Jensen (Ed.), *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 446–452). Columbus: Ohio State University.

Halamek, L. P., Kaegi, D. M., Sowb, Y. A., Smith, B. C., Smith, B. E., & Howard, S. K. (2000). Time for a new paradigm in pediatric medical education: Teaching neonatal resuscitation in a simulated delivery room environment. *Pediatrics, 106,* Article e45. Retrieved November 19, 2003, from http://pediatrics.aappublications.org/cgi/reprint/106/4/e45

Helmreich, R. L. (1984). Cockpit management attitudes. *Human Factors, 26,* 583–589.

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.

Holzman, R. S., Cooper, J. B., Gaba, D. M., Philip, J. H., Small, S. D., & Feinstein, D. (1995). Anesthesia crisis resource management: Real-life simulation training in operating room crises. *Journal of Clinical Anesthesia, 7,* 675–687.

Howard, S. K., Gaba, D. M., Fish, K. J., Yang, G., & Sarnquist, F. H. (1992). Anesthesia crisis resource-management training: Teaching anesthesiologists to handle critical incidents. *Aviation, Space, and Environmental Medicine, 63,* 763–770.

i Gardi, T., Christensen, U. C., Jacobsen, J., Jensen, P. F., & Ording, H. (2001). How do anaesthesiologists treat malignant hyperthermia in a full-scale anaesthesia simulator? *Acta Anaesthesiologica Scandinavica, 45,* 1032–1035.

Jacobsen, J., Lindekaer, A. L., Ostergaard, H. T., Nielsen, K., Ostergaard, D., Laub, M., et al. (2001). Management of anaphylactic shock evaluated using a full-scale anaesthesia simulator. *Acta Anaesthesiologica Scandinavica, 45,* 315–319.

Karp, M. R., Condit, D., & Nullmeyer, R. T. (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.

Karp, M. R., Condit, D., & Nullmeyer, R. T. (2000). *Survey of cockpit/crew resource management for F-16 pilots* (Rep. No. AFRL-HE-AZ-TR-1999-0253). Mesa, AZ: Air Force Research Laboratory.

Katz, L. (2003). Army CRM training: Demonstration of a prototype computer-based program. In *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 648–650). Columbus: Ohio State University.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resources development* (pp. 18.1–18.27). New York: McGraw-Hill.

Komich, J. (1997). CRM training: Which crossroads to take now? In R. S. Jensen & L. A. Rakovan (Eds.), *Proceedings of the 9th International Symposium on Aviation Psychology* (pp. 541–546). Columbus: Ohio State University.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78,* 311–328.

Kurrek, M. M. & Fish, K. J. (1996). Anaesthesia crisis resource management training: An intimidating concept, a rewarding experience. *Canadian Journal of Anaesthesia, 43,* 430–434.

Lainos, J. S., & Nikolaidis, E. D. (2003). Presuppositions for the effective introduction of changes to the aviation safety culture, through CRM training program: The case study of the Olympic Airways. In R. S. Jensen (Ed.), *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 694–699). Columbus: Ohio State University.

Morey, J. C., Simon, R., Jay, G. D., Wears, R. L., Salisbury, M., Dukes, K. A., et al. (2002). Error reduction and performance improvement in the emergency department through formal teamwork training: Evaluation results of the MedTeams project. *Health Services Research, 37,* 1553–1581.

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]. Arlington, VA: National Training Systems Association.

O'Connor, P., & Flin, R. (2003). Crew resource management training for offshore oil production teams. *Safety Science, 41,* 591–609.

O'Connor, P., Flin, R., Fletcher, G., & Hemsley, P. (2002). Methods used to evaluate the effectiveness of flightcrew CRM training in the UK aviation industry. *Human Factors and Aerospace Safety, 2,* 235–255.

O'Connor, P., Flin, R., Fletcher, G., & Hemsley, P. (2003). *Methods used to evaluate the effectiveness of flightcrew CRM training in the UK aviation industry* (Civil Aviation Authority Paper 2002/05). Retrieved October 19, 2003, from http://www.caa.co.uk/docs/33/CAPAP2002_05.PDF

O'Donnell, J., Fletcher, J., Dixon, B., & Palmer, L. (1998). Planning and implementing an anesthesia crisis resource management course for student nurse anesthetists. *CRNA: The Clinical Forum for Nurse Anesthetists, 9,* 50–58.

Orasanu, J., Murray, L., Rodvold, M. A., & Tyzzer, L. K. (1999). Has CRM succeeded too well? Assertiveness on the flight deck. In R. S. Jensen (Ed.), *Proceedings of the 10th International Symposium on Aviation Psychology* (pp. 357–361). Columbus: Ohio State University.

Povenmire, H. K., Rockway, M. R., Bunecke, J. L., & Patton, M. W. (1989). Cockpit resource management skills enhance combat mission performance in a B-52 simulator. In R. S. Jensen (Ed.), *Proceedings of the 5th International Symposium on Aviation Psychology* (pp. 489–494). Columbus: Ohio State University.

Robertson, M. M., & Taylor, J. C. (1995). Team training in aviation maintenance settings: A systematic evaluation. In B. J. Hayward & A. R. Lowe (Eds.), *Applied aviation psychology: Achievement, change, and challenge. Proceedings of the Third Australian Aviation Psychology Symposium* (pp. 373–383). Aldershot, UK: Avebury Aviation.

Salas, E., Bowers, C. A., & Edens, E. (Eds.). (2001). *Improving teamwork in organizations: Applications of resource management training.* Hillsdale, NJ: Erlbaum.

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, 43,* 641–674.

Salas, E., & Cannon-Bowers, J. A. (2000). Designing training systems systematically. In E.A. Locke (Ed.), *The Blackwell handbook of principles of organizational behavior* (pp. 43–59). Malden, MA: Blackwell.

Salas, E., & Cannon-Bowers, C. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52,* 471–499.

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.

Salas, E., Prince, C., Bowers, C., Stout, R., Oser, R. L., & Cannon-Bowers, J. A. (1999). A methodology for enhancing crew resource management training. *Human Factors, 41,* 161–172.

Salas, E., Wilson, K. A., Burke, C. S., & Bowers, C. A. (2002). Myths about crew resource management training. *Ergonomics in Design, 10*(4), 20–24.

Simmon, D. A. (1997). How to fix CRM. In R. S. Jensen & L. A. Rakovan (Eds.), *Proceedings of the 9th International Symposium on Aviation Psychology* (pp. 550–553). Columbus: Ohio State University.

Small, S. D. (1998). What participants learn from anesthesia crisis resource management training [Extended Abstract]. *Anesthesiology, 89,* U153.

Small, S. D., Wuerz, R. C., Simon, R., Shapiro, N., Conn, A., & Setnik, G. (1999). Demonstration of high-fidelity simulation team training for emergency medicine. *Academic Emergency Medicine, 6,* 312–323.

Spiker, V. A., Tourville, S. J., Bragger, J., Dowdy, D., & Nullmeyer, R. T. (1999). Measuring C-5 crew coordination proficiency in an operational wing. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference* [CD-ROM]. Arlington, VA: National Training Systems Association.

Spiker, V. A., Wilson, D. D., & Deen, G. C. (2003). CRM and mission performance during C-130 mission-oriented simulator training. In R. S. Jensen (Ed.), *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 1108–1114). Columbus: Ohio State University.

Tannenbaum, S., & Yukl, G. (1992). Training and development in work organizations. In M. Rosenzweig & L. Porter (Eds.), Annual review of psychology (Vol. 43, pp. 399–441), Palo Alto, CA: Annual Reviews.

Taylor, J. C. (1998, August). *Evaluating the effectiveness of maintenance resource management (MRM).* Paper presented at the 12th International Symposium on Human Factors in Aviation Maintenance, Washington, DC.

Taylor, J. C. (2000, October). A new model for measuring return on investment (ROI) for safety programs in aviation: An example from airline maintenance resource management (MRM). Paper presented at the Advances in Aviation Safety Conference, Daytona Beach, FL.

Taylor, J. C., Robertson, M. M., Peck, R., & Stelly, J. W. (1993). Validating the impact of maintenance CRM training. In R. S. Jensen (Ed.), *Proceedings of the 7th International Symposium on Aviation Psychology* (pp. 538–542). Columbus: Ohio State University.

Taylor, J. C., & Thomas, R. L. (2003). Written communication practices as impacted by a maintenance resource management training intervention. *Journal of Air Transportation, 8,* 69–90.

Thompson, J. S., Tourville, S. J., Spiker, V. A., & Nullmeyer, R. T. (1999). Crew resource management and mission performance during MH-53J combat mission training. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference* [CD-ROM]. Arlington, VA: National Training Systems Association.

Wiegmann, D. A., & Shappell, S. A. (2000). Human error and crew resource management failures in naval aviation mishaps: A review of U.S. Naval Safety Center data, 1990–96. *Aviation, Space, and Environmental Medicine, 70,* 1147–1151.

Wiegmann, D. A., & Shappell, S. A. (2001). Human error analysis of commercial aviation accidents: Application of the human factors analysis and classification system (HFACS). *Aviation, Space, and Environmental Medicine, 72,* 1006–1016.

Eduardo Salas is program director at the Institute for Simulation and Training and trustee chair and professor in the Psychology Department at the University of Central Florida. He received his Ph.D. in industrial/organizational psychology in 1984 at Old Dominion University, Norfolk, Virginia.

Katherine A. Wilson is a research assistant at the Institute for Simulation and Training at the University of Central Florida, where she received her M.S. in modeling and simulation in 2002.

C. Shawn Burke is a research scientist at the Institute for Simulation and Training at the University of Central Florida. She received her Ph.D. in industrial/organizational psychology in 2000 at George Mason University, Fairfax, Virginia.

Dennis C. Wightman is chief of the Rotary Wing Aviation Research Unit at the Army Research Institute in Fort Rucker, Alabama. He received his Ph.D. in industrial/organizational psychology at the University of South Florida, Tampa, Florida.

Date received: March 11, 2004
Date accepted: January 25, 2005