Methods used to evaluate the effectiveness of CRM training: A literature review.

Cite as: O'Connor, P., Flin, R. & Fletcher, G. (2002). Methods used to evaluate the effectiveness of CRM training: A literature review. *Journal of Human Factors and Aerospace Safety*, 2(3), 217-234.

Techniques used to evaluate Crew Resource Management training: A literature review.

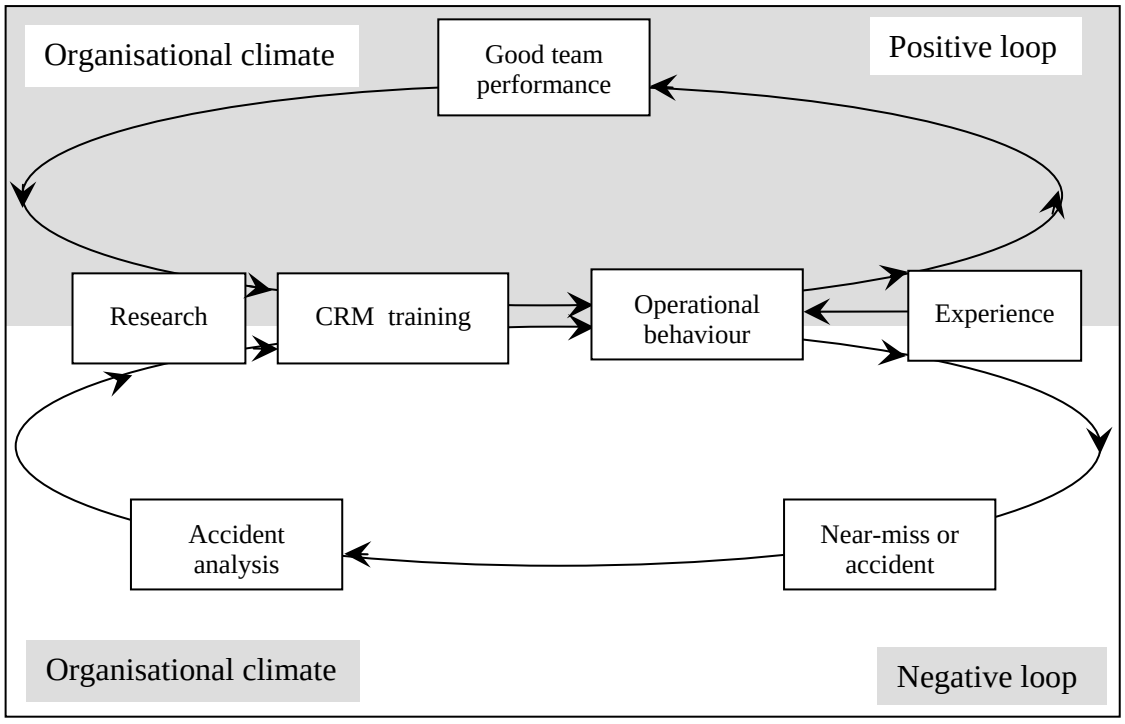
Abstract

This review paper examines the methods used to evaluate Crew Resource Management (CRM) training in 48 published studies from aviation (40) and other industries. The training evaluation techniques are categorised in terms of reactions, learning, attitudes, behaviour and organisational effects. It was found that in general CRM training was well received, resulted in a positive change in CRM attitudes, and had the desired effect on CRM behaviours. However, from the evidence available it is not possible to be as certain about the influence of the training on the organisation as a whole. This is because there are few studies that have made a rigorous assessment of the effects of CRM training on organisational metrics such as safety or productivity.

Introduction

The aviation industry has been instrumental in the development of human factors training programmes known as Crew Resource Management (CRM) designed to reduce error and increase the effectiveness of flight crews (Wiener, Kanki & Helmreich, 1993). CRM can be defined as "*using all the available resources- information, equipment, and people- to achieve safe and efficient flight operations*" (Lauber, 1984: 20). Figure 1 shows how CRM training needs are identified from both positive and negative input. In the positive loop, analysis of successful team performance reveals behaviours which are instrumental in success and need to be encouraged. The negative loop illustrates that in-depth analysis of accidents and confidential incident reports can pinpoint failures in CRM skills that can be addressed through training.

Figure 1. 'Figure eight' CRM training model.



CRM training is now used by virtually all the international airlines and is recommended by the major civil aviation regulators (e.g. FAA, 1998; JAA, 2001). A recent survey of International Air Transport Association affiliated airlines indicated that 96% of respondents were running CRM courses. Over 60% of these had been in existence for five years or more (O'Leary, 1999). In the UK, human factors training and examination are mandatory for a Flight Crew Licence, and the CAA requires that CRM training be carried out annually by commercial pilots (CAA, 1998a).

An introductory CRM course generally takes two or three days. Teaching methods include lectures, classroom training, practical exercises, case studies, and films. The topics covered, "*are designed to target knowledge, skills, and abilities as well as mental attitudes and motives related to cognitive processes and interpersonal relationships*" (Gregorich & Wilhelm, 1993: 173). A course typically covers six core topics: team work, leadership, situational awareness, decision making, communication, and personal limitations. Refresher training is also advised- normally a half or whole day course focusing on a specific CRM topic.

A framework for evaluating training effectiveness

The CAA has stated that '*the variability of CRM standards and the lack of common practical reference criteria have indicated the need for research into means of assessment*' (CAA, 1998b: 1). The Federal Aviation Administration (FAA) also recognises the crucial role of CRM evaluation. "*It is vital that each program be assessed to determine if it is achieving its goals*" (FAA, 1993: 9). There are a number of important reasons for evaluating the effects of CRM training programmes. Firstly are they meeting their stated goal of improving safety and efficiency? Secondly are they teaching the appropriate knowledge, attitudes and skills for an

ever-changing technology and a dynamic risk environment? Thirdly are companies receiving a return on investment for the development and delivery costs of these programmes?

The fundamental question of whether CRM training can fulfil its purpose of increasing safety and efficiency does not have a simple answer (Helmreich, Merritt & Wilhelm, 1999). Although research has been devoted to the ongoing development of CRM training courses (e.g. Helmreich & Foushee, 1993; Wiener, Kanki & Helmreich, 1993), only a small proportion of this has been devoted to evaluation (Edkins, under review; Holt, Boehm-Davis & Beaubien, 2001). Gregorich and Wilhelm (1993) argued for the need to evaluate the outcomes of CRM training using multiple assessment methods and longitudinal designs. Their approach categorised CRM training evaluation methods based on several levels of training effects (attitudes, behaviours, learning, motivation and organisational) in line with those advocated by training researchers (e.g. Hamblin, 1974; Kirkpatrick, 1976, 1998).

In this review a similar framework was adopted by using Kirkpatrick's (1976) hierarchy for training evaluation to examine the impact of CRM training interventions at four different levels: reactions, learning (attitudes and knowledge), behaviour, and organisational effects.

Level 1: Reactions. This is concerned with how the participants react to the training. It is important to indicate that a positive response does not ensure learning, although, a negative reaction almost certainly reduces the likelihood that this has taken place (Kirkpatrick, 1998).

Level 2: Learning. Learning refers to "*the principles, facts, and skills which were understood and absorbed by the participants*" (Kirkpatrick, 1976: 11). This level is concerned with

whether the participant has acquired knowledge or has modified their attitudes or beliefs as a result of attending the training course.

Level 3: Behaviour. The evaluation at the behaviour level is the assessment of whether knowledge learned in training has transferred to change actual behaviours on the job. (It is sometimes measured in a simulated work environment). This level is key, as the purpose of CRM training is to maintain safety-related behaviours at the workplace.

Level 4: Organisation. This is the highest level of evaluation. The ultimate aim of any training programme is to produce tangible evidence at an organisational level, such as an improvement in safety and productivity.

Method

In order to identify published CRM evaluation studies, information was drawn from a number of different sources:

- Online database searches using Web of Science and PsychLIT.
- The contents of conference proceedings in which this type of research is published (e.g. International Symposium of Aviation Psychology, European Association of Aviation Psychology) and journals (e.g. *International Journal of Aviation Psychology*).
- Bibliographies of research articles and chapters in relevant papers.

A total of 48 published studies were found in which CRM training was evaluated (see O'Connor, Flin & Fletcher, 2002 for details of the studies). The criterion for the inclusion of a study was that it must include an empirical evaluation of a CRM or equivalent human factors training course which was designed to improve performance. It was decided not to

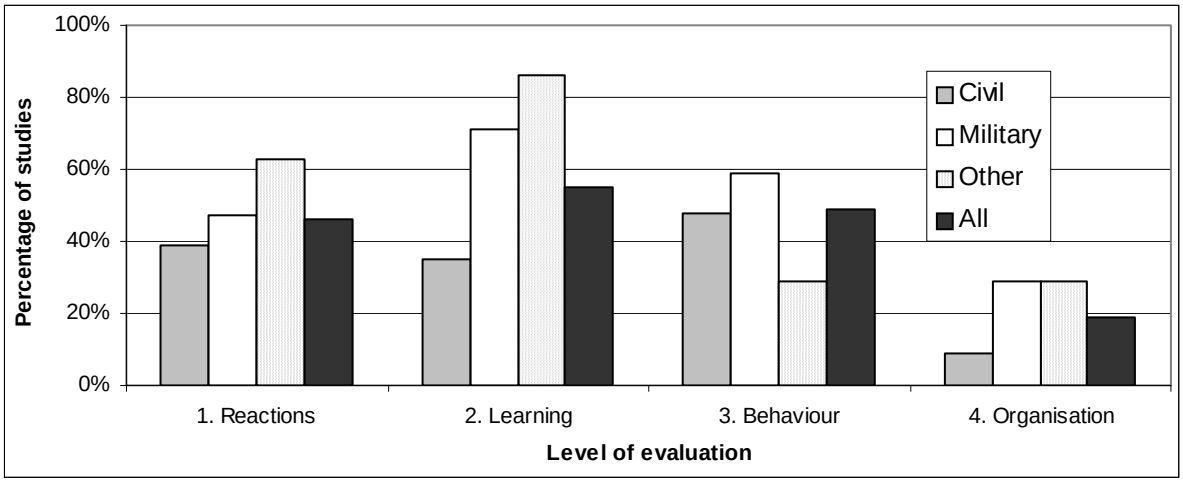
only concentrate on studies in civil aviation, but also to include military aviation, and other high reliability industries (aviation maintenance, maritime, offshore oil production, and medicine). The rationale for the inclusion of these other industries is that the domain in which the training is being applied is not relevant to this review. Rather, the purpose of the review is to concentrate on how the CRM training course is evaluated and what results were reported.

Of the 48 studies, 23 were from civil aviation (48%), 17 from military aviation (35%), and eight from other high reliability industries (17%; air traffic control, aviation maintenance, offshore oil and gas production, anaesthetics, nuclear power generation, and the maritime industry). The majority of the studies were carried out by US research teams (36 studies, 75%), with 10 being carried out by European researchers (21%), and two carried out by Australian and Japanese researchers (4%). The sample sizes ranged from 17 to 6,354 participants.

Evaluation Evidence

From Figure 2 it can be seen that the majority of studies examined the effectiveness of the CRM training at the learning level (i.e. attitudes and/or knowledge), with few researchers examining the effectiveness of the training at the organisational level. From an examination of those studies carried out in civil aviation, it appears that the most common level of evaluation was at the behaviour level, with only two of the studies making an assessment at the level of the organisation.

Figure 2. Percentage of studies carrying out CRM evaluation at each level



In the following sections, the studies describing CRM training evaluation will be examined in relation to each of the four method types and a summary given of the results of the analysis.

Reactions

A total of 22 (46%) of the studies carried out an evaluation of participants' reactions to CRM training: nine from civil aviation (39%), eight from military aviation (47%), and five in other high reliability industries (63%; see Figure 3). All of these used a paper-based questionnaire method.

Studies measuring reactions have generally reported an overall positive evaluation. To illustrate, Taggart and Butler (1989) assessed the reactions of over 2000 flight deck crew members to *Pan-Am's* Flight Operations Resource Management (FORM) training. It was found that 71% thought the seminar to be either very, or extremely useful, and all but 11% indicated that there would be some change in their behaviour on the flight deck. The only study to report some negative reactions was in Schiewe's (1995) assessment of 777 German *Lufthansa* cockpit crew. It was found that the modules that included training methods using

case studies or which allowed participants to act in a job related scenario were rated very positively. However, those modules based exclusively on lectures were not rated favourably.

Learning

Attitude assessment

An attitude assessment was carried out in 27 (56%) of the studies. The most commonly used experimental design was to compare attitude changes before and after training (83% of the attitude assessments). The remaining studies either carried out an evaluation after training (8%), or compared the post-training attitudes with a control group (8%).

The most frequently used tool for assessing pilots' attitudes to CRM is the Cockpit Management Attitudes Questionnaire (CMAQ; Helmreich, 1984). The CMAQ is a well established training, evaluation and research tool developed to assess the effects of CRM training for flight crew (Gregorich & Wilhelm, 1993). It comprises 25 items chosen to measure a set of attitudes that are either conceptually or empirically related to CRM. These cover 'Communication and co-ordination', 'Command responsibility', and 'Recognition of stressor effects'. This was used to assess attitude change in 75% of those studies carried out in civil aviation and 42% of the studies carried out in military aviation which assessed the effects of training at the learning level. The CMAQ has also been used as the basis of attitudes questionnaires designed to assess the attitudes of personnel to the concepts covered in CRM training. These questionnaires have been developed for use in military aviation (Leedom & Simon, 1995), the nuclear power industry (Harrington & Kello, 1991), aviation maintenance (Taylor, 1998), air traffic control (Woldring & Isaac, 1999) and the offshore oil and gas production industry (O'Connor & Flin, under review).

Using the CMAQ, researchers have consistently found there to be a positive shift in CRM attitudes as a result of CRM training (e.g. Helmreich, Wilhelm, Kello, Taggart & Butler, 1990; Helmreich & Wilhelm, 1991). Gregorich, Helmreich and Wilhelm (1990) found a positive change in responses towards CRM related attitudes in a survey of 4216 flight deck crew as a result of attending CRM training.

Irwin (1991) found a decay in positive attitudes to CRM over time. Attitudes were measured using the CMAQ at five different time intervals: a baseline (two years prior to initial CRM training), immediately prior to initial CRM training, immediately after initial CRM training, immediately prior to recurrent CRM training (one year later), and immediately after recurrent CRM training. It was found that there was an overall decline in attitudes on all three CMAQ sub-scales during the intervals between training interventions. The recurrent training resulted in another positive shift in attitudes back to the level found after initial training. Therefore, Irwin (1991) concluded that the reinforcement of CRM concepts through recurrent training is important for attitude maintenance and the stability of attitudes over time. This is also recognised by both the British and US regulators (CAA, 1995; FAA, 1998).

Knowledge assessment.

In CRM training learning can be assessed by testing students on their retention of the curriculum. Of the studies reviewed, only seven (15%) report any knowledge assessment. A comparison of knowledge from before and after training (Howard, Gaba, Fish, Yang & Sarnquist, 1992; O'Connor & Flin, under review) or with a control group who have not received the training (Salas, Fowlkes, Stout, Milanovich & Prince, 1999; Stout, Salas & Kraiger, 1996) provides an indication of what parts of the course have been retained by the participants. Of the papers examined, only Incalcaterra and Holt (1999) assessed any

knowledge change in civil aviation pilots. They examined the CRM knowledge gained by 166 Advanced Crew Resource Management (ACRM) trained pilots who were assessed on their knowledge to the new procedures and nomenclature in ACRM training. On the eight question multiple choice test, seven questions were answered above chance. This was despite the fact that the training had been delivered two years prior to the knowledge assessment. Thus, it was concluded that the ACRM training lead to an increase in CRM knowledge.

In military aviation, Salas et al (1999) found that although CRM training did not show an effect on the pilots' attitudes, it did appear to increase their knowledge of teamwork principles. Those who had participated in the CRM training scored significantly better than the baseline group that had not received any training (a mean of 12.6 out of 17, compared to 9.8 respectively). Stout et al (1996) assessed knowledge gain with military personnel but found no significant change on a multi-choice knowledge test between the trained and control groups. However, this could be attributed to the very small number of participants (12 trained and 10 controls). Another approach is to use case report analysis to ascertain whether the students have improved in their ability to identify human factors causes of incidents (O'Connor and Flin, under review).

Behaviour

A widely used technique for assessing CRM skills in flight crew is for training captains to use a behavioural rating system (see Klampfer et al, 2001). Of the studies evaluating behaviours, five carried out in civil aviation (45%), and six from military aviation (60%) employed a CRM skills rating scale with behavioural markers. A number of different behavioural marker systems were used, each of these is discussed below.

Line/LOS Checklist. The first set of behavioural markers was developed by Helmreich and colleagues in the 1980s. The markers were incorporated in a form for systematic observations known as the Line/LOS checklist (LLC). Clothier (1991) used an early version (2.0) of the LLC to assess the behaviours of crews on both the line and in LOFT for a US airline before and after CRM training. In this version, ratings are made on a five point scale of 14 different behaviours. On the line, a comparison between trained (1,000 crews) and untrained (2,000 crews) showed that there was a significant difference after training on 12 of the 14 categories. In LOFT, the 485 trained crews significantly outperformed the 1,625 untrained crews in all 14 categories of CRM behaviour. The LLC system has now been integrated into the Line Operations Safety Audit (LOSA; Helmreich, 2000) instrument.

Targeted Acceptable Responses to Generated Events or Tasks (TARGETs). The TARGETs system was developed by Fowlkes, Lane, Salas, Franz and Oser (1994) specifically for military crews. As with the LLC, this is a measure of crew performance rather than individual performance. Salas et al (1999) used the TARGETs approach to assess US Navy helicopter aircrew and pilots. It was found that the CRM trained crew performed 15% better than the untrained crew during the pre-flight brief and 9% better during high workload segments.

Line Operational Evaluations (LOE) worksheet. Holt and his colleagues at George Mason University used a behavioural marker system called a 'LOE worksheet' that was designed specifically to assess a particular situation or scenario. By observing 50 line flights in a US regional airline, it was found that those crews who had received Advanced Crew Resource Management (ACRM) behaviour training showed superior performance on 13 out of the 20 items evaluated (Ikomi, Boehm-Davis, Holt & Incalcaterra, 1999).

Aircrew Coordination Evaluation checklist. The Aircrew Coordination Evaluation checklist (ACE) was developed to assess military aviators (Leedom & Simon, 1995). Evaluation of team performance is organised around a set of 13 dimensions (e.g. establish and maintain flight team leadership and crew climate, permission planning and rehearsal accomplishment). Using the ACE checklist, Leedom and Simon (1995) found that after a week of CRM training, US military helicopter crews showed a significant improvement on 12 of the 13 team co-ordination dimensions. They displayed improved communication patterns within the cockpit, more efficient management of crew resources for critical flight tasks, and fewer team errors of the type previously implicated in aviation accidents.

Non-Technical Skills (NOTECHS). The NOTECHS system is a taxonomy of pilots' non-technical (CRM) skills developed and tested by a consortium of European research organisations and airlines (see Avermaete & Krujsen, 1998 and O'Connor et al, in press). The NOTECHS system is divided into four categories of behaviours, two of social skills (co-operation; leadership and managerial skills) and two of cognitive skills (situation awareness; decision making). Each category is then further subdivided into three or four elements (supporting others, anticipation) and for each of the 15 elements, positive and negative exemplar behaviours are provided.

This system has been recently developed and so far has been used once to assess the effectiveness of CRM training. As part of the PHARE Air Safety Improvement Project (Goeters, 2000), 17 aircrew from an Eastern European airline, prior to the introduction of their CRM training participated in the study. When a comparison was made between LOFT performance before and after the CRM training, it was found that there was an improvement

in the pilots' skills in all four NOTECHS categories. This difference was found to be statistically significant for both the situation awareness and decision making categories. The training was not a standard CRM course, but one specifically designed to address the areas of weakness identified in the first LOFT session. Thus, the CRM training was able to shape the professional behaviour in the manner intended (Goeters, 2000).

All of the studies reviewed reported an improvement in CRM behaviour as a result of participating in CRM training. However, for these systems to be used effectively for CRM evaluation purposes, it is necessary to ensure the system is valid, reliable, and that the evaluators have been trained to use the system to an acceptable level of reliability (see Klampfer et al, 2001; Baker, Mulqueen & Dismukes, 2001 for details on rater training).

Other behavioural evaluation methods

A number of other methods were used to assess the extent to which CRM behaviours had changed as a result of training. These methods generally consisted of measuring responses to questions about whether the course participants' would- or had- changed their behaviour as a result of the training. To illustrate, Naef (1995) reported that in a follow up survey six months after a nine day CRM training course in *Swissair*, 97% of the flight deck crew reported one or more positive behavioural changes. However, the problem with relying on subjective assessment is that the individual will be affected by whether they enjoyed the training, or by how well they felt they performed. Also, the participants will only describe what can be verbalised, which results in information that cannot be verbalised, or is difficult to verbalise, being ignored.

Organisational effects

Evidence of an impact at the organisational level is the most valuable evidence of the effect of CRM training. However, only eight of the studies (17%) carried out any evaluation at this level, with only two of these from civil aviation. The low number of studies carrying out an evaluation at this level reflects the difficulty in obtaining this type of information.

In military aviation, where there is a higher accident rate than civil aviation, five of the studies examined the effect of CRM training at the organisational level. Diehl (1991) found that CRM training reduced aircrew error accident rates by as much as 81 percent in the US military. Alkov (1989, 1991) found that CRM training reduced mishaps due to aircrew error for US naval aircrews. Over a 4 year period there was a reduction in error rates for helicopter pilots from 7.0 (per 100,000 flight hours) to 5.1; for bomber crews from 7.6 to 1.4; and multi-crew fighter aircraft from 13.9 to 6.3. Furthermore, the cost of the five year CRM programme was less than a million dollars. This represents a large financial saving when considering the cost of aircraft and human lives saved. Grubb, Morey and Simon (2001) attributed a reduction in accident rates from 1.8 (1993) to 0.8 (1996) per 100,000 flight hours in US Army aviation to Aircrew Co-ordination Training. Similarly, in the maritime industry, Byrdorf (1998) found that incidents and accidents in the *Maersk* shipping company decreased by a third from one major accident per 30 ship years in 1992 (before the introduction of CRM training) to one major accident per 90 ship years in 1996 (after the introduction of CRM training). Furthermore, at the beginning of 1998 insurance premiums were lowered by 15 percent. This reduction in accidents and incidents was attributed to the CRM and simulator training. In aviation maintenance, Taylor and Watson (2000) found that where safety trends were available, there was a reduction in aircraft ground damage and serious personal injuries that coincided with the Maintenance Resource Management (MRM) training.

In civil aviation, the accident rate is so low that it does not provide a robust test for the effectiveness of CRM training. Of the studies reviewed, only two examples of CRM evaluation at the organisational level were from civil aviation. Kayten (1993) cites several examples of NTSB reports in which good CRM practices were reported to limit the effects of either human or mechanical failures. Byrnes and Black (1993) found that the CRM programme at Delta Airlines resulted in a reduction in the quarterly air carrier discrepancy reports.

Multi-level analysis

The best approach used by CRM research teams in aviation is one that is multifaceted and considers several separate methods of assessment (Gregorich & Wilhelm, 1993; Kraiger, Ford & Salas 1993). However, from Figure 3 it can be seen that this was not generally found to be utilised in the studies examined. Particularly in the case of studies carried out in civil aviation, researchers only tended to undertake an evaluation at one of the levels. In civil aviation, a mean of 1.3 levels were evaluated per study, 2.1 in military aviation, and 2.1 in other high reliability industries.

Figure 2. Percentage of studies carrying out CRM evaluation at each level

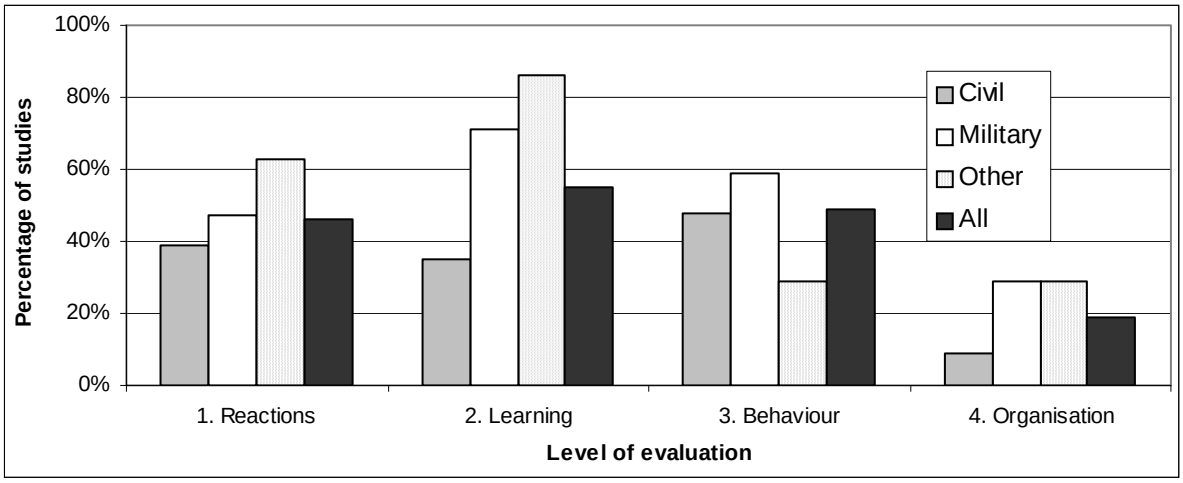
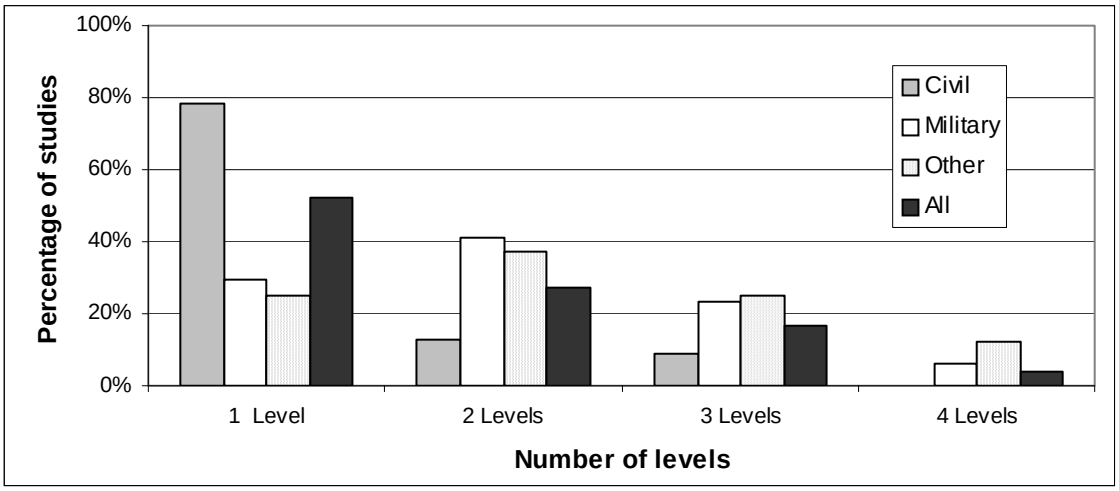


Figure 3. Number of levels of evaluation performed by the studies.



The only studies that looked at the effectiveness of the training at all four levels of the hierarchy were Grubb et al. (2001) in military aviation and Taylor (1998, 2000) with aviation maintenance personnel. Using a multi-level analysis approach allows the return on investment (ROI) of the training to be calculated. Taylor (2000) proposes the following equation for calculating the ROI of CRM training.

$$\text{ROI} = \left(\frac{[\text{Net CRM Benefits}] \times \text{Causal Operator}}{\text{CRM costs}} \right) \times 100$$

The *net CRM benefits* are the benefits of the training minus the cost of the development of the training (*CRM costs*). These are financially calculable savings such as reduction in accidents, reduction in damage to aircraft, increased productivity. The *causal operator* is a corrective calculation that is used to take into account for the many different factors which may act to improve the behaviours or outcomes. The value of the causal operator is the square of the correlation between the CRM training outcome (i.e. knowledge gained, improvement in attitudes or behaviours) and subsequent safety results. This provides a "*conservative, quantitative estimates of 'credit' to be allocated to organisational effectiveness (e.g. MRM) interventions*" (Taylor, 2000: 4).

Taylor (2000) gives the following example to illustrate how this equation is used. The development of a Maintenance Resource Management (MRM) course and delivery to 1,600 employees costs $251,660. A post-training survey showed a significant improvement in attitudes, and in the two years following the training there was a decrease in lost time injuries of 80%. A correlation of -.24 was found between lost time injuries and attitudes towards participative leadership and assertiveness in a post-training survey. Therefore, the coefficient of determination is $.24^2$ or 0.0576. Taylor (2000) estimated the cost of an injury as $13,465. Thus, a reduction of 80% on 91 incident per year is a saving of $1,314,150. Therefore, the ROI calculation is as follows:

$$\text{ROI} = \left(\frac{[\$1{,}314{,}150 - \$251{,}660] \times .0576}{\$\,251{,}660} \right) \times 100 = 24.3\%$$

It follows that even with the conservative estimate of 5.76% LTI benefit accounted for by the MRM training, the training paid for itself plus an additional 24% return in two years.

Edkins (under review) suggests that Taylor's (2000) ROI approach should be used in other areas of the aviation industry. It is crucial to justify capital expenditure on CRM training. As shown above, evaluations at a number of levels can be correlated to provide evidence of the effectiveness of the training.

Discussion

The 48 studies examined, showed that in general CRM training was well received by trainees, resulted in a positive change in their CRM attitudes, and had the desired effect on their CRM behaviours. However, there could be a reporting bias: journals or researchers may be less inclined to publish studies in which a positive training effect had not been found. It is difficult to draw firm conclusions on the effects of CRM training on knowledge and on the organisation. Only six of the studies reviewed assessed knowledge, and only one of these was carried out in civil aviation. Also, little evidence was found of evaluations carried out at the organisational level. Those studies that did examine effectiveness at this level, tended to have been carried out in the military, and were based on accident and incident frequencies. Thus there is a need to identify other metrics to allow any evaluation of the effects of CRM training at an organisational level. Similar conclusions have been made in a forthcoming paper by Salas, Burke, Bowers and Wilson (in press) following a review of studies of CRM training evaluation carried out in both civil and military aviation (mainly from the USA).

A recent survey of UK aviation companies (O'Connor, Flin, Fletcher & Hemsley, 2002) showed that the majority attempt to evaluate the impact of the company's CRM training, but

rarely at an organisational level. Many of the methods used would appear not to be based on formal evaluation techniques, and thus are unlikely to provide sufficient information to assess whether CRM training is actually transferring to the flight deck. When companies do not evaluate CRM training, the main reasons relate to limited resources and a lack of simple, concise material on how to carry out formal training evaluation.

CRM training is no longer confined to the training of pilots, but is beginning to be used to instruct a variety of personnel in a range of high reliability industries (Flin, O'Connor & Mearns, in press). In the eight studies included in this literature review it was found that the techniques used to evaluate CRM training in these industries have tended to be adaptations of the methods which have previously been used in aviation. The development of valid and reliable measures to evaluate CRM training is crucial to allow baseline measurements to be obtained. Prior to the extensive use of CRM training in the aviation industry, few baseline measures were obtained to allow an assessment to be made of the effects of the training (O'Connor et al, 2002).

It is important to track the effects of CRM training to allow for the identification of topics for recurrent programmes, and to ensure that it continues to improve performance despite changes in aircraft design, operational conditions, emerging risks and pilot demographics. As Gregorich and Wilhelm (1993) have argued, any evaluation should be carried out at multiple levels, with assessment techniques of proven validity and reliability. Proper evaluation data could be used for internal performance auditing, as well as for benchmarking across companies and industrial sectors to ensure an optimal return on CRM training investment.

References

Alkov, R. A. (1989). The US Naval aircrew coordination program. In R. Jensen (Ed.) *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 483-488). OH: Ohio State University.

Alkov, R. A. (1991). US Navy aircrew coordination training- a progress report. In R. Jensen (Ed.) *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 368-371). OH: Ohio State University.

Avermaete, van. J. & Kruijsen, E. (Eds.). (1998). *The evaluation of non-technical skills of multi-pilot aircrew in relation to the JAR-FCL requirements. (NOTECHS). Project report: CR-98443*. Amsterdam, The Netherlands: NLR.

Baker, D., Mulqueen, C.& Dismukes, R.K. (2001). Training rater to assess resource management skills. In E. Salas, C. A. Bowers & E. Edens (Eds.), *Improving teamwork in organizations: Applications of resource management training* (pp. 131-146). Mahwah, NJ: Lawrence Erlbaum Associates.

Byrdorf, P. (1998 September). *Human Factors and Crew Resource Management: An example of successfully applying the experience from CRM programmes in the Aviation World to the Maritime World.* Paper presented at the 23rd Conference of the European Association for Aviation Psychology, Vienna.

Byrnes, R. E. & Black, R. (1993). Developing and implementing CRM programmes: The Delta experience. In E. Wiener & B. Kanki & R. Helmreich (Eds.), *Cockpit resource management* (pp. 421-443). San Diego: Academic Press.

CAA. (1995). *United Kingdom aeronautical information circular: Crew Resource Management Training* (AIC 37/1995). London: Civil Aviation Authority.

CAA. (1998a). *United Kingdom aeronautical information circular: crew resource management*, AIC 117/1998. London: Civil Aviation Authority.

CAA. (1998b). *United Kingdom aeronautical information circular: flight crew CRM training standards*, AIC 114/1998. London: Civil Aviation Authority.

Clothier, C. (1991). Behavioural interactions across various aircraft types: Results of systematic observations of line operations and simulations. In R. Jensen (Ed.) *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 332-337). OH: Ohio State University.

Diehl, A. (1991). The effectiveness of training programs for preventing aircrew "error". In R. Jensen (Ed.) *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 640-655). OH: Ohio State University.

Edkins, G.D. (under review). A review of the benefits of aviation human factors training.

FAA (1993). *Advisory circular 120-51A crew resource management training* Washington: US Department of Transport.

FAA (1998). *Advisory circular 120-51C: crew resource management training* Washington: US Department of Transport.

Flin, R., O'Connor, P. & Mearns, K. (in press). Crew resource management: Improving teamwork in high-reliability industries. *Team Performance Management*

Fowlkes, J., Lane, N., Salas, E., Franz, T. & Oser, R. (1994). Improving the measurement of team performance: The TARGETs methodology. *Military Psychology, 6(1),* 47-61.

Goeters, K. M. (2000 September). *Validation of CRM training by NOTECHS: Results from the PHARE ASI project.* Paper presented at the 24th Conference of the European Association for Aviation Psychology, Crieff, Scotland.

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

Gregorich, S. E. & Wilhelm, J. A. (1993). Crew resource management training assessment. In E. L. Wiener & B. G. Kanki & R. L. Helmreich (Eds.), *Cockpit Resource Management* (pp. 173-196): San Diego: Academic Press.

Grubb, G., Morey, J. C. & Simon, R. (2001 February). *Sustaining and advancing performance improvements achieved by crew resource management training.* Paper presented at the 11th International Symposium of Aviation Psychology, Columbus, Ohio.

Hamblin, A. C. (1974). *Evaluation and control of training.* London: McGraw Hill.

Harrington, D. & Kello, J. (1991) Systematic evaluation of nuclear operator team skills training. *Transactions of the American Nuclear Society, 64,* 182-183.

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

Helmreich, R. (2000). *The Line Operations Safety Audit (LOSA). (Version 9).* Austin: NASA/University of Texas/Federal Aviation Administration Aerospace Group.

Helmreich, R. L. & Foushee, H. C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E. L. Wiener & B. G. Kanki & R. L. Helmreich (Eds.), *Cockpit Resource Management* (pp. 3-41). New York: Academic Press.

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

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

Helmreich, R., Wilhelm, J., Kello, J., Taggart, E. & Butler, R. (1990). *Reinforcing and evaluating crew resource management: evaluator/LOS instructor manual*. Austin: NASA/UT/FAA Aerospace Group.

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

Howard, S. K., Gaba, D. M., Fish, K. J., Yang, G. S. & Sarnquist, F. H. (1992). Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents. *Aviation, Space, and Environmental Medicine, 63*(9), 763-770.

Incalcaterra, K. A. & Holt, R. W. (1999). Pilot evaluation of ACRM programs. In R. Jensen (Ed.) *Proceedings of the Tenth International Symposium on Aviation Psychology* (pp.285-291). OH: Ohio State University

Ikomi, P. A., Boehm-Davis, D., Holt, R. W. & Incalcaterra, K. A. (1999). Jump seat observations of Advanced Crew Resource Management (ACRM) effectiveness. In R. Jensen (Ed.) *Proceedings of the Tenth International Symposium on Aviation Psychology* (pp. 292-297). OH: Ohio State University

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

JAA. (2001). *JAR-OPS ,1 amendment 3 (November 2001), Crew resource management-flight crew*. Hoofddorp, Netherlands: EC Joint Aviation Authorities.

Kayten, P. J. (1993). The accident investigator's perspective. In E. Wiener & B. Kanki & R. Helmreich (Eds.), *Cockpit resource management* (pp. 199-230). San Diego: Academic Press.

Kirkpatrick, D. L. (1998). *Evaluation training programs*. San Fransisco: Berrett-Koehler.

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

Klampfer, B., Flin, R. Helmreich, R, Häusler, R., Sexton, B., Fletcher, G., Field, P., Staender, S., Lauche, K., Dieckmann, P. & Amacher, A. (2001) *Enhancing performance in high risk environments: Recommendations for the use of behavioural markers. Report from the behavioural markers workshop, Zürich, June*. Berlin: Damler Benz Foundation (available on the internet from www.psyc.abdn.ac.uk/homedir/poconnor/GIHRE21.pdf).

Kraiger, K., Ford, J. K. & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcome to new method of training evaluation. *Journal of Applied Psychology, 78(2),* 311-328.

Lauber, J. K. (1984). Resource Management in the cockpit. *Air Line Pilot, 53,* 20-23.

Leedom, D. K. & Simon, R. (1995). Improving team coordination: a case for behavior -based training. *Military Psychology*, *7*(2), 109-122.

Naef, W. (1995). Practical applications of CRM concepts: Swissair's human aspects development program (HAD). In R. Jensen (Ed.) *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 597-602). OH: Ohio State University.

O'Connor, P. & Flin, R. (under review). Crew resource management training for offshore production teams.

O'Connor, P., Flin, R. & Fletcher, G. (2002). Literature review: Methods used to evaluate the effectiveness of CRM training in the aviation industry. CAA research project (121/SRG/R&AD/1) report . London: Civil Aviation Authority.

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. CAA research project (121/SRG/R&AD/1) report . London: Civil Aviation Authority.

O'Connor, P., Hoermann, H. J., Flin, R., Goeters, K. M., Lodge, M. & the JARTEL group. (in press). Developing a method for evaluating CRM skills: A European perspective. *International Journal of Aviation Psychology*.

O'Leary, M. (1999 June). *The British Airways human factors reporting programme.* Paper presented at the Human error, safety, and system development conference, Liege.

Salas, E., Burke, C. S., Bowers, C. A. & Wilson, K. A. (in press). Team training in the skies: Does Crew Resource Management (CRM) training work? *Human Factors*.

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(2),* 326-343.

Schiewe, A. (1995). *On the acceptance of CRM-methods by pilots: Results of a cluster analysis.* In R. Jensen (Ed.) *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 540-545). OH: Ohio State University

Stout, R. J., Salas, E. & Kraiger, K. (1996). The role of trainee knowledge structures in aviation psychology. *International Journal of Aviation Psychology, 7*(3), 235-250.

Taggart, W. R. & Butler, R. E. (1989). CRM validation program. In R. Jensen (Ed.) *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 468-475). OH: Ohio State University

Taylor, J. C. (2000 April). *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, Paper number 2000-01-2090, Daytona Beach, Florida.

Taylor, J. C. (1998 May). Evaluating the effectiveness of maintenance resource management (MRM). Paper presented at the 12th International Symposium on Human Factors in Aviation Maintenance, Canada: Vancouver.

Taylor, J. C. & Watson, J. (2000). Evaluating the effectiveness of maintenance resource management (MRM) in aviation safety. Paper presented at the 14th International Symposium on Human Factors in Aviation Maintenance, Canada: Vancouver.

Wiener, E., Kanki, B. & Helmreich, R. (1993). *Cockpit resource management*. San Diego: Academic Press.

Woldring, M. & Isaac, A. (1999). *Team Resource Management test and evaluation* (HUM.ET1.ST10.2000 -REP-01): EUROCONT. Brussels: European Air Traffic Management Programme.