

Cite as: O'Connor, P., Flin, R., Fletcher, G., & Hemsley, P. (2002) Methods used to evaluate the effectiveness of CRM training: A survey of UK aviation operators. *Journal of Human Factors and Aerospace Safety*, 2(3), 235-256.

Methods used to Evaluate the Effectiveness of Flightcrew CRM Training in the UK Aviation Industry.

Abstract

This study examines the techniques which UK aviation operators are employing to evaluate crew resource management (CRM) training. A questionnaire designed to elicit information on the methods used in industry to evaluate CRM training was sent to all aviation operators in the UK. A total of 113 questionnaires were returned (65% response rate). A representative sample of 20 interviews was also carried out with companies who had completed the questionnaire. The survey showed that most UK operators attempt to evaluate the impact of their CRM training. However, many of the methods used would appear not to be based on formal evaluation techniques and do not provide sufficient information to assess whether CRM training is actually transferring to the flight deck. The main reasons why companies are not evaluating CRM training are a lack of availability of resources and of simple, concise material on how to carry out training evaluation.

Introduction

In recent years the regulatory bodies of the aviation industry have become increasingly concerned with the evaluation of the effectiveness of Crew Resource Management (CRM) training (CAA, 1998; FAA, 1998). A review of 48 CRM training studies showed that CRM training generally results in positive reactions from participants to the training, as well as desired changes in attitude, knowledge and behaviour reported (O'Connor, Flin & Fletcher, under review). However, it is not possible to be as certain about the influence of the training on the organisation as a whole and whether it has had the ultimate effect of increasing safety. Similar conclusions have been drawn 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 USA).

The purpose of this study is to establish the methods that are currently being used in the UK aviation industry to evaluate flight crew CRM training and to ascertain whether guidance on evaluation techniques is required.

Method

In order to collect information from all the aviation operators in the UK a mailed questionnaire survey was used, followed by a series of interviews with a sub-sample of respondents.

Questionnaire design

The questionnaire was developed on the basis of a literature review (see O'Connor et al, under review), and a previous survey conducted at a workshop on human factors

training evaluation at the Australian Aviation Psychology Conference in Manly, Australia in November 2000. The questionnaire consisted of nine sections. The first section was concerned with whether pilots received initial and recurrent CRM training, and who provided it. The next five sections were concerned with whether an evaluation was carried out with regard to Kirkpatrick's (1976) hierarchy of four levels of evaluation, reactions, learning (attitudes and knowledge), behaviour, and organisation. Section 7 asked whether any other evaluation techniques were employed. The purpose of section 8 was to gather information about the reasons why companies are not evaluating CRM training. Respondents were asked to rank reasons for not evaluating CRM training. Section 9 asked respondents to rank methods of evaluating CRM training. They were also given the opportunity to describe in their own words what could be done to help operators to evaluate the effectiveness of their CRM training.

Questionnaire sample

The questionnaires were sent out by post to a named individual responsible for CRM training within the particular organisation. The target organisations were UK Air Operator Certificate (AOC) and Police Air Operator Certificate (PAOC) holders. An AOC is issued by the CAA and is required by aircraft operators flying scheduled or charter public transport flights, including cargo, air-taxi and pleasure-flying work. Police forces are issued with a PAOC licensing them to perform their specialist air operations.

A total of 113 questionnaires were returned (65% response rate). To make comparisons between the companies, it was decided to differentiate between large, medium, and

small companies (see Table 1). Following discussions with the CAA, it was decided to use the type of licence held and the capacity in ‘Percentage of all available Seat Kilometres’ between April 2000 to March 2001 (CAA, 2001).

Table 1 Sample demographics (number of respondents in brackets).

Type of certificate		
AOC holders.... 57% (90 responses) PAOC holders...73% (22 responses)		
No details 1% (1 response)		
Type of licence (% of total sample)		
Type A.... 30% (34)	Type B.... 50% (56)	Police 19% (22)
No details 1% (1)		
Aircraft Type		
Fixed-wing.... 59% (67)	Rotary.... 40% (45)	No details 1% (1)
Size of Company		
Large.... 10% (11)	Medium.... 20% (23)	Small 69% (78)
No details 1% (1)		

Type A AOC holders are permitted to carry passengers, cargo and mail on aircraft with 20 or more seats, and Type B AOC holders are permitted to carry passengers, cargo and mail on aircraft with fewer than 20 seats and/or weighing less than 10 tonnes (CAA, 2001). Available Seat Kilometres represents the number of seats available for passengers multiplied by the number of kilometres those seats are flown. The method of distinguishing between large, medium and small was as follows:

- Large: Type A licence operators with 1% (or greater) of the percentage of all available seat-kilometres.

- Medium: Type A licence operators with less than 1% of the percentage of all available seat-kilometres.
- Small: Type B licence operators and PAOC holders.

Interview survey.

Information obtained from the questionnaire was augmented with interview data from a sample of 20 respondents who had completed the questionnaire survey. The interviews were carried out by four researchers between July and September 2001 (10 face-to-face interviews, and 10 telephone interviews). Nine interviews were carried out with large companies, seven with medium sized companies, and four with small companies.

Results

The responses to each of the nine sections of the questionnaires and interview will be examined in turn with respect to the size of the company.

Flight crew CRM training

For the large operators, basic/foundation and recurrent training is almost exclusively provided by an in-house training department. This figure falls to 74% (17 respondents) for the medium sized companies, with the remainder employing specialist consultants to carry out the training. For the small companies, half the basic/foundation training is carried out by specialist consultants, with the remainder split almost equally between in-house training departments and other airlines employed to carry out the training. However, for the recurrent training, 56% (44) of the small companies carried out this training in-house, with only 9% (7) of them using another airline and 15% (12)

employing specialist consultants. For the large and medium companies 73% (8) and 57% (13) of the CRM training respectively was designed for the specific operations of the flight crew. For the small companies almost 50% (38) of the training was generic and not designed specifically for the operations carried out by the crews. The findings from the interviews largely endorsed the information obtained from the questionnaire responses. The reasons given in the interviews for the benefits of carrying out the training in-house were that it allows courses to be designed for the specific operations and cultures of the airline, and is cost effective.

Assessment of reactions to CRM training

A total of 61% (69) of respondents to the questionnaire reported carrying out an assessment of the reactions of participants to the training. All 11 of the large operators reported that they carried out an assessment of the reactions. For the medium and small operators, this figure reduced to just over 50% (13 and 44 respondents respectively). There were also differences in the methods used by the three sizes of operators to assess the reactions of the participants. Almost 75% (8) of the large operators reported that they used a reaction sheet to obtain feedback. However, only 31% (4) of medium sized operators, and 14% (6) of small operators who were carrying out any evaluation at this level used reaction sheets. The companies who were not using a reaction sheet relied on oral feedback from the participants. It was commented that this information was not only obtained after training but whenever the instructors came into contact with the crews.

The follow-up interviews indicated that those companies that used reaction sheets had generally developed the questionnaire in-house and the course participants filled them out anonymously. The reaction sheets were normally fairly short and consisted of closed questions on each of the topics covered in the course. Although the data tended not to be formally analysed to identify trends, the information was examined and used to make changes to the training course if required. In some companies, however, it transpired from interviews that the 'reaction sheet' was a blank piece of paper where the participants were asked to record any comments. Companies using oral feedback from participants considered this to be sufficient for course instructors to assess how the training was being received and to establish what changes needed to be made to the course. All of the companies interviewed stated that, in general, CRM training was received positively by most participants.

Assessment of attitudes to CRM training

Of all the respondents to the questionnaire, 21% (24 respondents) reported carrying out an evaluation at the attitude level. Looking at the three different sizes of company separately, 18% (2) of large companies, 43% (10) of medium companies, and 15% (12) of small companies assessed flight crew attitudes to CRM. One of the large companies used a company specific attitude questionnaire, the other used the Cockpit Management Attitude Questionnaire (CMAQ) designed by Gregorich, Helmreich and Wilhelm (1990; see discussion). An attitude questionnaire was used by 20% (2) of the medium sized operators, and only 8% (1) of the small operators that carried out an assessment of attitudes. Thus, as with the reactions to the training course, the majority of the companies rely on informal oral feedback. In addition, of those companies that carry out

an evaluation of attitudes to CRM concepts, only 75% (18) have used this information to evaluate the company's CRM training.

Only 10% (2) of the companies interviewed carry out a formal assessment of the attitudes of the aircrew. One of the companies used an adaptation of the CMAQ. The course participants completed the survey anonymously. The other company used an attitude questionnaire designed specifically for use by the company and which the course participants completed anonymously. The response from the companies who did not carry out a formal assessment of attitudes was to question the relevance of this type of analysis and the lack of suitable measurement systems.

Assessment of knowledge of CRM concepts

A total of 36% (41 respondents) reported carrying out an assessment of knowledge. The largest proportion of companies who reported carrying out a knowledge assessment were small (42%, 33) in comparison with 26% (6) of medium companies and 9% (1) large company (the responses of the one anonymous respondent are also included). The most common technique used by those companies carrying out a knowledge test was oral feedback from flight crew (88%; 36), with multiple choice tests and written exams accounting for 12% (5) in each case. A total of 61% (25) of the companies who carried out a knowledge assessment reported using the information obtained from the feedback to evaluate their CRM training.

Three of the companies interviewed carried out a formal assessment of flight crews' knowledge of the concepts covered in CRM training. One of the medium sized

companies required flight crew to complete a written exam two weeks prior to each annual CRM training period. This test was developed in-house and is open book and non-jeopardy. Two small companies also reported that they used a formal assessment of knowledge. The majority of companies interviewed either did not carry out any assessment of knowledge or carried out an informal assessment only.

Assessment of CRM behaviour

A total of 53% (60 respondents including the one anonymous response) of respondents reported carrying out an evaluation of flight crew behaviours. For large operators 82% (9) report that they had carried out an assessment of flight crews' CRM skills in the past two years. For the medium sized operators, this was the case for 70% (16) of respondents, and 44% (34) for smaller operators.

Behavioural marker systems were used by 67% (6 respondents) of large companies, with 33% (3) relying on informal feedback. For medium sized and small operators who conducted a behavioural assessment, 19% (3) and 12% (4) respectively used behavioural markers, 13% (2) and 24% (8) reported using technical checklists, and 69% (11) and 65% (22) used informal feedback for both sizes of operators. Examining the 14 operators that used behavioural markers, 64% (9) used company specific behavioural marker systems, with 22% (3) using NOTECHS (a European behavioural markers system; see Avermaete & Kruijsen, 1998), and 14% (2) using the Line/LOS checklist (a behavioural marker system developed at the University of Texas, see Helmreich, 2000; Klampfer et al, 2001).

Companies carrying out an evaluation of flight crew behaviours were asked when the assessment took place. The responses for the large and medium companies were similar with a fairly equal split between base/proficiency checks, line/route checks, and simulator/Line Oriented Flight Training (LOFT). The majority of large carriers carried out an evaluation at all three occasions (78%; 7 respondents), whereas only 50% (8) of the medium carriers did so. For the 34 small companies who carried out CRM behaviour assessment, there was an almost even split between base/proficiency checks and line/route checks. The majority (50%; 17) of small companies carried out the evaluation on only one of the occasions listed. The one medium and three small companies that selected the 'other' response category provided in the questionnaire reported that the CRM skills of the flight crew were discussed after every flight. It was also reported that 49% (29) of all companies assessing behaviour use the information about the CRM skills of flight crew to evaluate the training.

One operator also identified a behavioural level evaluation technique called a 360° appraisal. An outside expert developed a questionnaire in which the managers, pilots and crew were given the opportunity to provide ratings of their own performance, as well as those of the other members of the company. This process provided individuals with the opportunity to receive feedback on their skills, and how their perception of them differed from that of their work colleagues.

A total of 35% (7) of the companies interviewed reported using a behavioural marker system. Two reported carrying out a formal assessment using systems based on the NOTECHS framework. The training captains undertaking the assessments had received

no formal tuition in using the system but they felt able to carry out the ratings accurately. Three companies used systems based on the Line/LOS Checklist (LLC). Companies using it appear not to formally train raters, and the information is retained as part of a crew members' training record. This evaluation occurred at base, on the line, and/or during simulator training. The other two companies were using behavioural marker systems developed in-house (by training captains) at line and simulator checks. Their assessments are recorded in the flight crews' training records. 15% (3) of the other companies used technical checklists that included one or two items relating to CRM skills. The remaining companies interviewed (50%;10) carried out an informal assessment of behaviours using oral feedback from the trainers after simulator/LOFT training. The reason given for not using behavioural markers was that the trainers felt this type of system to be complicated and would be unable to use it accurately.

Assessment of organisational effects of CRM training

Of all of the respondents to the questionnaire, 33% (37 respondents) reported carrying out an organisational performance evaluation. In 36% (4) of the large companies surveyed, 52% (12) of the medium companies and 27% (21) of the small companies an evaluation was carried out at the organisational level. The respondents were given a list of six evaluation methods and asked to indicate if they were using any of the evaluation techniques. However, a number of the methods, such as company climate surveys and business performance measures, were not specific to the flight crew alone and may have limited utility for the evaluation of CRM training. Other measures being used, such as safety performance, incident reporting, confidential reporting, CRM training audit and

technical performance, have more direct relevance to the flight crews and CRM training (see Table 2).

Table 2. Percentage of organisational performance evaluation techniques reported to have been used.

	Size of company		
	Large	Medium	Small
Safety performance	30	23	18
Incident reporting	10	23	25
Confidential reporting	20	13	18
Technical performance	30	7	5
CRM training audit	0	7	9
<i>Non-flight crew specific evaluations</i>	<i>10</i>	<i>27</i>	<i>25</i>

The largest proportion of the 37 companies who collected information at an organisational level reported using one technique only (41%; 15 respondents). The most common of the techniques were safety performance data, incident and confidential reporting. The information has been used to evaluate CRM training by 41% (15) of the companies who obtained organisational data.

Although none of the small companies interviewed reported carrying out an evaluation of CRM training at an organisational level, four of the large companies and four of the medium sized companies reported carrying out such evaluations. However, in the interviews, it was found that three of these 8 companies were referring mainly to business and internal audits with limited implications for CRM training.

The five companies that reported carrying out an evaluation at this level which was relevant to flight crew collected information such as confidential reporting, incident reporting and safety performance. Two companies reported using an aviation safety management tool called the British Airways Safety Information System (BASIS).

Use of other evaluation techniques

None of the companies surveyed or interviewed reported using any other novel method of evaluation not described above.

Reasons why CRM training effectiveness is not evaluated

The responses given to the questions about why CRM training is not evaluated were common across all company sizes. When asked to choose the factors that prevent companies from evaluating CRM training, a mean of 2.1 options were chosen by the respondent. The most common choices were time (26%; 60 respondents) and resources (25%; 58), followed by availability of measurement systems (19%; 44), expertise (18%; 42), and management support (10%; 24).

Respondents were asked to rank a number of possible problems relating to the evaluation of CRM training. The overall scores for each of the methods were calculated as follows: the number of respondents that ranked a method in position 1 was multiplied by 3, the number who ranked it in position 2 was multiplied by 2, and these numbers were added to the number of respondents who ranked it in position 3. This gave an overall score for each method. To illustrate, 42 respondents ranked *availability of measurement systems* as the best method (position 1), 22 ranked it as second, and 12

ranked it as third. Therefore, the overall score is $(42 \times 3) + (22 \times 2) + 12 = 182$. The overall scores were then ranked in order from 1 to 7 (see Table 3). It should be noted that a limitation of this question was that time was omitted as a possible problem. Nevertheless, the question provided useful insight into some of the difficulties identified by the respondents.

Table 3. Ranking of problems relating to the evaluation of CRM training

(1= greatest problem).

	Large	Medium	Small	All
Availability of measurement systems	1	1	1	1
Competence/expertise of evaluators	2[#]	2	3[#]	2
Quality of measurement systems	=3	3[#]	2	3[#]
Financial	6	5	4	4
Flight crew's attitude/acceptance	=3	4	5	5
Management	5	6	6	6
Other	7	7	7	7

[#]Indicates a large drop in the overall scores after this point.

Table 3 shows clearly that the main problems with regard to CRM evaluation for all operators relate to the availability of measurement systems, competence of evaluators, and the quality of measurement systems. There were differences between company size as to whether evaluating CRM training is judged to be useful. For large companies, 82% (9) of respondents considered evaluation to be beneficial, for medium sized companies 70% (16), but only 46% (36) for small companies. The majority of comments favoured CRM evaluation in that it was perceived to be useful in providing feedback to trainers and in facilitating changes to training. However, some respondents thought that it was

of limited use to their organisation as, for example, their company was a single pilot operation.

The interview responses supported the evidence obtained from the questionnaire study. The greatest perceived problem reported by the companies relating to CRM evaluation was the lack of availability of measurement systems. There was less support for the benefit of CRM training evaluation to flight crew in the small companies. There was also scepticism in these companies that it is possible to evaluate CRM training at all. There appears to be a consensus that CRM training is generally a multi-crew training mechanism and of limited relevance to single crew aircraft. Nevertheless, in the large and medium companies the respondents were more positive about the benefits of CRM training.

Methods to support CRM training evaluation

To gain an understanding of the most preferred method for evaluating CRM training, respondents were given the opportunity to rank three methods of evaluation from a list of 13. The same technique for calculating the overall score was used as before. A summary of the overall scores ranked from 1 to 13 is shown in Table 4.

Table 4. Ranking of methods of evaluating CRM training (1= best method).

	Large	Medium	Small	All
Line/route checks	2	1	1	1
Base/proficiency checks	3[#]	=4	2[#]	2
Simulator/LOFT checks	1	2[#]	4	3[#]
Self/peer/360° appraisal	=10	=6	3	4
Confidential reports	=7	=4	5	5
Feedback questionnaire	9	3	7	=6
Interview sessions	=10	9	6	=6
Attitude surveys	4	=6	10	8
Knowledge assessment	5	10	=8	9
Incident reports	=7	8	=8	10
Accident data	=10	11	11	11
Technical performance	6	13	13	12
Other	13	12	12	13

[#] Indicates a large drop in the overall scores after this point.

From Table 4 it can be seen that respondents across all companies believe that an evaluation of CRM training should be carried out of the behaviour of the flight crew during either real or simulated flight. Other types of assessments addressing reactions (feedback questionnaires), learning (attitude surveys, interview sessions, knowledge assessment) or organisational levels (confidential reports, incident reports, accident data, technical performance) were chosen with a much lower frequency.

The preferred methods, identified in the interviews by the large and medium companies, for evaluating CRM training were simulator/LOFT checks, line/route checks and base/proficiency checks. These were chosen because the assessment is carried out in a

realistic environment in which the trainers can actually see how the flight crew are behaving.

Summary of results

The overall conclusion that can be drawn from the questionnaire survey and interviews is that the majority of companies are making attempts to evaluate CRM training. Only 14% (16) of companies reported no evaluation of CRM training. All 11 of the large companies, 83% (19) of medium companies and 83% (65) of small companies carried out some type of CRM evaluation. However, the majority of companies did not undertake formal evaluations of CRM training at more than one level of analysis (see Table 5). Further, the proportion of companies carrying out formal evaluations of CRM training is lower for the smaller companies.

Table 5. Number of levels of evaluation carried out by companies (in percentages).

Number of levels of evaluation	% of companies performing evaluations		
	Formal	Informal	Formal or Informal
No evaluation carried out	59	26	14
Evaluation made at only 1 level	29	32	26
Evaluation carried out at 2 levels	5	25	27
Evaluation carried out at 3 levels	7	8	15
Evaluation carried out at 4 levels	0	7	10
Evaluation carried out at 5 levels	0	2	8

Table 6 allows an examination of the proportion of informal and formal evaluations of CRM training reported to be carried out by the three sizes of company. The data also shows that a greater proportion of large companies carry out formal evaluations of

training and that an assessment at the reactions and behaviour levels are the most frequently performed across all three sizes of company.

Table 6. Percentage of companies carrying out evaluations.

	All (n=113)		Large (n=11)		Medium (n=23)		Small (n=78)	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Reactions	16	44	73	27	13	44	9	47
Attitudes	4	17	18	0	9	34	1	14
Knowledge	9	27	0	9	9	17	10	32
Behaviour	23	31	55	27	26	48	15	29
Organisation	33	0	36	0	52	0	27	0

Discussion

This section reviews the results from each of the nine topics identified in the analysis, before making a number of recommendations relating to CRM training evaluation.

Flight crew CRM training

The results from both the questionnaire and interviews indicate general support for CRM training. However, it is recognised by the industry that for CRM training to be most effective, it should be customised specifically to the culture of the company and to the type of operations being undertaken. The survey showed that this is a particular problem for the small operators in which only 23% of basic and 56.5% of recurrent training is provided in-house.

Assessment of reactions to CRM training

An evaluation of the reactions to training is valuable for the trainers as it provides feedback as to its relevance and where improvements can be made. It is also important to establish whether the participants actually liked the course or thought that it was useful, as crew are unlikely to change their attitudes or behaviour if this is not the case (Kirkpatrick, 1998). The survey indicated that feedback is normally obtained either orally or from a reaction sheet.

The limitations of relying on oral feedback to assess the reactions to training are that it is not anonymous, the outcome can be swayed by a few dominant people and there is a lack of a structure in the discussion. This is not to say that oral feedback should not be used. It should, however, be taken in addition to a reaction questionnaire as it can provide useful information about aspects of the training that have not been covered in the questionnaire. The advantage of using a written feedback questionnaire over oral feedback is that this allows the course participants to make objective comments about the training anonymously.

Assessment of attitudes to CRM training

An evaluation of the attitudes of flight crew to CRM training can be carried out to assess the extent to which the participants have modified their attitudes or values as a result of the training. From the survey and interviews, it was found that the majority of companies did not carry out an evaluation of attitudes and those companies that did relied on informal oral feedback. It is important to measure the attitudes of the flight crew to the concepts covered in the training, as a modification of behaviour is more

likely if a change in attitudes has occurred. The limitations of relying on oral feedback of attitudes are the same as for oral feedback on reactions to training. Therefore it is considered that companies should adopt a more formalised survey method for assessing attitudes to CRM training.

A questionnaire offers an objective and reliable technique for assessing attitude change before and after training and for comparison between different groups. A standard questionnaire that has been developed to do this is the Cockpit Management Attitude Questionnaire (CMAQ). It was developed solely to assess attitudes regarding '*interpersonal components*' of the flight crew's job performance and to link these attitudes to behaviour (Gregorich et al, 1990). The CMAQ was designed more than ten years ago when flight crew were receiving first and second generation CRM training (see Helmreich, Merritt & Wilhelm, 1999). However, modern CRM training has a greater focus on the *cognitive* aspects of the role of the flight crew such as situation awareness, decision making and workload management, which are not explicitly addressed by the CMAQ. Therefore, it may no longer be a relevant assessment technique for measuring the attitudes of course participants to the latest CRM training.

It is important that any attitude instrument is tailored to a particular CRM course so that only those attitudes towards concepts which have been covered in the training are assessed. There is a requirement for a questionnaire with a number of different subscales that have been tested to ensure that they are reliable and discriminatory and which cover all of the possible concepts that are included in CRM training. Thus, if a training course consisted of each of these topics, the operator would use the relevant

sub-scales to construct an attitude questionnaire specifically tailored to the company's training course.

Assessment of knowledge of CRM concepts

Testing students on their recall and understanding of the CRM curriculum can assess learning. The results from the survey and interviews show that if operators carry out knowledge assessment it tends to be on an informal basis. The standard method is to assess acquisition of knowledge using a paper-based test. This could be a reasonably quick and simple way of receiving feedback on knowledge acquisition, and is reported as being used by some operators. The questionnaire could be completed by self-scoring, and used as a mechanism for assessing trends over time and identifying where the training required improvement, as opposed to a technique for evaluating the flight crew.

Assessment of CRM behaviour

Assessment of whether the desired behaviours are being demonstrated was found to be carried out by informal observations rather than by using a more formal behavioural rating system. The survey and interview study found that many of the large airlines have developed their own behavioural marker systems and that these are mostly used for training. However, the survey also found that the majority of companies (88%) do not use a behavioural marker checklist and if any evaluation of behaviour is carried out it tends to be based on informal observational and oral feedback. Whilst this is better than no feedback, there are drawbacks of relying on informal feedback as outlined in the discussion of reactions assessment. Further, a single item relating to CRM training on the technical checklist is not able to provide sufficient detail to allow useful information

to be disseminated to the flight crew. Therefore, to ensure detailed feedback is obtained from CRM training, it is suggested that a behavioural marker system is used.

The lack of a widespread use of formal CRM behaviour assessment in Europe has been corroborated by other studies. A survey of 11 major UK airlines in 1997 showed that only five of them had developed a CRM behavioural markers list, and none of these was used for formal CRM assessment (Flin & Martin, 2001). In the study reported by O'Connor, Höermann, Flin, Goeters et al (in press) it was found that, in a sample of 104 European training captains, only 53% were familiar with a behavioural marker system and only 31% had any experience of evaluating CRM skills. These findings indicate a need for a valid and reliable generic behavioural marker system that could be made available to those operators which do not have the resources or expertise to develop their own systems. However, the provision of a behavioural marker system alone is not sufficient to assess the CRM skills of flight crews. It is also important that instructors who conduct an evaluation have been properly trained and calibrated (Baker, Mulqueen & Dismukes, 2001; Klampfer et al, 2001).

Assessment of organisational effects

The objective of a CRM training programme should be to produce tangible evidence of an effect at an organisational level, such as an improvement in safety and productivity. Therefore, arguably, evidence of an effect at the organisational level is the most valuable evidence of the utility of CRM training. However, it can be very difficult to attribute measured organisational effects to CRM training only given the wide range of

other factors which can have an influence (e.g. changes in regulations, organisational restructuring, aircraft type; Gregorich & Wilhelm, 1993).

The Royal Aeronautical Society Human Factors Group (RAeS, 1999) identified a number of other potential measures of assessing the effects of CRM training. These include: fuel management, punctuality, job satisfaction, insurance costs, and damage to aircraft. However, although large companies are able to track such trends, the smaller companies may not have sufficient flight crew or resource for this to be effective. Incident reporting systems, which may be confidential, provide a mechanism for assessing the effectiveness of CRM training and for identifying areas where recurrent CRM training could address these issues. However, adverse events tend to be infrequent and caution should be taken before redesigning a training course on the basis of one or two incidents.

Another measure which could potentially be used to assess the effects of CRM training is information from flight data quick access recorders (FDQAR). These can produce a multi-channel recording of over 100 flight parameters for every second flown (Bannister, 2001). This information can be used to assess the frequency with which particular events, such as heavy landings, are occurring in a given company or fleet of aircraft. However, it is important to appreciate that it is not possible to assess the human factors causes of the situation (e.g. low situation awareness, poor communication) which can only be inferred.

Methods to support CRM training evaluation

The results from the survey illustrate that the vast majority of the participants thought that the flightdeck, during routine operations, is the most appropriate place for carrying out an assessment of the effectiveness of CRM training. However, as described above, the use of CRM rating scales (e.g. behavioural marker systems) is not widespread. While there is already a clear culture within aviation for assessing technical behaviour on the flightdeck, operators need to be provided with more information and training on how to use behavioural marker systems for assessing non-technical skills.

Reasons why CRM training effectiveness is not evaluated

The main difficulties for aviation operators in carrying out evaluations of CRM training are that the training personnel have limited time, resource and expertise. The individuals involved in CRM training are generally enthusiastic line pilots who have a particular interest in it. Nevertheless, they often feel that they do not have the skill to carry out an in-depth assessment of CRM training. In addition, there is a lack of guidance on how to carry out evaluation of training and tools to facilitate an evaluation. A useful aid would be a resource pack providing information on how to carry out formal evaluations of reactions, attitudes, behaviours and organisational outcomes.

Conclusion

This survey has demonstrated that, despite the recognition by most respondents that CRM training evaluation is beneficial, few companies are actually carrying out a multi-level, formal evaluation of CRM training. However, it is hoped that more guidance and

better availability of techniques for evaluating CRM training will aid companies in the design and implementation of the next generation of CRM training.

References

- Avermaete, v., J.A.G. & Kruijssen, E. (Eds.). (1998). *The evaluation of non-technical skills of multi-pilot aircrew in relation to the JAR-FCL requirements*. NLR: EC NOTECHS (Project report: CR-98443). Amsterdam, The Netherlands.
- 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.
- Bannister, J. (2001). *Airline insurance: An introductory guide*.
Available: www.generali.com/library/airline/impagino.pdf [2001, 3/12/01].
- Civil Aviation Authority. (1998). *United Kingdom aeronautical information circular: Crew Resource Management (AIC 117/1998)*. London: CAA.
- Civil Aviation Authority. (2001). *Size of UK airlines by available capacity: April 2000-March 2001*. Available: www.caaerg.co.uk/ADUPublishedTables/Airline%20Data/200103/012_Size_of_UK_Airlines_in_seat_km_available_and_used.XLS.
- FAA. (1998). *Advisory Circular 120-51C: Crew Resource Management Training (AC No: 120-51C)*: US Department of Transport. Washington: FAA.
- Flin, R. & Martin, L. (2001). Behavioural Markers for CRM: A review of current practice. *International Journal of Aviation Psychology*, 11, 95-118.
- 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.
- Helmreich, R. (2000). *The Line Operations Safety Audit (LOSA). (Version 9)*. Austin: NASA/University of Texas/Federal Aviation Administration Aerospace Group.
- 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.
- Kirkpatrick, D. L. (1998). *Evaluating training programs*. San Fransisco: Berrett-Koehler.
- 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
- O'Connor, Flin & Fletcher (under review). Techniques used to evaluate Crew Resource Management training: A literature review.
- 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*.
- RAeS. HFG (1999). Discussion document: Crew Resource Management, London: RAeS.

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*.