Fall 1994

An Evaluation of Cockpit Resource Management Training in Qantas

P. J. Moore

R. A. Telfer

R. L. Wilkinson

AN EVALUATION OF COCKPIT RESOURCE MANAGEMENT
TRAINING IN QANTAS

P. J. Moore, R. A. Telfer, and R. L. Wilkinson

This study reports an evaluation of a Cockpit Resource Management (CRM) training program at Qantas Airlines in Australia. Four sets of survey data were gathered: Two from crew involved in CRM and a specific Line Oriented Flight Training (LOFT) scenario conducted in a simulator; and two from instructors, one from the LOFT instructors, the other from the CRM instructors. The crew results showed that CRM was perceived to have improved teamwork and leadership and that LOFT was viewed favorably. The instructor results showed generally successful implementation of the decision-making strategies but indicated that certain groups were not sufficiently assertive in communicating their preferred options. The instructors were generally positive in their evaluation of the CRM program. Implications are discussed.

INTRODUCTION

After two years of research and development, CRM training was introduced by Qantas in May 1989. Management consultants (Coopers and Lybrand) were commissioned to aid the airline in designing the course. A committee drawn from the aircrew association and line and training staff synthesized and adapted the consultants' recommendations, using observations from full mission simulator checks of command trainees. It was concluded that the course would be designed to complement standard operating procedures (SOPs). Just as the SOPs ensured the safety standards of routine operations and anticipated emergencies, so the CRM would supply tools to support pilots in the management of these activities.

An outline of the Qantas CRM course

The focus of the course was on aircrew teamwork and decision-making. A model of decision-making was represented by the mnemonic GRADE (Gather information; Review it; Analyze alternatives; Decide; Evaluate outcome). Another mnemonic, TORA, represented the situational factors affecting decision-making.

Represented as a square figure, the situational factors could be represented in the diagram. TORA was derived from the initial letters of the four points of a quadrant formed from the two dimensions of the expertise of crew and the critical nature of the task being undertaken (make the decision Together; Offer the decision-making process to another team member; Refer to others while the captain leads the process; the captain makes the decision Alone). The "refer" area of TORA was seen as the most important aspect.

Attitude was described in terms of a graphical relationship between the effectiveness of an individual's input and the forcefulness of that input, showing a low of "withdrawn submissive" rising to a peak at "supportive assertive," then progressive deterioration to "dominant aggressive."

Leadership was conceptualized in the familiar terms of the Blake and Mouton model (concern for people versus concern for performance) conveyed in terms of three overlapping foci: task, team, and individual (Blake and Mouton, 1982).

The notion of managing upward was clarified by the provision of appropriate support to the leader by judging the degree of assertiveness required. This language of assertiveness was incorporated in the company's Operations Policy Manual in the following four ways:

1. Express personal concern.
2. Define preferred alternatives.
3. Ask for an evaluation.
4. In an emergency use the phrase: "Captain, you must listen!"

The LOFT program

The CRM course was supported by a LOFT program

that met the approval of the Australian Civil Aviation Authority as a cyclic, 12-hour training/license renewal package. It occurs in four simulator sessions per year as part of a three-year program.

A two-day course was designed to train instructors for the LOFT program and to ensure that they consolidated the CRM approach. The first day reviewed the CRM course, and the second day concentrated on LOFT scenario construction, LOFT briefing and debriefing, and a recognition of the issues and principles involved. The LOFT program was implemented in three B747-400, two B767, and two B747-200/300 simulators.

EVALUATION METHODOLOGY

Because baseline data were not available, a pre-post survey design was not possible. Instead, the following four sources were tapped for descriptive and judgmental data:

1. Crew Survey: This was a six-item survey of crew in-line operations, with their responses classified according to aircraft type (747/744/767) and to rank (Captain; F/O; S/O; E/O). The items asked whether, since the introduction of CRM training:

• The use of the GRADE process had significantly improved decision-making skills in the cockpit;

• The skills taught in CRM training had significantly improved crew teamwork;

• The skills taught in CRM training had significantly improved leadership;

• Operational decisions were reached by the crew as a team;

• Crew members were more inclined to be assertive in expressing their operational opinions;

• Senior crew members were more likely to respond to the recommended assertive behavior from other crew members.

2. LOFT Survey: This was a 20-item survey of crew who had undertaken a specific LOFT scenario. The scenario was a night flight from San Francisco to Los Angeles in a complicated weather pattern. There are operational problems associated with the five airports written into the scenario: San Francisco has a crosswind problem; Los Angeles has low clouds and poor visibility; Las Vegas has a problem with approach aids; Sacramento has marginal weather and aid unserviceability, and

Oakland has a runway problem. None of these airports presents any problems under normal conditions and there is no reason to suspect that the flight cannot be completed routinely.

At briefing the crew are warned of a security threat from Gulf War agitators. At about halfway through the flight, a bomb explosion occurs in the aft hold, decompressing the aircraft and failing the number one hydraulic system of nose and body gear extension and inboard trailing edge flaps.

After completing the forced descent, the crew must decide on a course of action. No airport is entirely suitable because of various problems: crosswind with possible control damage; a night approach to a runway not approved by the company; committing to an airport requiring an alternate although the extended gear could not be raised; en route safety height problems, and so on. There is no unqualified correct solution. Responses were classified by aircraft type (744/747/767) and by position of respondent (Captain; F/O; F/E).

The NASA/University of Texas LINE-LOS checklist (Butler, 1991) provided the items for the questionnaire, with responses made on a seven-point Likert scale. This questionnaire, in use with several airlines, assesses eight dimensions of behavior (including communications; decision making; team building and maintenance; workload management; and situational awareness) as well as proficiency and overall effectiveness. Questions on the following 20 topics were asked:

• Awareness of the LOFT scenario and problem;

• The effect of any awareness on the training value of the scenario;

• The extent of the realism;

• The level of difficulty;

• The level of crew performance on the mission;

• The level of personal performance;

• The value of the LOFT session for crew co-ordination training;

• The technical training value of the session;

• The extent of applicable learning from the session;

• Team work;

• Extent of talking or arguing;

• Extent of irritation or frustration;

• Extent of positive contributions to the decision;

Table 1
Survey of crew in-line operations (n = 243)
(means and standard deviations)

ITEM	744	747	767	CAPT	F/O	S/O	E/O	TOTAL
1 Improved cockpit decisions	4.82 (1.42)	4.74 (1.32)	4.79 (1.23)	4.71 (1.42)	4.77 (1.24)	5.12 (1.20)	4.29 (1.53)	4.75 (1.35)
2 Improved teamwork	5.04 (1.32)	4.74 (1.50)	4.91 (1.26)	4.91 (1.36)	4.98 (1.22)	4.97 (1.34)	4.53 (1.66)	4.93 (1.31)
3 Improved leadership	4.24 (1.36)	4.48 (1.35)	4.59 (1.23)	4.29 (1.36)	4.58 (1.27)	4.41 (1.21)	4.29 (1.49)	4.42 (1.32)
4 Team decisions	4.61 (1.47)	4.48 (1.47)	4.97 (1.31)	4.70 (1.55)	5.00 (1.14)	4.35 (1.37)	4.18 (1.38)	4.74 (1.36)
5 Greater assertiveness	5.41 (1.37)	5.35 (1.21)	5.24 (1.16)	5.22 (1.29)	5.47 (1.17)	5.56 (1.19)	5.12 (1.17)	5.34 (1.22)
6 Senior crew more responsive	4.71 (1.24)	4.85 (1.38)	4.80 (1.42)	4.61 (1.29)	4.94 (1.35)	4.88 (1.25)	4.65 (1.50)	4.79 (1.30)

Scale: 1 = Strongly disagree through 7 = Strongly agree

- Unilateral decisions made by the captain;
- Crew assertiveness about their opinions;
- Crew evaluation of decisions;
- Knowledge and helpfulness of the instructor;
- Value of the debriefing;
- Value of the feedback from the videotaping;
- Value of LOFT, overall, as a training technique.

3. LOFT Instructor Survey: This was a 12-item survey of instructors involved with the LOFT exercise covered by the survey above. The responses were categorized according to aircraft type (744/747/767). The items covered the following aspects of crew performance:

- Use of the GRADE process and each of the six stages;
- Team performance on GRADE;
- Choice of diversion;
- Assertiveness of crew in communicating preferred options;
- The captain's role (in terms of the TORA model);
- The crew's operation as a team in the whole process.

4. CRM Instructor Survey: This was an 11-item survey of CRM instructors. The responses were categorized by position: simulator instructor; senior check captain; check

3

captain; check and training captain; promotional training captain; or senior check engineer. The items covered the following:

- Perceived value of the GRADE model as a decision-making tool;
- View of leadership in decision-making;
- Appropriateness of the use of GRADE in operational decisions in emergency situations not covered by a checklist;
- Need for decision-making training;
- Value of the language of assertiveness in CRM training;
- Effect of the REFER process of TORA on the leadership of captains;
- The primacy of task over team or individual in flight, pre-flight and post-flight;
- Extent of groupthink as an aircrew problem;
- Value of the DARTS checklist at briefing;
- Effect of managing upward on flightdeck harmony;
- The balance of emphasis on CRM in Qantas.

RESULTS

It should be noted that in some cases pilots failed to indicate their rank and/or aircraft type and hence their data were not examined from those perspectives, although their data were included in the totals. Given recent trends in the reporting of data like ours (see Lundeberg, Fox, & Puncochar, 1994), we decided to use means and standard deviation to describe our data.

Crew survey

Table 1 shows the results of the survey of 243 flight crew involved in line operations. Given that the scale used a seven-point scale ranging from "Strongly disagree" (1) to "Strongly agree" (7), the overall result is on the positive side of neutral, but not markedly so. The responses bordered toward "slightly agree" for the view that the GRADE process had improved cockpit decision making; that CRM skills had improved crew teamwork and leadership; that operational decisions were being reached by the crew as a team; and that senior crew were more responsive to the recommended assertive behavior from other crew members.

Stronger agreement (between "slightly agree" and "agree") was gained in response to the view that crew were more inclined to be assertive in expressing their operational decisions.

ANOVAs conducted on item scores for the three types (744, 747, 767) and four rank categories (captain, and so on) showed that there were no significant differences between aircraft types or crew status. Worth noting, however, is that the most subordinate crew member (second officer) was most supportive of the effects of GRADE on improved decision-making (Item 1). This result indicates a greater involvement of all team members, perhaps to an unprecedented degree. Flight engineers, however, provided the least support for improved decision-making, their scores being the lowest on this item and others. This finding could have been due, at least in some part, to their attitudes toward change in a context where their redundancy seemed imminent. In addition, there are historical (and spatial) reasons for engineers perceiving themselves to be out of the loop.

A similar pattern of response can be found in Item 4, which stated that operational decisions were reached by the crew as a team. The first officer was most in agreement, the engineer least in agreement.

LOFT survey

The findings are presented in **Table 2.** For Items 3-20 a mean of more than 4.0 indicates that the response has been positive. Items 11, 12, and 14 require reversal because they express a negative view. Given those criteria, an overview of the total results shown in Table 2 supports the view that the LOFT exercise is a valuable training technique. The responses show that the exercise gained a reasonably high approval rating of 5.45 out of 7 (Item 3); that the value of the learning exercise was rated at a reasonably high 5.30 out of 7 (Item 9); that the crews gave themselves a result of 80% (5.63 out of 7, Item 5); and that individuals gave themselves an average mark of 71% on their performance (Item 6). Again, statistical analyses showed no significant differences due to aircraft type.

Narrative additions to the questionnaire indicated the enthusiasm of pilots for LOFT as a form of training. What many found particularly useful was the opportunity to see a situation through to its conclusion, rather than having the simulator re-set after an immediate emergency had been dealt with.

Table 2
Survey of aircrew after LOFT excercise (n = 163)
(means and standard deviations)

ITEM		744		747		767		TOTAL	
		Y	N	Y	N	Y	N	Y	N
1	Awareness of scenario and LOFT problem	56%	45%	21%	79%	51%	49%	68%	32%
2	This reduced training value of LOFT	40%	60%	31%	69%	33%	67%	31%	69%
3	Realism of the scenario	5.55	(1.10)	5.56	(0.81)	5.48	(0.73)	5.45	(0.99)
4	Difficulty of the scenario	4.80	(0.83)	4.82	(0.90)	4.68	(0.91)	4.80	(0.92)
5	How well did the crew perform on the mission?	5.15	(1.04)	5.56	(1.00)	5.57	(1.01)	5.63	(0.98)
6	How well did you personally perform?	4.63	(1.01)	4,85	(0.92)	4.92	(0.91)	4.97	(0.95)
7	Rating of the LOFT session for crew co-ordinating training	5.80	(0.89)	5.68	(1.29)	5.86	(0.95)	5.67	(1.18)
8	Rating of technical training value of this LOFT session	5.20	(1.40)	5.46	(1.28)	5.57	(1.04)	5.41	(1.22)
9	How much have you learned to use on the line?	5.20	(0.77)	5.39	(1.04)	5.44	(0.94)	5.30	(1.09)
10	Our crew really worked as a team	5.84	(0.59)	5.95	(0.64)	6.11	(0.57)	5.99	(0.67)
11	We spent too much time talking or arguing	1.75	(0.44)	2.00	(0.91)	1.84	(0.55)	1.87	(0.89)
12	Dealing with crew left me irritated and frustrated	1.55	(0.60)	1.68	(0.90)	1.95	(1.43)	1.67	(1.00)
13	All crew made a positive contribution to decisions	5.95	(1.05)	5.89	(1.10)	6.30	(0.57)	6.00	(1.05)
14	Captain made most decisions, not involving others	3.20	(1.54)	2.80	(1.55)	2.68	(1.55)	2.94	(1.64)
15	When necessary, crew members asserted opinions	5.75	(1.16)	5.75	(0.74)	5.97	(0.55)	5.82	(0.97)
16	Our crew evaluated its decisions	4.90	(1.65)	5.23	(1.22)	5.30	(1.10)	5.23	(1.36)
17	Our LOFT instructor was knowledgeable and helpful	5.90	(0.91)	6.05	(0.80)	6.35	(0.59)	6.14	(0.77)
18	The debriefing after LOFT was highly useful for all crew	5.95	(0.89)	5.67	(1.23)	5.81	(1.00)	5.66	(1.18)
19	The videotape of the LOFT provided important feedback	4.90	(1.59)	4.62	(1.72)	4.44	(1.80)	4.47	(1.75)
20	Overall, LOFT is an extremely useful training technique.	5.90	(1.59)	5.64	(1.00)	6.08	(1.10)	5.81	(1.00)

Note: Response on a 7-point scale, negative low, positive high, for Items 3-20

When LOFT scenarios are being designed it is difficult for both the designer and/or the instructor to design one that is both realistic and novel. Most participants were aware of the LOFT scenario, but the awareness differed widely across fleets (compare the 747 and 744 results for Item 1). Such a difference could be linked to company operations for the various fleets and the extent to which the scenario is part of these operations. That more than a third of the participants saw the lack of novelty as a problem may indicate a need for greater exposition by instructors on the inevitability of awareness if scenarios are to be as realistic as possible.

The response was not as predictable on Item 14 (The captain made most decisions, not involving others). In the previous reversal items (11 and 12), the means were in the range of 1.55 to 2.0, with totals of 1.87 and 1.67. In Item 14, however, the range was 2.68 to 3.2, with a total of 2.94. The upper extreme in the range was the 744 crew, whose mean rating of 3.2 meant that they were on the "slightly disagree" side of "neutral." This is evidence that GRADE/TORA were not being implemented in the 744 two-crew crews as well as they were in the three-crew 747 or the two-crew 767. The captains who were more likely to make more decisions alone were generally the older, more senior captains. However, in the absence of more detailed knowledge of the individuals involved, it is difficult to postulate other bases for these differences.

LOFT instructor survey

The results are presented in **Table 3**. Items 1 and 5 had a dichotomous response, clearly indicating a perception of successful implementation of the GRADE process and an observation by instructors of crews actually implementing the decision model. Item 8 indicates the range of selected diversions, the spread indicating the robustness of the scenario in that crews were able to justify three different choices of alternate.

For the remaining items, the midpoint of 4 on the 7-point scale provides a criterion for interpretation. Despite the fact that most instructors perceived (81 of the 95, Item 1) that the GRADE process was identifiable and that the crew "decided" (89, Item 5), the instructors were far less positive in their ratings of the ways in which the process was employed by the crew as a team. Gathering (Item 2 mean = 3.31) and reviewing (Item 3

mean = 3.10) were performed better than the analysis of the information (Item 4 mean = 2.78), evaluation of that information (Item 6 mean = 2.36), and crew performance as a team in the decision-making process (Item 7 mean = 2.94). Yet the instructors perceived that the crews operated as teams (Item 12 mean = 3.80). The data suggest that the gathering process was performed better than the succeeding stages of the GRADE process, in the eyes of the instructors.

The results of Items 9, 10, and 11 can be linked by their consistency. The instructors thought that neither the F/O or E/O were assertive in communicating their preferred options, and that the captain tended to make the decisions alone.

It is possible, too, that the order of the presentation of the questionnaire items—starting with GRADE, then dealing with its components—may have induced a response set. It may be useful in future studies to reverse this sequence.

CRM instructor survey

The CRM instructors were far more positive in their evaluation of the effectiveness of decisional and leaderships functions introduced by the CRM course. Modal scores shown in **Table 4** can be interpreted by means of a midpoint of 4 on the 7-point scale. All results fall above this point, suggesting a response set when one considers that Item 4 (Qantas crews did not require any training in decision-making); Item 8 (Groupthink is never a problem); Item 10 (Managing upward leads to disharmony); and Item 11 (There is too much emphasis on CRM) could reasonably have been expected to have a modal response at least closer to the midpoint, if not below it.

DISCUSSION

The results raise several questions on evaluation methodology, attitudes and change, crew behaviors in different aircraft types, the effect of video replay on participants' self-evaluation of their performance, scenario construction for LOFT, and the consequences of a perceived dichotomy between theory and practice. Such a dichotomy precludes uniform implementation of CRM within an airline and across the industry. As Helmreich (1993) has pointed out: "The implementation of CRM will not be complete until acceptance of its contents is

more uniform among the organizations that use it."

First, for evaluation of CRM to be optimized, base-line data are needed. The critical point for the design of evaluation is when an airline decides on its rationale for CRM. With the CRM objectives come the criteria for its evaluation. From the criteria come the sources of evaluative data, and the scheduling of their periodic collection by means of observation, surveys, interviews, company operating statistics, and so on. The same criteria are the bases for the derivation of evaluation instruments. The instruments then need to be applied in a field trial, as tests of reliability and validity, before being applied prior to the initial CRM to establish base-line data.

Formative evaluation, conducted systematically during the progress of CRM, can be of value to participants and presenters during the course of a CRM program. It can help participants to ensure that aspects of high personal relevance and significance are incorporated, and it can provide vital feedback to presenters. In brief, the evaluation of CRM is as critical as the program itself. This importance needs to be reflected in both the budget and in critical-path planning before the implementation of CRM programs.

Second, CRM programs are essentially concerned with improving the quality of crew behavior. This is an attempt to change the way human beings behave. It is insufficient to simply show people a "better" way. There is a need for sustained impetus in a variety of ways: support of management and peers, bulletins, notices, recognition by superiors, and so on.

In their 1986 survey of Qantas captains, consultants concluded that 75% of the captains did not include other flight crew in operational decision-making. Neither did crew members act assertively to support the leader. The Cockpit Resource Management Course was introduced and, although receiving enthusiastic support from the aircrew associations and those who volunteered to help compile the course, the majority greeted the introduction of CRM training with either indifference or hostility.

Response to the course itself was positive. The participants were supportive and, from all reports, enjoyed and actively participated in the course. A survey carried out on completion of each course indicated that most participants believed the course was very professionally compiled and presented. They almost unanimously agreed that the concepts were valuable and useful. However, when asked whether the CRM course would change the way they conducted their day-to-day aircraft operations, the answer was a resounding "negative!" (Wilkinson, 1991).

Perhaps one of the implications is that this demonstrable gap between concept and practice has to be bridged. LOFT training is one means of aiding this transition. Others could include brief follow-up refresher courses and on-the-job activities.

Crew behavior in decision-making appeared to vary with group size. In dyads (744, 767) captains tended to make fewer decisions alone than those in the 747 triads. Further, 767 first officers tended to be more assertive than 747 first officers. This result supports further research of the type described by Clothier (1991) into behavioral interactions across various aircraft types. It is relevant to note the extension of this line of research into cross-cultural perspectives of CRM (Johnston, 1992) in which some of the major contextual variables of aeronautical decision-making and judgment are provided.

Third, there is the question of pilots' self-evaluation. Are they their own worst critics? Crew rated their individual performances lower than the team performance. Ideally, this would be the result of observation of synergy in action, when the achievement of group power applied to problem-resolution was graphically evident. To test this hypothesis it would be valuable to ensure when video is employed in CRM that participants were desensitized to its application, and that a degree of detachment could be introduced into self-evaluation. Perhaps a coarse rating scale or some fairly relaxed group trials at evaluation of, say, the instructor on videotape may be a suitable orientation to remove the possibility that such results could be gained by self-imposition of unnecessarily rigorous standards.

Fourth, there is the difficulty involved in scenario construction for LOFT. On the one hand, line-orientation implies the probability of actual occurrence in line flying. On the other hand, the scenario has to provide the challenge of novelty and quintessential decision-making in terms of process and consequence.

Table 3
Survey of instructors of the LOFT excercise (n = 95)
(means and standard deviations)

	ITEM	RESPONSE
1	Was GRADE process identifiable?	Y - 81 N - 14
2	Did crew "Gather Information"? (new - old)	3.31 (1.52)
3	Did crew "Review the Information"?	3.10 (1.40)
4	Did crew "Analyse the Alternatives"?	2.78 (1.60)
5	Did crew "Decide"?	Y - 89 N - 5
6	Did crew "Evaluate"? (not at all - fully)	2.36 (1.75)
7	Was Grade process performed as a team?	2.94 (1.47)
8	Which diversion was selected?	SFO - 33 LAX - 23 OAIC - 3 NGO - 30 Others - 6
9	Was F/O assertive in communicating preferred option? (not at all - very)	2.92 (1.45)
10	Was E/O assertive in communicating preferred option?	2.67 (1.56)
11	Did the Captain involve others in problem solving?	3.03 (1.46)
12	Did the crew operate as a team? (badly - well)	3.80 (1.12)

Note: Items 2-12, 7-point scale, extremes indicated in brackets

Table 4
Survey of instructors in CRM (n = 79)

ITEM	RESPONSE	
1. The GRADE decision making process is a useful tool in decision making.	6.29	(0.60)
2. A good leader always makes quick and effective decisions by himself.	5.67	(1.24)
3. The GRADE process is appropriate for operational decision in emergency situations not covered by a checklist.	5.41	(1.45)
4. Qantas crews did not require any training in decision making.	5.82	(1.06)
5. Teaching the recommended language of assertiveness is an important aspect of CRM training.	5.92	(1.33)
6. The encouragement of Captains to operate in the REFER box of the TORA chart leads to weak leadership.	5.75	(1.19)
7. Captains should generally concentrate upon the task to be achieved. The TEAM and/or INDIVIDUAL are a pre-flight and post-flight consideration.	5.53	(1.22)
8. GROUPTHINK is never a problem amongst Qantas aircrew.	5.76	(1.26)
9. The DARTS checklist is a useful tool at briefing.	4.94	(1.35)
10. MANAGING UPWARDS leads to disharmony on the flight deck.	5.43	(1.17)
11. There is too much emphasis on CRM in Qantas.	4.85	(1.64)

Reconciling these two extremes is further complicated by the fact that the professional pilot is continually discussing and updating skills, knowledge, and procedures. Recent research into pilot learning indicates that successful airline pilots use discussion with their colleagues as an important source of learning (Moore, 1991; Telfer, 1991). Perhaps LOFT scenarios based on actual occurrence stimulate such discussion among pilots.

Finally, the evaluation of this CRM program demonstrated the gap that can often exist between theory and practice. The gap is one that is usually more likely in educational rather than training programs. A training program usually has a narrow focus on a well-defined task or skill, with a demonstrable application. Education,

however, usually deals with reasons for human performance, as well as the nature of the activity itself. Given the broader focus, it has to incorporate vagaries and variables associated with human motivation, self concept, norms, attitudes, and values. It cannot be dogmatic or sweepingly universal. For those associated with the specificity of flight training and professional aviation, the lack of prescription can be frustrating. The lack of a unitary theory to govern the way people behave, in contrast to why wings generate lift, is a matter that merits explicit (rather than implicit) inclusion in CRM. This discussion may help participants recognize how the designers chose to emphasize the aspects included in the course, and the constraints under which they operated.

CONCLUSION

There are a number of methodological factors to be considered if evaluation of CRM is to be valid and reliable. Base-line data are vital, and instruments require demonstrable reliability derived from rigorous field testing. The need for formative and summative evaluation is a critical component in the design of CRM programs, and requires consideration in the initial stages rather than some time after the program has been operating.

EPILOGUE

In 1990 a Qantas 747 departed Cairns on a seven and a half hour flight to Narita. Near the top of the climb the aircraft began to vibrate and control difficulties were experienced because, unknown to the crew, part of the wing had detached. Over the next hour the crew grappled with wrong advice, pressure from schedulers on the disposition of the aircraft for servicing, and an attempt by air traffic control to have another airliner formate on them to inspect the damage.

The aircraft landed safely at a diversion airfield. Analysis of the decision showed that both the process and result were exemplary. The captain pointed out that he had deliberately applied CRM principles that he had practiced in LOFT, and regarded his ability to handle the in-flight emergency as a direct result of that training. That is the ultimate evaluation of CRM.□

Phillip J. Moore holds a Ph.D. in Educational Psychology from the University of Newcastle, Australia, where he is an Associate Professor in the Department of Education.

Ross A. Telfer is an Emeritus Professor of the University of Newcastle, Australia, and works as an airline consultant.

Richard L. Wilkinson is a line pilot on 747 aircraft. He has experience flying fighters in the Royal Navy and was a Qantas official.

REFERENCES

Blake, R. R., and Mouton, J. S. (1982). *Cockpit resource management*. Scientific Methods Inc.

Butler, R. (1991). Lessons from cross-fleet/cross airline observations: Evaluating the impact of CRM/LOS training. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology*. Columbus: Ohio State University.

Clothier, C. C. (1991). *Behavioral interactions across various aircraft types*. Paper presented at the Sixth International Symposium on Aviation Psychology, Ohio State University, Columbus.

Helmreich, R. L. (1993, September). Future directions in crew resource management training. *ICAO Journal, 48*,(7), 8-9.

Johnston, N. (1992). CRM: Cross cultural perspectives. Dublin: Aviation Psychology Research Group, Department of

10

Psychology, Trinity College.

Lundeberg, M. A., Fox, P. W., & Puncochar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology, 86,* 114-121.

Moore, P. J. (1991). *Approaches to learning in aviation: Does the difficulty of the topic make a difference?* Paper presented at the Human Factors in South Pacific Aviation Symposium, New Zealand Psychological Society Annual Conference, Massey University, Palmerston North.

Telfer, R. (1991). *How airline pilots learn.* Paper presented at the Human Factors in South Pacific Aviation Symposium, New Zealand Psychological Society Annual Conference, Massey University, Palmerston North.

Wilkinson, R. L. (1991). *The simulator in Qantas cockpit resource management training.* Paper presented to the Royal Aeronautical Society, London.☐