Running head: FACILITATING LOFT DEBRIEFINGS

Facilitating LOFT Debriefings: Instructor Techniques and Crew Participation

R. Key Dismukes

NASA Ames Research Center, Moffett Field, CA

Lori K. McDonnell and Kimberly K. Jobe

San Jose State University

NASA Ames Research Center, Moffett Field, CA

Abstract

This study analyzes techniques instructor pilots (IPs) use to facilitate crew analysis and evaluation of LOFT performance. We analyzed IP facilitation and crew participation for 36 debriefing sessions conducted at five U.S. airlines. For this analysis we developed a rating instrument called the Debriefing Assessment Battery (DAB) and demonstrated that it can be used reliably. IP facilitation skill varied dramatically, suggesting a need for concrete hands-on training in facilitation techniques. All measures of crew participation correlated significantly with IP effectiveness in facilitation. Crews responded to IP guidance but did not lead their own debriefings. We suggest ways to improve debriefing effectiveness.

Facilitating LOFT Debriefings: Instructor Techniques and Crew Participation

Line-Oriented Flight Training (LOFT) provides aircrews the opportunity to practice Crew Resource Management (CRM) techniques in realistic and challenging simulated flight situations. How much crews learn in LOFT and take back to the line depends in part on the effectiveness of the debriefing that follows the LOFT (Helmreich & Foushee, 1993). The simulation itself is a busy, intense experience; thoughtful discussion of the experience helps crews sort out and interpret what happened and why. The FAA Advisory Circular on Line Oriented Simulation (1990) states that instructor pilots (IPs) should lead debriefings in a way that encourages crew members to analyze their LOFT performance for themselves. Rather than lecture to the crew on what they did right and wrong, the IP should facilitate self-analysis and self-critique by the crew (Butler, 1993; Hawkins, 1987; Klair, 1997; Smith, 1994). The adult-learning literature suggests that this crew-centered approach provides deeper learning and better retention (Duvall & Wicklund, 1972: Gow & Kember, 1993; Jones, 1982; see also Dismukes, Jobe, & McDonnell, 1997, and references therein). Crew-centered debriefings may also help crews develop the habit and skill of analyzing their own CRM performance on the line and conducting their own crew debriefings following line operations (Butler, 1993).

Although the concept of facilitated debriefings is widely espoused in the CRM literature, little empirical research has examined what actually happens in debriefings. This study addressed five major questions:

1) To what extent do IPs attempt to facilitate crew participation and self-analysis in LOFT debriefings?
2) What techniques do IPs use to facilitate, and how effective are these techniques?
3) How much variation occurs among IPs and among airlines in the conduct of debriefings?
4) What is the character of crew participation, especially in terms of analyzing and evaluating their own performance?
5) Is facilitation a viable approach to encouraging crew participation and self-analysis?

Method

Participants and Procedures

We observed, made audio recordings from, and later transcribed the debriefings of 36 U.S. airline crews who were completing their annual recurrent LOFT training between June 1994 and May 1995 (Dismukes, et al., 1997). The crews were from five large, well-established international companies; four were passenger airlines and one was a cargo company[1]. The letter codes V, W, X, Y and Z were substituted for airline names in the data, and all references to individuals and airlines were deleted. The number of crews observed at each airline ranged from four to nine (Table 1). Twenty-five of the 36 were two-person crews, and 11 were three-person crews.

We interviewed training department managers and IPs for general comments about LOFT debriefing and IP training. The airlines' preexisting training schedules for the period of data collection determined the selection of IPs and crews. The observed debriefings represented all or most of the fleets operated by each airline, and at least one LOFT simulation of each scenario flown in each fleet was observed. We generally observed one debriefing per IP and crew; however, for the purpose of comparison, we observed four of the instructors debriefing a second crew. Each IP and each crewmember gave permission for their debriefing to be observed and audiotaped. After each debriefing the IP rated the crew's CRM performance and technical performance in the LOFT on five-point scales ranging from poor (1) to exemplary (5).

Measures

Descriptive measures. The descriptive measures discussed in this paper are grouped into three categories: LOFT Ratings, Participation Measures, and Discussion Topics (Table 2). LOFT Ratings are the IP ratings of crew performance in the LOFT. Participation Measures include ten

---

[1] Our findings may not be representative of smaller, regional, or newly-started airlines, especially those still developing CRM and LOFT programs. Also, several airlines have recently started upgrading their facilitation training for instructors, partly in response to the preliminary reports of this study.

variables which characterize the nature of IP and crew participation. Discussion Topics include nine variables which characterize the content of IP and crew discussion.

Debriefing Assessment Battery (DAB). We developed a battery with which to assess IP and crew participation in the debriefing (Appendix). The battery was developed from a longer version previously described in Dismukes, et al., 1997. A secondary purpose for developing the short form of the battery was to provide an instrument that could be used in laboratory and field settings for either research or training.

Five of the eight IP battery items address how the IP conducts the debriefing and uses facilitation techniques; the other three items concern the content on which the instructor focuses. Three of the six crew items concern the content of crew discussion; the other three items concern the character of crew participation. Each item is rated on a seven-point Likert scale ranging from poor (1) to outstanding (7). The *IP Overall Score is* the average of the scores on the eight IP items and the *Crew Overall Score* is the average of the six crew items. The IP items and the crew items correspond to the objectives of debriefing stated in AC 120-35B (FAA, 1990) and the concepts of facilitation and adult learning (Dismukes et al., 1997).

Two raters scored the debriefings after listening to the audiotapes without pausing. While listening to the audiotapes, raters read the debriefing transcript to clarify indistinct utterances but did not preview the transcript or review it afterwards. Each audiotape was heard twice; the first pass was used to score the eight IP items and the second pass was used to score the six crew items.

Two raters trained and then independently rated 13 tapes. Interrater reliability (Pearson correlation coefficients) for these 13 tapes ranged from .56 to .99 for IP items (.89 for overall score) and ranged from .69 to .99 for crew items (.99 for overall score). One rater then coded an additional seven tapes for a total of 20. After an interval of 18 months the rater repeated the rating of four audiotapes. Retest reliability of instructor items ranged from .64 to .99 (.99 for overall score) and ranged from .33 to .98 for crew items (.91 for overall score). The rater then coded the remaining 16 tapes. The DAB data reported below are the ratings of the 36 audiotapes by that rater.

Results

General Observations

At all five airlines most debriefings did not occur immediately after the LOFT. Instead, after a short break, the IP and crew first returned to the simulator to conduct about two hours of practice maneuvers as rehearsal for the proficiency check that would follow the next day. A few IPs, apparently on their own initiative when scheduling allowed, reversed the order so they could debrief the LOFT first.

At all airlines most debriefings followed the same general format. The IP gave either a very short introduction or none at all[2] and then led a discussion of segments of the LOFT in chronological order. Rarely did the IP engage the crew in setting an agenda for discussion, although some IPs invited general comments on the LOFT before starting discussion of specific segments. IPs from the four airlines with video equipment generally used a video segment to begin the discussion of related portions of the LOFT. The average duration of the debriefings was 30.7 minutes, with a range of 8 to 82 minutes.

Descriptive Data

For the sake of brevity, data are collapsed across airlines and across two and three person crews, except for variables for which significant differences occurred.

IP ratings of crew performance. Ratings averaged 3.6 (SD = .90) for CRM and 3.5 (SD = .89) for technical on the 1 to 5 scale. Duration was negatively correlated[3] with IP ratings of crew CRM performance ($r = -.49$, $p < .01$) and technical performance ($r = -.39$, $p < .05$) and positively correlated with the proportion of IP words directed to negative aspects of crew performance or ways

---

[2] Typically, before the LOFT the IP conducted a 90 minute briefing that reviewed a wide range of topics and activities that took place in the two-day session of which the LOFT was a part. We did not collect data from these briefings but observed about a dozen. All of the IPs mentioned that a debriefing would occur but only one described in detail how the debriefing would be conducted.

[3] Spearman coefficients were used for all correlations.

to improve ($r = .51$, $p < .01$). This suggests that IPs spent somewhat more time with crews whose performance in the LOFT was problematic.

Participation Measures. On average, IPs talked[4] much more than any of the crewmembers; with two-person crews IPs talked more than the Captain (CA) and First officer (FO) combined (Table 3). However, individual IPs differed substantially; for example, the percentage of talking by IPs with three-person crews ranged from 17 to 87%. CAs talked slightly more than FOs in two-person crews, but in three-person crews the amount of talking by CAs and FOs was nearly identical. Flight engineers (FEs) talked less than CAs or FOs, but none of the differences between crewmembers were statistically significant[5].

To examine the amount of discussion directly between crewmembers, we analyzed sequences of crew utterances in blocks, each block beginning with the first crew utterance after an IP utterance and continuing until the IP spoke again. These blocks were mostly very short; 80% of them consisted of only one utterance by a crewmember before the IP spoke again; thus, in these blocks there was no verbal crew interaction at all. Eleven percent of the blocks contained only two crew utterances and 8% of the blocks contained three or more utterances by crewmembers[6].

Most IPs asked a large *number of questions*, averaging 48 (SD = 23) per hour among two-person crews and 59 (SD = 65) per hour among three-person crews. Sixty percent of these questions were *directed* to specific crewmembers in two-person crews and 54% were *directed* in three-person crews.

Thirty-three percent of crew utterances were *responses* to IP or crew questions, 30% were *substantive statements* that added content to the discussion, 30% were *non-substantive statements*, most of which involved maintenance of discourse (e.g., "I see what you mean"), and 7% were *questions*. The distribution of words (percent of all crew words) among these categories was:

---

[4] Percentage of all group words uttered by each participant.

[5] This paper uses a rejection criterion of .05 throughout.

[6] Roundoff accounts for one percent.

responses, 44%; substantive statements, 45%; nonsubstantive statements, 7%; and questions, 4%.

Thirty-one percent of crew questions were *"proactive"* (addressed debriefing content, raised new issues or brought new information into the discussion); the remainder involved discourse maintenance or were extraneous.

Content of discussion. We considered utterances to address CRM if they involved any of the topics listed in AC 120-51B, Crew Resource Management Training (FAA, 1995) and to address technical issues if they involved techniques of flying, navigation, or systems operation. Many utterances involved some combination of CRM and technical issues. We arbitrarily coded utterances as *CRM* if 2/3 or more of the words pertained to CRM topics, *Technical* if 2/3 or more of the words pertained to technical topics, and *Mixed* if the mixture was between 1/3 and 2/3. The average percentage of words directed to CRM by IPs varied from 19 to 64, with significant differences among the airlines (Table 4). The average percentage directed to *CRM* by crews varied from 25 to 68, also with significant differences among airlines. However, when *CRM* and *Mixed* scores are added together, the differences among airlines are substantially less. At all five airlines the percent of crew words directed to *CRM + Mixed* topics was substantially greater on average than the percent directed to *Technical* topics. The percent of IP words was also substantially greater at four of the five airlines.

Discussion of crew performance. On average, 41% of IP words and 52% of crew words were directed to the performance (CRM, technical, or both) of the crew in the LOFT (Table 5). IPs devoted substantially more discussion to positive aspects of crew performance than to negative aspects or ways to improve performance. Most of the crews' words concerning performance were neutral descriptions of what they did, with much less discussion of positive aspects, negative aspects, or ways to improve.

The percentages of crew words were correlated to the percentages of IP words directed to discussion of CRM topics ($r = .76$, $p<.01$), technical topics ($r = .85$, $p<.01$), positive aspects of performance ($r = .35$, $p<.05$), negative aspects of performance ($r = .61$, $p<.01$), and ways to improve performance ($r = .67$, $p<.01$).

Debriefing Assessment Battery

The item *Explains the Debriefing Process scored* significantly lower than the other IP items. IPs of three-person crews scored significantly higher[7] than IPs of two-person crews for the items *Encourages Crew Analysis*, *Encourages Crew Evaluation*, *Use of Questions*, and *Overall Score* (Table 6). The average *Overall Score* was 3.0 for IPs of two-person crews and 4.0 for IPs of three-person crews. (The midpoint of the scale, *Adequate*, is 4.0). Three-person crews scored significantly higher than two-person crews on all items except *Analyze in Depth* and *Respond in Depth*, and these two items showed the same trend (Table 7). The average *Overall Score* was 3.1 for two-person crews and 4.4 for three-person crews. The item *Interact with Each Other* scored lower than other crew items, however the differences were not significant except between *Interact with Each Other* and *Focus on CRM*, the highest scoring item.

Very few of the differences between airlines on the DAB items were statistically significant. Among IP items Airline Y scores were significantly higher than Airline Z scores on *Explains Debriefing Process*, and higher than airline W scores on *Use of Questions*. Among crew items Airline Y scores were significantly higher than Airline W items on *Evaluate Performance, Respond in Depth to Questions*, and *Overall Score.*

Battery correlations. All IP items except *Use of Silence* were significantly correlated with several crew items, and all crew items were significantly correlated with several IP items (Table 8). The crew item *Analyze in Depth* was most strongly correlated with the IP items *Encourages Crew Analysis* (.72) and *Overall Score* (.72). The item *Evaluate Performance* was most strongly correlated with *Encourages Crew Evaluation* (.57) and *Encourages Crew Analysis* (.55). *Focus on CRM was* most strongly correlated with *IP Focus on CRM* (.72), *Use of Questions* (.70), and *IP Overall Score* (.73). *Proactive Participation* was most strongly correlated with *Focus on Crew* (.50) and *IP Overall Score* (.48). *Respond in Depth was* most strongly correlated with *Use of*

[7] Significance was examined by t-test of the means of the combined airline scores. ANOVA could not be used because some cells from three of the airlines contained no sample or only one sample.

Questions (.62). *Interact with Each Other* was most strongly correlated with *Encourages Crew Evaluation* (.47), *Focus on Crew Participation* (.43) and *Overall Score* (.43). *Crew Overall Score* was most strongly correlated with *Use of Questions* (.68) and *IP Overall Score* (.68).

The average of the intercorrelations among the eight IP items was .56 (range: .08 to .88), and the average of the intercorrelations among the six crew items was .59 (range: .37 to .77).

Discussion

Large differences occurred within each airline for most IP and crew variables. Substantial differences also occurred between airlines in average values for some variables, though few of these differences across airlines were statistically significant (one major exception is emphasis on CRM, discussed below). The large variability among IPs and the relatively small sample size preclude conclusions about differences among airlines from these data. The large variability among IPs at each airline has major implications for standardization and training (discussed below).

On several occasions crewmembers spontaneously mentioned that they had trouble remembering relevant aspects from the LOFT. The common practice of delaying the debriefing two hours or more until after the maneuver practice may have contributed to this memory difficulty. Performing the practice maneuvers before the debriefing, in the same cab as the LOFT and under similar conditions, is likely to interfere with the memory of the preceding LOFT. Unfortunately, we have no data addressing this issue, but suggest that it deserves empirical study.

IP Facilitation and Crew Participation

Most IPs talked substantially more than any of the crewmembers. In many cases the IPs talked more than all the crewmembers combined. Discussion typically revolved around the IPs' questions and comments. Back and forth discussion among crewmembers was quite limited, and crewmembers asked few questions.

Most IPs asked a large number of questions, almost a question per minute on average. Most crewmembers responded readily to questions, however, crew responses and self-initiated statements tended to be brief. Long, probing, or deeply analytical utterances were rare.

One might wonder if the percent participation by the IP might be driven by the crew; an IP

might be forced to do more of the talking if he or she tried unsuccessfully to induce the crew to participate. However, the data suggest this is not the case: percent participation by IPs was significantly negatively correlated with all eight of the IP items on the battery (average correlation with items was .57). Thus, for example, IPs who talked a lot were less likely to use questions effectively.

The data suggest that the character of crew participation in many debriefings was not optimal for learning and retention. Average crew scores on the DAB items fell around the midpoint of the scale ("adequate") or below. We designed the crew items of the DAB on the premise that experienced pilots would learn best by analyzing the situations that confronted them in the LOFT, evaluating their performance, and interacting directly with fellow crew members (FAA, 1990). Direct interaction among crewmembers may have several benefits: it puts greater responsibility on the crew, which may enhance the depth of what they learn; and it provides them with an opportunity to practice debriefing as they would conduct it on the line. Also, IPs sometimes note that pilots will accept an observation from a fellow crewmember more readily than from an IP. To varying degrees, the IPs led the crews through analysis and evaluation in a step by step fashion, and the crews undoubtedly benefited, but their participation was limited in depth and more reactive than proactive.

We interpret the combined descriptive and DAB data as indicating that IPs attempted to elicit crew participation but retained moment-to-moment control of the discussion, rather than attempting to assist the crew conduct their own debriefing. Some IPs may have felt that it was appropriate or necessary to keep close control of the discussion. Facilitating crew discussion takes longer than lecturing to the crew, and some of the IPs may have been responding to time constraints. Other IPs might not have realized that they were dominating the discussion or might not have known how to get the crew to take a more proactive role.

The DAB scores reveal that some IPs used facilitation effectively. However, most IPs, especially those with two-person crews, showed substantial room for improvement in use of

facilitation techniques[8]. Our subjective observations are consistent with the DAB data. The most striking deficiency was that most IPs failed to explain to the crew at the beginning of the debriefing how the debriefing would be conducted, what role the IP would play, and what the crew was expected to do. Even if IPs discuss the debriefing process during the pre-LOFT briefing, it is important to begin the debriefing with an explicit reminder of how it will be conducted. Other common mistakes included failing to pause (use silence) when the crew did not respond immediately to questions, keeping the discussion centered on the IP instead of encouraging the crew to interact with each other, making long soliloquies, evaluating crew performance before eliciting crew self-evaluation, failing to push beyond superficial description of events, and not getting the crew to analyze why things went well.

Clearly, these debriefings were not led by the crews. It is not certain how effectively the crews might have led their own debriefings had they been charged to do so by the IPs. However, beyond the issue of who should lead the debriefing, it is clear that IPs substantially influenced the amount of crew participation and the focus of the crew's remarks. The six crew items on the DAB were moderately to highly correlated with IP items. Nearly half of the variance of the crew Overall Score can be accounted for in terms of IP use of facilitation techniques. These data support the argument that facilitation is an effective way to encourage and guide crew participation in debriefing.

Effects of Crew Position and Crew Size

Percent participation by CAs and FOs was quite similar, although CA participation was slightly higher in two-person crews. FE participation was substantially less than that of CAs and FOs, however only the latter difference was statistically significant. If airlines desire equal participation from FEs, IPs may need to pay special attention to eliciting their participation.

The combined participation of three-member crews was greater than the combined participation

---

[8] Our subjective impression is that the IPs' traditional instructional skills were quite high: their technical knowledge, people skills, and commitment were clearly very strong. Thus the issue seems to be IP training in facilitation, discussed in a later section.

of two-member crews. The participation added by the FE in three-person crews was compensated for by a corresponding decrease in IP participation, rather than by decreased participation by CAs and FOs.

Three-person crews scored significantly higher than two-person crews on most DAB items. The reason for this difference is not clear, but it might be partially accounted for by the greater combined participation of the three-person crews. In scoring the crew items the two raters attempted to respond to both the quality of discussion involving each item and the amount of discussion per unit time. The greater participation by three-person crews would increase the amount of crew discussion measured by each item, other things being equal. This interpretation is supported by significant correlation of each crew item with the descriptive variable *percent crew participation* (average correlation was .64).

Another potential explanation is that the nature of crew duties in three-person cockpits may lead them to interact with each other more than two-person crews do. For example, in trouble-shooting abnormal situations, CAs often work the problem with the FE while the FO flies the airplane. Three-person crews might carry this habit of verbal interaction over into the LOFT debriefing. Also, having three crewmembers changes the dynamics of discussion in the debriefing and affords more opportunities for crew interaction.

IP scores for several DAB items were also significantly higher with three-person crews. It may be that these IPs took advantage of the greater level of crew participation and interaction to play a less dominant role.

Content of Discussion

Large, statistically significant differences occurred among the airlines in the percent of discussion scored as *CRM*, however when the categories *CRM* and *Mixed* were added together the differences among airlines were much smaller. It is appropriate to discuss CRM and technical issues in an integrated manner so that crews come to understand CRM not just as an abstraction but as a practical approach to managing tasks in line operations. On average, at all but one of the

airlines, the *CRM + Mixed* categories occupied substantially more of the discussion than did the *Technical* category, which is consistent with the goals of LOFT

At each of the five airlines very large differences occurred among IPs in emphasis on CRM; at one airline, for example, *CRM + Mixed* ranged from 30 to 94% of IP words. In some cases the IP may have emphasized technical issues because of inadequacies in the crew's technical performance, however even in such cases it seems inappropriate to spend only a few minutes discussing how CRM or the lack of CRM affected the flight.

Discussion of crew LOFT performance was emphasized in most debriefings, averaging about half of IP and crew words. This figure was fairly consistent across airlines. A large portion of IP comments on performance were positive, and this is consistent with the objective of reinforcing the crews with positive feedback. In contrast, IPs and crews directed only a small percentage of their discussion to problematic aspects of crew performance or ways to improve performance, even though IPs tended to hold longer sessions for crews who did not perform as well. This lack of emphasis on ways to improve seems inconsistent with the objectives of LOFT.

The content of IP utterances correlated significantly with the content of crew utterances. Although correlation does not indicate causality, our subjective impression is that the IPs predominantly drove the content and emphasis of the debriefing. This impression is supported by the pattern of discourse, discussed above.

Duration of Debriefings

Most debriefings were fairly short, 31 minutes on average, including time spent watching videos (typically about 1/3 of the total session was spent watching video segments). Half-hour sessions allowed group discussion of only a few examples of crew performance, and often did not provide adequate time for in-depth analysis. Given all that occurs in a typical two-hour LOFT and the importance of deep analysis of what happened and how the crew managed the situations confronting them, it seems highly desirable to spend more than 31 minutes on debriefing. Although these data do not indicate what duration would be optimal, a thorough discussion was often accomplished in debriefings lasting about an hour. Obviously, airlines must consider cost-benefit

tradeoffs in the amount of time scheduled for debriefing. IPs do need to vary the length of the session according to the training needs of the crew, but the 10-fold range of duration observed in this study is clearly problematic.

Duration of debriefings correlated negatively with IPs' ratings of crew CRM and technical performance, suggesting that IPs spent somewhat more time with crews they perceived to have more problems. During interviews after debriefings, some IPs indicated that they found less to discuss with a crew that performed well and that they wanted to avoid criticizing good performance. We suspect this attitude may shortchange high performing crews, since analyzing why things went well could help them distinguish the factors and behaviors that led to success. Explicitly discussing what makes certain behaviors effective may help crews apply these behaviors more successfully on the line, even in difficult situations such as incompatibility with fellow crewmembers. Even high-performing crews need a chance to practice the skill of self-debriefing.

Debriefing Assessment Battery

The audiotapes were considerably less intelligible than the actual debriefings because the tapes lack some of the audio cues (e.g., 3D localization) and all of the visual cues that aid speech comprehension. To enable reliable assessment of IP or crew performance in a single pass through the audiotapes without pausing, we had the raters follow the tapes with the written transcripts (without preview or review). Although this arrangement obviously differs from using the battery in a field setting, the level of difficulty for the rater may, arguably, be roughly comparable. Guy Smith and Tony Sasso, who are conducting a field evaluation of the battery at Northwest Airlines, have been able to rate IP facilitation of live debriefings reliably with the battery (personal communication, 1998).

Intercorrelations among IP items and among crew items ranged from low to high. Two possibilities might account for these intercorrelations: (1) For some pairs of items the behaviors measured inherently overlap to some degree; for example, using questions is one of the ways IPs focus crews' attention on CRM. However, other pairs of items do not overlap; for example, it would be hard for raters to confuse explaining the debriefing process with use of questions. (2) In

this particular data set the independent variables measured by the battery items may covary as a function of some underlying variable. This might occur if IPs tended to either grasp and accept the fundamental concepts underlying facilitation or fail to grasp or accept those concepts. Crew behaviors might also have covaried as a function of how well they grasped their role and because of the influence of IP behaviors.

Use of Silence was the only IP item not significantly correlated with several crew items. Raters found this item more difficult to score than other items, and in seven of the 36 debriefings this item was not scored because the crews quickly filled in silences. Our subjective impression is that silence can be an effective facilitation tool, thus it may be worthwhile to revise this item to make it easier to score.

Implications for Training

The uniformly low IP scores on *Explains Debriefing Process* indicate that this is an area in which IPs have not been adequately trained. It seems a matter of common sense to tell crews explicitly what is expected of them. What the IP says in the introduction and how he or she says it sets the tone for the entire debriefing[9]. A good introduction is easy to provide; thus, training departments may be able to improve crew participation with relatively little effort by emphasizing this topic to IPs. Ideally, the introduction should describe how the debriefing will be conducted, explain how the crew is expected to participate and what the IP's role will be, and provide an explicit rationale for the benefits of crew-centered debriefings.

The wide variability of IP effectiveness in facilitation and emphasis on CRM and crew participation suggests that the airlines face an issue of standardization and quality control of debriefings. Although we did not collect data on instructor performance other than use of facilitation, the great majority of IPs were clearly highly competent technically, were conscientious, and displayed strong interpersonal skills. All seemed comfortable with and committed to the

---

[9] Sandy Lozito pointed out to us that this may be analogous to Ginnett's (1987) finding that what the CA says when first meeting the crew in line operations strongly influences team formation.

concepts of CRM. Thus, the variability may reflect inadequate training of IPs in the techniques of facilitation. Unfortunately, we have no data on the amount and content of training individual IPs received. When interviewed, several IPs spontaneously volunteered that they did not feel adequately trained to facilitate. Facilitation, especially because it departs radically from the instructional techniques traditionally used in aviation, requires hands-on training in which IPs observe expert facilitators, practice facilitating, and receive feedback[10].

IPs sometimes mistakenly assume that using facilitation requires relinquishing their role as teacher in the debriefing. This misunderstanding may arise because in some business settings, facilitators lack the deep subject matter expertise of the group members; in these settings the facilitator is primarily concerned with maintaining a process that will enable the group to work together effectively. In contrast, the LOFT IP has deep subject matter expertise in both technical and CRM matters; furthermore, from his or her position in the simulator cab the IP can see things happen in the LOFT that the crew are unable to observe. This expertise and these observations play an important role in the debriefing. IPs who are effective facilitators learn how to integrate their observations into group discussions in which the crewmembers are full participants.

These findings also suggest that the frequently expressed goal of having crews debrief themselves, using the IP as a resource, though worthwhile, is rather idealistic. Only one of the IPs observed attempted to have the crew lead their own debriefing. Though that debriefing was one of the better ones in terms of the level of crew participation, the crew only partially understood what constituted a good debriefing and needed considerable help

The importance of crews participating substantially in LOFT debriefings is widely recognized. Some authors have gone beyond this principle to argue also that the crews themselves should lead

---

[10] We have recently published a training manual that describes in detail specific facilitation techniques, how to integrate these techniques in the debriefing, and ways to avoid common facilitation mistakes (McDonnell, Jobe, and Dismukes, 1997). Also see the facilitation method developed by Klair, 1997.

the debriefings, using the IP as a resource (see, for example, Butler, 1993). Only one of the IPs in our study attempted to have the crew lead the debriefing. Although that debriefing was one of the better ones in the level of crew participation, the crew only partially understood what constituted a good debriefing and needed considerable help. Before crews can debrief themselves, they must be taught how to conduct a debriefing; this could be the subject of classroom training that precedes the LOFT. This classroom training might include videos of crews debriefing themselves effectively and provide crews opportunities to practice debriefing.

With the one exception noted above, the debriefings we observed revolved around the IP's questions and comments, even when the IP used facilitation techniques effectively to elicit crew participation. Is it better for the IP or the crew to lead the discussion? No data exist to answer this question. Also we are not aware of any model that explicitly describes how a crew should conduct a debriefing. Leading the discussion provides the IP greater moment to moment control of content, and still permits using facilitation to elicit substantial levels of crew participation. On the other hand, crews may learn more from questions they pose themselves, and taking more responsibility for the LOFT debriefing may encourage crews to debrief themselves on the line. This issue deserves further exploration.

Realistically, many crews may not yet have the skills and motivation to lead their own debriefings without substantial assistance from the IP, however this situation might change over time if IPs consistently encourage crews to take a proactive role in debriefing their own LOFT and to consider the benefits of debriefing after line operations.

At the current state of industry practice, IPs who attempt to encourage crews to self-debrief, or to at least take greater responsibility for the direction of the debriefing, will encounter widely varying levels of crew responsiveness. McDonnell, Jobe, and Dismukes (1997), drawing upon a concept expressed by Continental Airlines (1992), suggest that facilitation can be conducted at levels ranging from low to high, depending on the level of initiative and the self-debriefing skill of the particular crew. In the highest level of facilitation the IP assists the crew in their own analysis, blending his or her own observations into the group discussion. In low-level facilitation the IP leads

the debriefing, directs the crew's attention to critical issues, and may need to lecture to insure points are understood, but the IP still attempts to foster as much self-analysis as possible.

It has been a matter of faith among training departments that facilitation is an effective tool to encourage crews to analyze their performance in LOFT along CRM dimensions in a way that will benefit them in line operations. This study provides empirical evidence that this faith is correct.

Conclusions and Recommendations

The DAB can be used to assess IP facilitation and crew participation reliably and has potential for use in field settings by airline personnel. This is a substantial advantage over measuring descriptive variables, which requires a tedious amount of data reduction from audiotapes in the laboratory.

The following conclusions and recommendations address the specific questions posed at the beginning of this paper and reflect both the objective data and our subjective impressions:

1) Most instructors attempted to facilitate crew participation, but their success ranged from very good to poor.

2) IPs used questions as their primary tool to elicit crew participation and guide the content of discussion. IPs used other facilitation techniques less frequently but significantly influenced crew participation when they did. The content of crew discussion was strongly driven by the content on which the IPs focused.

3) Within each of the five airlines, instructors varied widely in their conduct of debriefings, especially in terms of emphasis on CRM, emphasis on crew participation, and effectiveness in facilitation. This suggests a need for better standardization within companies. The great variability within individual airlines made it difficult to determine whether real differences occur among airlines.

4) Crews did not lead their own debriefings, but did participate actively. The concept of having crews debrief themselves using the IP as a resource may be unrealistic at present, deserves further explanation. The 31-minute average for debriefings in this study limited the depth and range of analysis of the LOFT. Crews would probably learn more from the LOFT with a longer debriefing;

however this issue involves cost-benefit tradeoffs that can only be evaluated by the airlines themselves.

5. This study provides empirical evidence that facilitation works. IPs who facilitated effectively elicited more participation and deeper analysis by crewmembers than did a substantial number of IPs who did not use facilitation effectively. Airlines would benefit by giving IPs explicit, hands-on training in facilitation, followed up with mentoring.

Acknowledgments

This study originated from requests from several airline training departments for help in analyzing the effectiveness of LOFT debriefings. Doug Daniel, Mary Connors, Mike Shafto and Steve Gregorich helped identify crucial issues and ways to study these issues.

This study could not have been conducted without the generous willingness of instructors and line crews to allow us to observe their debriefings. We are impressed with their high standards of professionalism. Training department managers from each of the airlines that participated in the study provided a wealth of background information and made valuable suggestions.

The FAA's Office of the Chief Scientist and Technical Advisor for Human Factors (AAR-100) funded this study. Eleana Edens, the program manager, provided support, encouragement and helpful suggestions. The NASA/San Jose State University Foundation Cooperative Research Agreement NCC 2-798 also provided support.

References

Butler, R. E. (1993). LOFT: Full-mission simulation as crew resource management training. In E. L. Wiener, R. L. Helmreich, & B. G. Kanki (Eds.), Cockpit resource management (pp. 231-259). San Diego: Academic Press.

Continental Airlines, Flight Operations Human Factors Group. (1992, January). LOFT facilitation techniques: Practical strategies and techniques for LOFT facilitators. Unpublished manuscript.

Dismukes, R. K., Jobe, K. K., & McDonnell, L. K. (1997). LOFT Debriefings: An analysis of instructor techniques and crew participation. (NASA Technical Memorandum 110442). Moffett Field, CA: NASA Ames Research Center.

Duval, S., & Wicklund, R. A. (1972). A theory of objective self awareness. New York: Academic Press.

Federal Aviation Administration (1990). Line operational simulations (Advisory Circular 120-35B). Washington DC: Author.

Federal Aviation Administration (1995). Crew resource management training (Advisory Circular 120-51B). Washington DC: Author

Ginnett, R. C. (1987). The formation process of airline flight crews. In R. S. Jensen (Ed.), Proceedings of the Fourth International Symposium on Aviation Psychology (pp. 399-405). Columbus, OH: Ohio State University.

Gow, L., & Kember, D. (1993). Conceptions of teaching and their relationship to student learning. British Journal of Educational Psychology, 63, 20-33.

Hawkins, F. H. (1987). Human Factors in flight. Brookfield, VT: Gower Publishing Company.

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, R. L. Helmreich, & B. G. Kanki (Eds.), Cockpit resource management (pp. 3-41). San Diego: Academic Press.

Jones, P. G. (1982). Adult learning in your classroom: The best of training magazine's strategies and techniques for managers and trainers. Minneapolis, MN: Lakewood Books.

Klair, M. (1997). Briefing and debriefing techniques for instructors in aviation. In R. S. Jensen & L. A. Rakovan (Eds.), Proceedings of the Ninth International Symposium on Aviation Psychology (pp. 568-573). Columbus, OH: Ohio State University.

McDonnell, L. K., Jobe, K. K., & Dismukes, R. K. (1997). Facilitating LOS debriefings: A training manual. (NASA Technical Memorandum 112192). Moffett Field, CA: NASA Ames Research Center.

Smith, G. M. (1994). Evaluating self-analysis as a strategy for learning crew resource management (CRM) in undergraduate flight training. Ann Arbor, MI: Dissertation Abstracts.

Appendix

Debriefing Assessment Battery - Short Form

Directions:

Use the scale below to rate the IPs and crews on each of the following elements. If a particular item does not apply to the IP being rated (e.g., no need to pause or rephrase questions because crew always answers right away), skip that item and divide the total IP score by the actual number of items scored.

Needs

Poor	**Marginal**	**Improvement**	**Adequate**	**Good**	**Very Good**	**Outstanding**
1	**2**	**3**	**4**	**5**	**6**	**7**

IP Facilitation:

IP Facilitation is a summary of the strategies and techniques IPs use to guide the debriefing session. The two main goals of the debriefing session are to 1) get the crew to perform an in-depth analysis of the situation that confronted them, how they understood and managed the situation, the outcome, and ways to improve, and 2) get the crew to participate in a proactive, rather than reactive, manner in which they initiate discussion and elaborate beyond the minimal.

_____ *Clarifies* up front that the *goal* is for the crew members to give their own in-depth analysis and evaluation, initiating their own discussion rather than merely responding to the IP, and *clarifies* that the IP's *role* is to facilitate their discussion. Explicitly states the *rationale* for this approach.

_____ Encourages crew to *analyze* in depth, including the situation that confronted them, *what* they did to manage the situation, and *why* they did what they did

_____ Encourages crew to *evaluate* their performance and, if appropriate, discuss ways to improve

_____ Encourages crew to explore *CRM* issues, how CRM affected their LOS performance, and/or how CRM affects line operations

_____ Keeps *focus on crew* rather than IP by using techniques such as active listening*, following up on crew-initiated topics, and encouraging crew to address each other

_____ Asks *questions* (as necessary) to introduce issues, elicit crew participation, and evoke deep analysis

_____ Allows at least 3 seconds of *silence* after IP questions and crew comments to allow crew to formulate thoughtful responses and encourage continued crew discussion and in-depth analysis. If necessary, rephrases questions to avoid answering for crew.

_____ *Refrains from lecturing* to crew or analyzing or evaluating crew performance without attempting to elicit substantial analysis and evaluation by the crew (unless absolutely necessary because crew will not participate or cannot grasp an essential concept)

_____ **Total score**

_____ **Final IP score** (Total score ÷ number of items scored)

* Can range from simple acknowledgment such as "uh huh" to reflective listening in which IP repeats or rephrases what crew member said to encourage crew to elaborate

Crew Participation:

Crew participation measures the degree to which 1) the crew performs an in-depth CRM analysis and evaluation of the situation that confronted them, how they understood and managed the situation, the outcome, and ways to improve, and 2) the crew participates in a proactive rather than reactive manner, initiating discussion, interacting with each other, and elaborating beyond the minimal. Note that crews should be rated on how well they actually performed on these dimensions, even if poor crew participation was due to the IP hindering their participation.

_____ Perform in depth *analysis* of the situation that confronted them, *what* they did to manage the situation, and *why* they did what they did

_____ *Evaluate* their performance as a crew and, if appropriate, ways they might improve

_____ Explore *CRM* issues and how they affect LOS outcomes and/or line operations

_____ Behave in a predominantly *proactive* rather than reactive manner, initiating dialogue rather than just responding to IP questions

_____ Go beyond minimal responses to IP questions

_____ *Interact* with each other rather than only with the IP

_____ **Total score**

_____ **Final Crew score** (Total score ÷ number of items scored)

TABLE 1

Number of Debriefings Observed and Analyzed

	Airline V	Airline W	Airline X	Airline Y	Airline Z	Total
2-person	6	0	5	5	9	25
3-person	2	4	0	4	1	11

TABLE 2

Descriptive Measures

LOFT Ratings	
IP Rating of Crew's CRM Performance in LOFT (1 to 5 scale)	
IP Rating of Crew's Technical Performance in LOFT (1 to 5 scale)	
Participation Measures	
Duration of Session	
Percentage of Participation (CA, FO, FE, IP)	
Percentage of Crew Utterances Coded as Questions, Responses, and Statements	
Percentage of Crew Words Coded as Questions, Responses, and Statements	
Percentage of Crew Questions Coded as Proactive	
Number of IP Questions Per Hour	
Percentage of IP Questions directed to specific crew members	
Discussion of Crew Performance	**Content of Discussion**
% Words Positive (Crew, IP)	% Words CRM (Crew, IP)
% Words Improve (Crew, IP)	% Words Technical (Crew, IP)
% Words Negative (Crew, IP)	% Words Mixed (Crew, IP)
% Words Neutral (Crew, IP)	% Words Non-specific (Crew, IP)
% Words Performance (Crew, IP)	

TABLE 3

Participation in Debriefing

(percent of all group words uttered by each participant)

	IP	CA	FO	FE
2 Person Crews	61 (15) [a]	21 (7.8)	18 (9.7)	_
3 person Crews	49 (20) [b]	20 (9.4)	19 (13)	13 (7.8)

[a]Instructor> captain, first officer ($p<.001$); [b] Instructor> captain, first officer, flight engineer ($p<.001$).

TABLE 4

Content of Debriefings (percent of words)

Mean (SD)

	Airline V (n=8)	Airline W (n=4)	Airline X (n=5)	Airline Y (n=9)	Airline Z (n=10)	Combined (n=36)
Instructor						
CRM	32 (25)	19 (15)	27 (13)	56 (13)	64 (17)	45 (24)[a]
Technical	22 (14)	13 (11)	38 (10)	8.1 (8.7)	10 (15)	16 (15)[b]
Mixed	24 (8.6)	33 (13)	9.8 (16)	5.6 (5.3)	6.2 (8.3)	14 (14)[c]
Non-specific	22 (11)	34 (12)	26 (7.6)	30 (6.8)	20 (10)	25 (10)
CRM + Mixed	56 (20)	52 (8.0)	36 (9.6)	62 (12)	70 (17)	58 (18)[d]
Crew						
CRM	25 (12)	25 (17)	36 (20)	68 (13)	68 (19)	49 (25)[e]
Technical	21 (11)	10 (4.2)	23 (8.6)	5.6 (5.3)	6.9 (10)	12 (11)[f]
Mixed	38 (13)	46 (12)	8.8 (10)	11 (10)	14 (12)	21 (18)[g]
Non-specific	16 (11)	18 (4.6)	32 (14)	16 (7.4)	12 (13)	17 (12)[h]
CRM + Mixed	63 (15)	71 (6.8)	45 (20)	79 (11)	82 (21)	71 (20)[i]

Note. Differences found among airlines were determined by one-way ANOVA, Bonferroni post-hoc tests: [a]Y>W; Z>V,W,X. [b]X>Y,Z. [c]V>Y,Z; W>X,Y,Z. [d] Z>X. [e] Y>V,W,X; Z>V,W,X. [f]V>Y,Z; X>Y,Z. [g]V>X,Y,Z; W>X,Y,Z. [h] X>Z. [i] Y>X; Z>X.

TABLE 5

Discussion of Crew Performance

	Mean (SD)	
	% of IP words	% of crew words
Positive aspects	18.0 (12.0)	8.0 (9.6)
Negative aspects	3.2 (5.5)	5.9 (6.7)
Ways to improve	4.1 (4.6)	4.8 (6.1)
Neutral description	17.0 (9.5)	33.0 (19.0)
Performance total	41.0 (15.0)	52.0 (21.0)

TABLE 6

Instructor DAB Scores

Mean (SD)

Item	Crew size	Airline V	Airline W	Airline X	Airline Y	Airline Z	Combined
Explain debriefing	*2*	1.8 (.45)		1.3 (.58)	3.5 (1.6)	1.8 (.83)	2.1 (1.3) [a]
process	*3*	NA [b]	1.8 (1.5)		3.3 (1.5)		2.4 (1.6)
Encourage crew	*2*	2.8 (2.2)		3.2 (1.6)	3.2 (1.8)	2.4 (1.7)	2.8 (1.7)
analysis	*3*	5.5 (.70)	3.0 (1.8)		4.5 (1.3)		4.1 (1.7)
Encourage crew	*2*	2.5 (1.7)		3.0 (1.6)	2.8 (1.3)	2.4 (1.2)	2.6 (1.4)
evaluation	*3*	5.5 (.71)	3.5 (1.3)		5.0 (.82)		4.5 (1.3)
Focus on	*2*	2.5 (1.5)		3.4 (1.3)	4.4 (2.1)	4.0 (2.3)	3.6 (1.9)
CRM	*3*	4.0 (1.4)	3.5 (1.3)		5.3 (.96)		4.3 (1.3)
Focus on crew	*2*	3.2 (2.5)		3.0 (1.9)	3.6 (1.8)	2.8 (1.5)	3.1 (1.8)
participation	*3*	3.0 (0.0)	3.5 (2.4)		5.5 (1.3)		4.2 (1.9)
Use of	*2*	3.0 (1.9)		3.0 (1.9)	3.8 (1.6)	3.3 (1.4)	3.3 (1.6)
questions	*3*	4.0 (1.4)	3.3 (1.5)		5.8 (.50)		4.4 (2.2) [c]
Use of	*2*	2.5 (1.9)		4.5 (1.3)	4.6 (1.3)	4.3 (1.5)	4.0 (1.6)
silence	*3*	3.0 (0.0)	4.0 (4.2)		4.0 (1.4)		3.9 (2.0)
Refrain from	*2*	3.0 (2.4)		2.8 (1.9)	3.6 (1.5)	2.8 (1.5)	3.0 (1.8)
lecturing	*3*	4.5 (.71)	3.0 (1.4)		4.8 (.56)		4.0 (1.2)
Overall	*2*	2.6 (1.7)		3.1 (1.5)	3.7 (1.0)	2.9 (1.2)	3.0 (1.3)
score	*3*	4.3 (.14)	3.1 (1.5)		4.8 (.82)		4.0 (1.3)

[a] Y>Z, $p < .05$; [b] Not available due to missing data; [c] Y>W (determined by one-way ANOVA, Bonferroni post-hoc tests).

TABLE 7

Crew DAB Scores

Mean (SD)

Item	Crew size	Airline V	Airline W	Airline X	Airline Y	Airline Z	Combined
Analyze	2	3.3 (1.8)		3.0 (1.6)	3.0 (2.1)	3.4 (1.9)	3.2 (1.7)
in depth	3	4.5 (2.1)	2.5 (1.9)		5.8 (.96)		4.2 (2.1)
Evaluate	2	3.2 (1.2)		3.2 (1.3)	3.0 (1.9)	2.6 (1.3)	2.9 (14)
performance	3	4.5 (.71)	3.5 (.58)		5.5 (.58)		4.5 (1.1) [a]
Focus on	2	3.0 (1.3)		3.6 (1.8)	4.4 (1.8)	3.7 (1.4)	3.6 (1.5)
CRM	3	5.5 (.71)	4.3 (1.5)		6.3 (.96)		5.3 (2.2)
Proactive	2	3.7 (2.2)		3.0 (1.6)	3.4 (1.8)	3.1 (1.8)	3.3 (1.8)
participation	3	4.0 (1.4)	4.3 (1.5)		6.0 (.82)		4.9 (1.4)
Respond in							
depth	2	2.5 (.83)		3.0 (1.6)	4.4 (1.1)	3.6 (1.7)	3.4 (1.5)
to questions	3	3.0 (0.0)	2.8 (.96)		5.3 (1.3)		3.8 (1.5) [b]
Interact with	2	1.5 (1.2)		2.4 (1.3)	2.8 (1.8)	2.2 (1.7)	2.2 (1.5)
each other	3	3.0 (0.0)	3.0 (2.3)		4.8 (.96)		3.7 (1.7)
Overall	2	2.9 (.97)		3.1 (1.4)	3.5 (1.6)	3.1 (1.2)	3.1 (1.2)
score	3	4.1 (.35)	3.4 (1.0)		5.5 (.61)		4.4 (1.2) [c]

[a]Y>W, $p < .01$; [b]Y>W, $p < .05$. [c]Y>W, $p < .05$ (determined by one-way ANOVA, Bonferroni post-hoc tests).