Facilitated Debriefing of Line Operational Simulation training events

Peter Bergh | LUND UNIVERSITY

Facilitated Debriefing of
Line Operational Simulation
training events

Peter Bergh

Lund 2022

Title: Facilitated Debriefing of Line Operational Simulation training events

Author: Peter Bergh

Number of pages: 42

ABSTRACT

This study analyses aspects of instructor behaviour and their possible effects on the ability of the participating crew to analyse their own performance following simulator training events for Flight Crews. Eight briefings and debriefings related to one airline's training sessions were observed. The debriefings were recorded and transcribed. The analysis was performed using the "Debriefing Assessment Battery", a subjective rating scale developed by McDonnell (1996) in order to systematically assess the instructor's facilitation skills and the behaviour of the participating crew during debriefing. Additionally, quantitative data was collected using transcription of audio recordings from the observed debriefings. Further, a questionnaire was used to collect information about the participants' opinions on the debriefing process and outcome. Due to the small sample size of this study, the analysis results should be treated cautiously. However, the results of this study indicate that instructor behaviour and level of facilitation skills are predictive of the resulting level of crew participation, self-analysis and self-reflection. These results suggest that development of instructor facilitation skills is likely to improve the learning value of simulator training events. Based on the observations, a few suggestions for improvement of facilitator techniques are provided.

Keywords: Facilitation, Facilitated debriefing, Line Operational Simulation, experiential learning, self-reflection, debriefing, simulator training

Acknowledgements

First, I would like to express my gratitude to my supervisor, Mark Milich, for your valuable advice, helpful suggestions and constructive feedback.

Also, my sincere thanks to the Training Management of the participating airline for allowing me to conduct this study. And in particular, a huge thank you to the "coordinator". Without your efforts, support, and enthusiasm, this study would not have been possible. And, many thanks to all the highly professional crews and instructors who allowed me to observe your work. Again, without you there would have been no study and no thesis.

To my partner, Pernilla Ekberg, much love and many thanks for your support, advice, clever ideas, and patience.

I also wish to show my appreciation to David Bergh for excellent IT support when it was needed the most.

Finally, I would like to express my gratitude to Capt. Colin Budenberg for introducing me to the subject of Facilitation during the Thomsonfly Training Pilot CRMI Accreditation Course all those years ago.

Table Of Contents

What is "Facilitation"? Why is "facilitative debriefing" or "facilitated debriefing" gaining ground as the preferred debriefing method in connection with flight simulator training (FAA, 2015 p. 15; EASA, 2019 p. 29) as well as clinical simulations in medical training (Gantt et al., 2018)? According to Kolb (1984), *reflection* (or "reflective observation") forms part of the "Learning Cycle". Experiential learning (learning from experience) is achieved when the "learning cycle" is completed. The cycle consists of four steps: active experimentation, concrete experience, reflective observation and abstract conceptualisation (see Figure 1 below). Several studies and publications support Kolb's theory and confirm the importance of reflection in connection with learning from experiences such as training events. (Wain, 2017; Boud et al., 1985). "A skilled facilitator, asking the right questions and guiding reflective conversation before, during and after an experience, can help open a gateway to powerful new thinking and learning" (Jacobson and Ruddy, 2004, p.2).

CONCRETE EXPERIENCE :

Doing or experiencing something

REFLECTIVE OBSERVATION :

Reviewing and reflecting on the experience

ABSTRACT CONCEPTUALISATION :

Concluding / Learning from the experience

ACTIVE EXPERIMENTATION :

Planning / Trying out what you have learned
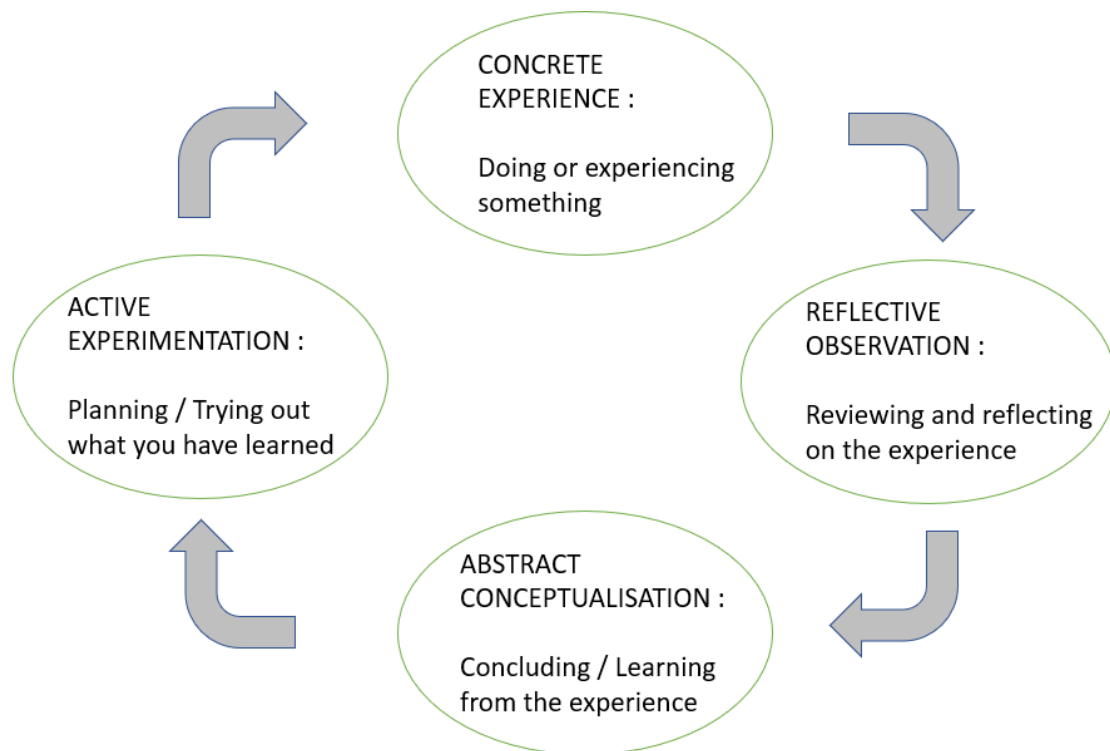
Figure 1. Illustration of The Learning Cycle (Kolb, 1984)

Initial and recurrent Flight Crew Training is performed in accordance with the regulations provided by, for instance, European Union Aviation Safety Agency (EASA) or Federal Aviation Administration (FAA) depending on the certification of the Air Operator (EASA, 2011). Part of this training consists of "Line-Operational Simulations" (LOS), such as Line-Oriented Flight Training (LOFT), Special Purpose Operational Training (SPOT) or Line Operational Evaluation (LOE). During LOS training the flight crew are exposed to "real-time scenarios that address normal, non-normal, abnormal, or emergency procedures and provides training in CRM" (FAA, 2015, p. 3). Being an example of experiential learning, the learning value of LOS training events should, in alignment with Kolb's (1984) theory, be

greatly enhanced by a successful debriefing that elicits self-reflection. Or, in the words of Helmreich and Foushee (2010, p. 28): "The debriefing of LOFT is also a critical element in achieving impact. Skilled instructors should guide crews to self-realization rather than lecture them on observed deficiencies". This view is echoed by Lederman (1992, p. 158): "The process of debriefing is not ancillary to the educational experience to which it is tied. Debriefing is an integral part of any learning experience that is designed to be experience based".

So, what are the characteristics of a successful debriefing? How can the resulting learning value be maximised? Research suggests that "adults learn and remember more when they participate actively and make their own analyses rather than when they listen passively to someone else's" (McDonnell et al., 1997, p. 2). This suggests that the debriefing should be "crew-centred" rather than "instructor-centred". According to Rogers (1969) on the subject of "facilitation of learning", the advantages are many:

- Active participation enhances the learning experience.
- Self-initiated learning that involves feelings as well as intellect is more lasting and pervasive.
- When self-evaluation is primary and evaluation by others is secondary, self-reliance will be facilitated. The resulting evaluation of our own behaviours will enable us to learn from our mistakes and successes.

This view is supported by several studies within clinical simulation in connection with medical training (Gantt et al., 2018; Wain, 2017). However, a literature search reveals few published Facilitation studies regarding Flight Crew Training. Three examples are: "Facilitation Techniques as Predictors of Crew Participation in LOFT Debriefings" (McDonnell, 1996), "LOFT Debriefings: An Analysis of Instructor Techniques and Crew Participation" by Dismukes, Jobe and McDonnell (1997) and "Improving Facilitated Debriefings - How are barriers to learning recognized by instructors and mitigated during post-simulator debriefings?" (Pruchnicki, 2018).

It is likely that several airlines throughout the world have developed their own briefing and debriefing concepts, including the use of facilitation techniques. Others have not yet implemented the concept of facilitated debriefing. Therefore, it is probable that many instructors worldwide still have had little opportunity to learn and develop facilitation skills. These skills are increasingly required by regulations. According to the International Civil Aviation Organization (ICAO), facilitation is "the primary [debriefing] technique that should be used for competency-based training" (2013, p. I-7-1). Additionally, improving the quality of debriefings may prove to be a highly cost-effective way of gaining additional training value from simulated training events.

Research questions

The purpose of this study is to evaluate possible effects of facilitated debriefing on the behaviour of the participants in connection with pilot training. It is partly replicating Dismukes et al. (1997). The study will attempt to answer the following questions: *How does the Instructor Pilot's (IP) behaviour and use of facilitation techniques affect the quality and outcome of the debriefing? What measurable and quantifiable aspects of IP behaviour are positively or negatively correlated with desired aspects of Crew debriefing behaviour such as high levels of Crew participation, interaction, analysis, and self-reflection?* In order to find answers to the above questions, five hypotheses were formulated:

Hypothesis 1 (H1):   *The more the IP talks, the less active the crew will be, which will in turn have a negative effect on crew interaction, self-reflection, and analysis.*

　　　To summarise the research cited above concerning experiential learning in general, and LOS training in particular, the aim of facilitated debriefing is to facilitate learning by creating a debriefing environment where the crew can proactively identify topics for discussion, self-reflect upon the events that occurred and why they occurred, and critically analyse their own performance in order to create valuable "lessons learned" that can be applied during line flying. For this to happen, it is reasonable to assume that the crew needs to feel their views and opinions are important. To demonstrate interest in the views and opinions of the crew, the instructor should allow them to speak without interruption, and listen actively to what they have to say. According to McNaughton et al. (2008, p. 224), "the goal in active listening is to develop a clear understanding of the speaker's concern and also to clearly communicate the listener's interest in the speaker's message". It is therefore likely that an instructor who talks excessively is discouraging crew activity, interaction, and analysis by implicitly demonstrating that the opinions of the instructor are more important than those of the crew. Active listening increases the feeling of being understood (Weger et al.,2014, p. 14). Therefore, allowing the crew to speak more may be beneficial for interaction, sharing of views and in-depth self-analysis.

Hypothesis 2 (H2):   *IP facilitation skills are conducive to crew self-reflection and in-depth self-analysis of crew performance*

　　　Similarly to Hypothesis 1, this hypothesis rests on the assumption that higher levels of facilitation skills are beneficial for the debriefing outcome. This is supported by previous studies (McDonnell, 1996; Dismukes et al., 1997). In this case, H2 asserts that the overall performance of the instructor during the debriefing will affect the behaviour of the crew and the quality of the outcome. "IP facilitation skills" covers a wide range of instructor behaviour in several different aspects of facilitation. The Debriefing Assessment Battery (DAB) developed by McDonnell (1996) and used by Dismukes et al. (1997) categorise these aspects as: Introduction, Questions, Encouragement and Focus on Crew Analysis and Evaluation (Appendix 1). While it may be difficult to correctly measure the relative impact of each skill, it is likely that a more proficient facilitator will be more successful in bringing the crew to self-reflect and analyse in-depth, using for instance a good introduction, questioning techniques, the use of silence, active listening and encouragement.

Hypothesis 3 (H3):   *IP use of Probing/Follow-up questions results in more and better crew analysis*

　　　This hypothesis suggests that the instructor can influence the amount and quality of analysis done by the crew by asking probing questions and follow-up questions. According to Legard, Keegan and Ward (Richie & Lewis, 2003, p.141) regarding in-depth interviewing: "an initial response is often at fairly 'surface-level': the interviewer will use follow-up questions to obtain a deeper and fuller understanding of the participant's meaning". For instance, following a crew statement such as "The CRM was good", a probing question could be "What is 'good CRM', in your opinion?". An example of a way to use a follow-up question after a crew statement such as "I felt rushed during the non-precision approach" is: "I understand. What caused you to feel rushed, do you think?". While the quality of the crew analysis may depend on many factors such as crew ability, motivation, and event complexity, it is not unlikely that proper use of probing and follow-up questions will benefit the amount and quality of crew analysis.

Hypothesis 4 (H4):   A good introduction to facilitated debriefing will increase Crew Interaction and Crew Analysis

According to the DAB (McDonnell, 1996), a good introduction to a facilitated debriefing will explain the reason why facilitated debriefings are more effective than "lecturing". It should outline the respective roles of crew members and facilitator. It will clarify that the crew are expected to do most of the talking and proactively bring up subjects for discussion. The instructor should also explain that the crew should "dig deep", go beyond just "what happened", and critically analyse their performance. A well performed introduction should therefore have a positive effect on the amount of crew interaction and analysis.

Hypothesis 5 (H5):   IP encouragement will result in more crew activity and interaction

Instructor pilot encouragement is described as "the degree to which the IP encourages and enables the crew to participate actively and deeply" (McDonnell, 1996, p. 11). This is accomplished by, for example, active listening, conveying a sense of interest in the crew's views, avoiding interruptions, and avoiding presentation of own analysis before the crew has given their view. If the crew members feel that their opinions are important and valued, it is likely that they will participate more actively.

Methods

General

This study aimed to identify some of the key factors that contribute to a successful facilitated debriefing, and how these are influenced by Instructor Pilot behaviour. First, a literature search was conducted using the Lund University research database, Google Search and Google Scholar. The following search words were used individually and in combination: "facilitation", "facilitated debriefing", "facilitative debriefing", "debriefing", "simulation", "flight crew training", "aviation", "LOFT", "LOS", "LOE". Plenty of relevant studies of facilitation and debriefing within the area of Medical Simulation are available. However, very few published aviation-related empirical studies concerning facilitation techniques and their effects on crew behaviour during debriefing were found. These were: "Facilitation Techniques as Predictors of Crew Participation in LOFT Debriefings" (McDonnell, 1996) and "LOFT Debriefings: An Analysis of Instructor Techniques and Crew Participation" by Dismukes, Jobe and McDonnell (1997). Based on the data obtained, "Facilitating LOS Debriefings: A Training Manual" (1997) by McDonnell, Jobe and Dismukes was issued as NASA Technical Memorandum 112192. A few years later, "Facilitation and Debriefing in Aviation Training and Operations" (2000) by Dismukes and Smith was published.

This study is partly replicating Dismukes et al. (1997), albeit using a smaller sample size. Similarly to the 1997 study, flight crews and Instructor Pilots were observed and audio recorded during post-simulation debriefings. Each debriefing was transcribed and analysed using a coding system to quantify certain aspects of the debriefings. The participants' behaviour during debriefing was also assessed using a standard system of behavioural markers- the DAB (McDonnell, 1996). The assessment data was compared to the analysis data to test the validity of several hypotheses regarding the effects of observable Instructor Pilot behaviour on the crews' ability to successfully analyse their performance and draw conclusions by self-reflection.

For the purpose of determining how "successful" a debriefing is, there needs to be a definition of desired aspects of crew behaviour with regards to debriefing. This has been accomplished through the definition of several quantifiable variables relating to the participants' communication and behaviour during the debriefing. While the relevance and suitability of these variables can (and should) be a subject for further study, they have in this study been replicated from the Dismukes et al. (1997) study with a few modifications. Where Dismukes et al. used *word count* as a means of quantifying the recorded debriefing data, this study has used *time*. For example, *Participation* in the Dismukes et al. study indicates the participant's number of words spoken as a percentage of the sum of all participants' total number of words spoken. In this study, *Participation* instead indicates the "talk time" of each participant as a percentage of the total "talk time" of all participants. The reason for the choice of *time* over *word count* for this study is that counting words is significantly more time-consuming than measuring the length (in seconds) of each utterance. This study has been performed over the course of about two months, by one person. Quantifying observed behaviour certainly has its challenges. In that perspective, it is possible to argue the merits of both word count and duration of utterance as quantitative measures.

Regardless of this difference, the desired crew behaviour and resulting outcome remains identical. The purpose of the debriefing is *learning*. Learning from the events that took place during the simulator training session. As explained above, "self-reflection" is considered a more effective way of learning from experience compared to "lecturing" (Helmreich & Foushee, 2010, p. 28). In the same manner, the "student-centred" debriefing (a high level of crew participation) is preferable to the "instructor-centred", lecturing way of debriefing. Therefore, factors that appear crew-centred and conducive to (or indicative of) self-reflection are in this study viewed as "desirable". These are: Crew Analysis and Evaluation (assessed score), Depth of Crew Activity (assessed score), Crew Participation (%), Crew S1 statements (self-initiated, substantive crew statements that raise issues, introduce topics, or add information to an existing topic), Crew Proactive Questions (not reactive or miscellaneous), Crew Analysis (number of utterances and time) and Crew Sequential utterances (indicating interaction between crew members rather than between Instructor Pilot and crew member).

This study doesn't cover the "Use of videos" as none of the observed debriefings included video recordings. Another difference compared to Dismukes et al. (1997) is that this study is focused on one single airline where facilitation has recently been introduced. Therefore, any findings may also indicate suitable topics for instructor support and development during the early stages of implementation.

To gain additional insight regarding the respective perceptions of the debriefing, a questionnaire was conducted with all participants immediately after each debriefing.

Participants

For the purpose of this study, the Training Department of a passenger airline was contacted with a request for observation of LOFT/LOE briefings and debriefings. The airline is well established, and the experience level of the pilots and instructors is very high- in most cases more than 20 years of continuous service. The airline has recently introduced facilitation as a debriefing tool. The concept has been presented at Instructor Meetings, and a printed folder has been distributed. No actual facilitation training of the instructors has been performed yet. Some of the instructors have (at their own initiative) incorporated facilitation techniques into their debriefings but for the most part, chronological and instructor-centred debriefings have been the norm until now.

Ethical considerations

The participating airline was informed about the planned purpose, scope and conduct of this study, which was approved by the Head of Training. Following this approval, all instructors were informed of the study. In advance of each planned session, all participants were contacted by the coordinator to inform them about the observations and ask for their consent. They were also informed (in writing and verbally) that their participation was voluntary, and that no identifiable data would be stored, used, or distributed. The audio recordings would be used for transcription and assessment purposes only, and would be deleted following the completion of the study. The participants were informed that they had the option of withdrawing their consent at any time, for any reason, without consequences.

Procedures

The participating airline assigned one of their instructors as a coordinator. The coordinator selected suitable sessions for observation, based on session content and observer availability. During each session, one Instructor Pilot (IP) participated along with one Captain (CA) and one First Officer (FO). The abbreviations "IP", "CA" and "FO" are used throughout this study in alignment with the terminology used by Dismukes et al. (1997). The simulator sessions were either Licence Proficiency Check/ Operator Proficiency Check (LPC/OPC) sessions (including LOFT/LOE) or Type Rating Training sessions with elements of LOFT/LOE (EASA, 2012). A typical simulator session started with a briefing which was performed in a designated briefing room adjacent to the actual Full-Flight Simulator (FFS). The duration of the briefing was 45-120 minutes, depending on whether the FFS session duration was 4hr or 2+2 hr. After the briefing, the simulator session commenced. The duration of the FFS sessions was either 4hr (with a 10-minute break after about two hours) or divided into two 2-hr sessions with a 2hr break in between. The debriefings took place directly following the 4hr sessions, and directly following each of the 2hr half-sessions. All 8 observations were performed by the author of this study. The pre-session briefings were only observed, and notes were taken to determine how the introduction to the facilitated debriefing was performed. The debriefings were observed, and audio recordings were made. After each session, all participants completed a questionnaire regarding their perception of the debriefing. The audio recordings from each debriefing were transcribed and coded using a similar format (Appendix 3) to the one used by Dismukes et al. (1997). All references to individuals or airlines were removed. Additionally, each debriefing was assessed and graded by the author/observer, using the Debriefing Assessment Battery [DAB] (Dismukes et al., 1997).

The questionnaires (Appendix 2) were completed anonymously by each crew member and the IP. The questionnaires for the crew and the IP were different. The participants were asked to review five statements, and the IP - eight statements. The statements were related to their recently completed debriefing. All participants were asked to select the most suitable answer option for each statement. The statements for the crew were:

I understand why facilitation is used as a debriefing tool.
The instructor gave us a sufficient introduction and explained how the debriefing should be conducted
My partner and I did most of the talking
Our analysis went deeper than simply "what happened". We explored the related CRM issues and how they affected our performance

After the debriefing was over, we had identified why things went well (or not so well) and what we might do differently next time we are in a similar situation

For the IP, the statements were:

I understand why facilitation is used as a debriefing tool
I am confident in my abilities as a facilitator
I gave the crew a sufficient introduction and explained how the debriefing should be conducted
The crew did most of the talking
The crew's analysis went deeper than simply "what happened". They explored the related CRM issues and how their performance was affected
After the debriefing was over, we had identified why things went well (or not so well) and what the crew might do differently next time they are in a similar situation
I would like to get more training/coaching as a facilitator
Having a basic guide or structure to base my facilitated debriefings on would be useful for me

The answer options for crew and IP were adapted from a Likert (1932) scale: "Strongly disagree", "Slightly disagree", "Slightly agree", "Strongly agree". As the interval between each option is unlikely to be perfectly linear, the modal value (most frequently occurring value) for the replies to each question was calculated, in accordance with the guidance by Jamieson (2004) concerning Likert scales.

Measures

Descriptive measures

Each of the 8 recorded debriefings was transcribed verbatim and coded for the following 10 factors: Utterance length (seconds), Speaker, Utterance (content word by word), Interruption/Interjection, Utterance type, Crew Proactive question, Topic Type, Analysis, Probing/Follow-up question and Interaction CA/FO. See Appendix 3 for a detailed description of the coding rules. In alignment with Dismukes et al. (1997), this data will be referred to as "descriptive". Based on the transcriptions and coding, the following variables were created:

Table 1.
Variables and definitions for descriptive data

Variable	Definition
IPParticip%	The sum of TIME for actual session with IP as speaker, as a percentage of total session time
IPInterrupt#	Total number of Interruptions/Interjections by IP for this session
IPInterrupt/hr	IPInterrupt# divided by total session time (hr)
IP_PFQuest#	Total number of Probing/Follow-up Questions by IP for this session
IP_PFQuest/hr	IP_PFQuest# divided by total session time (hr)

IPAnalysisTime	Total time of Analysis utterances by IP for this session
IPAnalysis%	IPAnalysisTime divided by total time of Analysis utterances by all participants
CrewPart%	Crew Participation %. The sum of TIME for this session where CA or FO is the speaker, as a percentage of total session time
Crew S1#	The total number of Crew S1 statements for this session. All self-initiated, substantive crew statements that raise issues, introduce topics, or add information to an existing topic are coded "S1"
Crew S1/hr	Crew S1# divided by total session time (hr)
CrewProact#	The total number of Crew Proactive Questions (not reactive or miscellaneous) for this session
CrewProact/hr	Crew Proact# divided by total session time (hr).
CrewAnalysis#	The total number of Analysing utterances by either CA or FO for this session
CrewAnalysis/hr	CrewAnalysis# divided by total session time (hr)
CrewAnalysisTime	The sum of TIME for all Analysis utterances by either CA or FO for this session
CrewAnalysisTime%	CrewAnalysisTime as a percentage of total session time
CrewSequential%	The number of Crew Sequential utterances as a percentage of the total number of utterances for this session. A Crew Sequential utterance is defined as an utterance by a crew member that follows an utterance by the other crew member. This indicates interaction between crew members rather than responses to IP questions, comments, or statements.
TopicCRM%	The total time for Topic:CRM, as a percentage of total session time. Topic:CRM is defined as: "Pertaining to the coordination and interaction of the crew and specifically relates to one or more CRM issues or topics".
TopicMixed%	The total time for Topic:Mixed, as a percentage of total session time. This topic is defined as: "Has between ⅓ and ⅔ of both CRM and Technical".
TopicTech%	The total time for Topic:Technical, as a percentage of total session time. This topic is defined as: "Pertaining to specific techniques of flying and navigating the airplane and/or managing the systems, without reference to

coordination, planning, communication, judgment, or decision

making among crew members".

TopicNon-Spec%	The total time for Topic:Non-specific, as a percentage of total session time

Debriefing Assessment Battery (DAB)

DAB (McDonnell, 1996) is a tool developed for the purpose of systematic assessment of instructor facilitation effectiveness and the nature of crew participation during debriefings. Provided the rater has sufficient CRM experience, it provides a way to subjectively measure performance using a 1-7 scale from "poor" to "outstanding" by comparing the observed performance to a set of descriptive behavioural markers. This assessment tool is identical to the one used by Dismukes et al. (1997). As none of the observed debriefings included video recordings, this aspect was not included in the DAB data for this study. The instructors' performance was assessed in four different categories: "Introduction", "Questions", "Encouragement" and "Focus on Crew Analysis & Evaluation". The crews' performance was assessed in two categories: "Crew Analysis&Evaluation" and "Depth of Crew Activity".

Each category was assessed and graded in four sub-categories. The average of the scores for the sub-categories was recorded as a combined score for each category. The scores for each of the four IP categories and the two Crew categories were used for statistical analysis, in addition to the IP Overall score (the average of the four IP category scores).

Further explanation of the categories and anchoring to the respective behavioural markers is found in Appendix 1. As the author of this study performed all assessments, no inter-rater reliability is available.

Statistical analysis

The descriptive data and the DAB data for all sessions was analysed using Pearson Bivariate Correlation, 2-tailed, (.05) in the SPSS statistical analysis tool. According to Cohen (1998), the Pearson product-moment correlation coefficient in psychological research can be interpreted according to the following table:

Table 2.
Pearson correlation - Effect size (Cohen, 1988)

Pearson Correlation Coefficient	Effect size
.10-.30	Small
.30-.50	Moderate
>.50	Large

Table 2 illustrates guidelines for assessing the strength of a possible association or correlation between two variables, depending on the Pearson correlation coefficient (Cohen, 1988). These guidelines have been used in this study. However, the sample size for this study

is too small (n=8) to provide any conclusive statistical material, so any findings should be treated as "possible indications" rather than evidence.

Verification of "Characteristics of Crew Participation" variables

To verify the relationship between the descriptive data regarding crew behaviour and the assessed (DAB) outcome, the following crew variables were categorised and analysed for possible correlations with DAB scores for Crew Analysis & Evaluation, and Depth of Crew Activity. The categories and their respective variables were:

1. Level of Activity: CrewPart% (The percentage of total "talk time" by the crew) and CrewSequential% (The percentage of crew utterances that followed another crew utterance, indicating interactivity between crew members rather than just responding to IP questions).
2. Analysis: CrewAnalysis# (the number of analysing utterances by the crew) and CrewAnalysisTime (the total time of crew analysing utterances). Other possible measurements are Crew Analysis/hr (the number of crew analysis utterances per hour) and Crew Analysis% (the percentage of total "talk time" spent on crew analysis utterances). The reason they were not used is the assumption that the quantity (number of crew analysis statements, and crew analysis time) is more indicative of the analysis effort than the rate per hour. For example, a very short debriefing with a high rate of crew analysis may still fail to address many interesting debriefing topics. On the other hand, a longer debriefing with substantially more analysis material discussed may have a lower analysis rate/hr due to an unsuitably high degree of IP participation but still reach further in terms of crew analysis. Therefore, CrewAnalysis# and CrewAnalysisTime were chosen as variables for this category.
3. Proactivity: CrewS1# (the number of self-initiated, substantive crew statements that raise issues, introduce topics, or add information to an existing topic) and CrewProact# (the number of crew proactive questions)

Hypotheses

Based on the descriptive data from the observation material as well as the DAB data, five hypotheses regarding the effects of instructor behaviour were tested using variables corresponding to the expected resulting crew behaviour. Table 3 (below) lists the respective hypotheses and variables.

Table 3.
Hypotheses and selected variables for analysis

Hypothesis	Variables
1. The more the IP talks, the less active the crew will be. This will have a negative impact on crew interaction, self-reflection, and analysis	IPPart%, Crew AnalysisTime, Crew Sequential%, DAB Crew Analysis&Evaluation
2. IP facilitation skills are conducive to crew self-reflection and in-depth self-analysis of crew performance	DAB IP Overall, DAB "Crew Analysis and Evaluation", DAB "Crew Depth of Activity"

3. IP use of Probing/Follow-up questions results in more and better crew analysis	DAB IP "Questions", IP_PFQuest#, Crew Analysis#, Crew AnalysisTime, DAB "Crew Analysis& Evaluation"
4. A good introduction to facilitated debriefing will increase Crew Interaction and Crew Analysis	DAB IP "Introduction", Crew Sequential%, Crew Analysis#, CrewAnalysisTime
5. IP encouragement will result in more crew activity and interaction	DAB IP "Encouragement", Crew Sequential#, CrewPart%

Results

General observations

The observed sessions were either Type Rating Training sessions or Licence Proficiency Check/Operator Proficiency Check sessions in accordance with EASA Part-ORO and Part-FCL (EASA, 2012). The LPC/OPC sessions involved LOFT/LOE-scenarios. The Type Rating Training sessions contained elements of LOFT/LOE. These sessions were standard, 4-hour simulator training sessions preceded by a briefing and followed by a debriefing. The LPC/OPC simulator sessions were instead performed in two 2-hour blocks. Before the actual simulator session started, a briefing was performed. Between the 2-hour blocks, a debriefing was conducted. This debriefing was followed by a break and a short briefing of the next part of the session. After the second 2-hour session was completed, a debriefing of the second part was performed, followed by a summary of the two "parts". Several crew members struggled to remember some of the events encountered during the simulator session. Most of the debriefings followed a strictly chronological format, i.e., all events were mentioned in chronological order, commented upon and - if deemed particularly interesting - became the subject of further discussion and/or analysis.

Descriptive data

Time
The average duration of the debriefings was 34.1 minutes ($SD = 14.9$) ranging between 14.1 and 55.2 minutes.
Participation
IP participation (share of total "talk time") was on average 58.1% ($SD = 8.7$), with a range between 42.3% and 66.3%.
IP Interruptions
The average rate of IP interruptions (IP interrupts either CA or FO) per hour was 18.0 ($SD = 19.7$). There was a very significant difference between the performance of the IPs in this respect - the value ranged between 4.0 and 63.5.
IP Questions
This parameter differs slightly from the data analysed by Dismukes et al. (1997). In this report, only IP Probing questions and IP Follow-up questions ("IP PF") were analysed. The average number of IP PF questions/hr was 21.6 ($SD = 10.5$) and the values varied between 10.7 and 35.9.

Topic

The debriefing topic also varied a lot between the debriefings. On average, 16.2% of the debriefing time was spent talking about CRM issues. The values ranged from 5.1% to 37.8% ($SD = 11.7$). The corresponding figure for Technical issues (pertaining to specific techniques of flying and navigating the airplane and/or managing the systems, without reference to coordination, planning, communication, judgment, or decision making among crew members) was 43.9% ($SD = 17.5$). Mixed CRM/Technical topics occupied 19.7% ($SD = 3.4$) and Non-specific 20.4% ($SD = 11.0$) of the total debriefing time.

Debriefing Assessment Battery (DAB)

The average overall IP score according to DAB was 3.4 (between 3 - Needs improvement and 4-Adequate). The overall scores ranged between 1.8 and 4.6. The four categories where the instructors' facilitation skills were assessed were: "Introduction", "Questions", "Encouragement" and "Focus on Crew Analysis & Evaluation". The average scores for Introduction were the lowest ($M = 2.7$) and the scores for "Encouragement" were the highest ($M = 4.3$). In three cases, the Instructor Pilots gave no introduction to facilitated debriefing.

The crew were assessed in two categories: Crew Analysis&Evaluation and Depth of Crew Activity. The average score for Crew Analysis&Evaluation was 3.3 and for Depth of Crew Activity: $M = 4.5$.

"Characteristics of crew participation" variables

For the category "Level of Crew activity", the following variables were analysed: Crew Participation% and Crew Sequential %. These two variables have a strong inter-correlation: $r = .916$ ($p < 0.001$). When analysed for correlation with DAB Crew Analysis & Evaluation and DAB Depth of Crew Activity, the only significant finding was a strong correlation between Crew Sequential% and Depth of Crew Activity, see Table 4 below.

Table 4.
Correlations (Pearson) in bold, Significance (p) within parenthesis

	DAB Crew Analysis&Evaluation	DAB Depth of Crew Activity
CrewPart%	**.624** (.098) ns	**.660** (.075) ns
CrewSequential%	**.671** (.069) ns	**.717*** (.045)

ns=not significant
* Correlation is significant at the 0.05 level

Concerning quantitative measures of Analysis, the variables Crew Analysis# and CrewAnalysisTime were analysed for possible correlation with DAB scores for Crew Analysis&Evaluation and Depth of Crew Activity. See Table 5 for results.

Table 5.

Correlations (Pearson) in bold, Significance (p) within parenthesis

	DAB Crew Analysis&Evaluation	DAB Depth of Crew Activity
CrewAnalysis#	**.765*** (.027)	**.617** (.103) ns
CrewAnalysisTime	**.708*** (.049)	**.612** (.107) ns

ns=not significant
* Correlation is significant at the 0.05 level

CrewAnalysis# and CrewAnalysisTime are both strongly correlated with the Crew DAB score for Analysis & Evaluation. There appears to be a positive but non-significant correlation with the Depth of Crew Activity score.

Regarding "proactivity", the variables CrewS1# and CrewProact# were compared to the DAB scores for Crew Analysis & Evaluation and Depth of Crew Activity. The results are presented in Table 6 below.

Table 6.

Correlations (Pearson) in bold, Significance (p) within parenthesis

	DAB Crew Analysis&Evaluation	DAB Depth of Crew Activity
CrewS1#	**.567** (.143) ns	**.341** (.408) ns
CrewProact#	**-.166** (.695) ns	**-.322** (.436) ns

ns=not significant

The results do not indicate any statistically significant correlations. Therefore, CrewS1# and CrewProact# are not suitable for any conclusions in this study.

To summarise the variables selected for further analysis of "Characteristics of Crew Participation" based on the above:
- Crew Participation% and CrewSequential% are strongly interrelated, the analysis shows a large positive correlation
- Crew Sequential% is positively correlated with Depth of Crew Activity
- Crew Analysis# and CrewAnalysisTime are positively correlated with DAB Crew Analysis&Evaluation
- CrewS1# and CrewProact# are discarded as variables for further analysis in this study due to lack of significant correlations

Hypotheses vs analysed data

Hypothesis 1: The more the IP talks, the less active the crew will be. This will have a negative impact on crew self-reflection and analysis.

Correlations IP Participation % - Crew variables

Table 7.
Correlations (Pearson) in bold, Significance (p) within parenthesis

	CrewAnalysis Time	Crew Sequential#	DAB "Crew Analysis & Evaluation
IPPart%	**-.071** (.868)　　ns	**-.916**** (.001)	**-.628** (.095)　　ns

ns=not significant
** Correlation is significant at the 0.01level

No correlation is found between IP "talk time" and the total time of "analysing" utterances by the crew. However, there is a very strong and significant negative correlation between the amount of IP "talk time" and the number of Crew Sequential utterances. There is also a strong but non-significant negative correlation between IP Participation% and DAB score for Crew Analysis & Evaluation.

> *Hypothesis 2: IP facilitation skills are conducive to crew self-reflection and in-depth self-analysis of crew performance.*

Correlations DAB IP Overall - Crew variables
Table 8.
Correlations (Pearson) in bold, Significance (p) within parenthesis

	DAB "Crew Analysis & Evaluation"	DAB "Crew Depth of Activity"
DAB IP Overall	**.789*** (.020)	**.599** (.117)　　ns

ns=not significant
* Correlation is significant at the 0.05 level

The analysis indicates a strong correlation between IP overall facilitation skills and the resulting crew performance in analysing and evaluating their performance. No statistically significant correlation is identified between the overall IP score and the score for Crew Depth of Activity.

> *Hypothesis 3: IP use of Probing/Follow-up questions results in more and better crew analysis.*

Correlations DAB IP "Questions" and IP_PFQuest# - Crew variables

Table 9.
Correlations (Pearson) in bold, Significance (p) within parenthesis

	Crew Analysis#	Crew AnalysisTime	DAB Crew Analysis & Evaluation
DAB IP "Questions"	**.159** (.707)　　ns	**.342** (.406)　　ns	**.419** (.301)　　ns
IP_PFQuest#	**.812*** (.014)	**.509** (.198)　　ns	**.467** (.243)　　ns

ns=not significant
* Correlation is significant at the 0.05 level

Regarding the IP DAB score for "Questions", the analysis results in no significant correlations. The number of Probing and Follow-up questions by the IP correlates positively with the number of Crew Analysis utterances.

Hypothesis 4: A good introduction to facilitated debriefing will increase Crew Interaction and Crew Analysis.

Correlations DAB IP "Introduction" - Crew variables
Table 10.
Correlations (Pearson) in bold, Significance (p) within parenthesis

	Crew Sequential%	Crew Analysis#	Crew AnalysisTime
DAB IP "Introduction"	**.441** (.274) ns	**.118** (.780) ns	**.440** (.276) ns

ns=not significant

The data indicates no significant correlations between the IP Introduction score and the analysed crew variables.

Hypothesis 5: IP encouragement will result in more crew activity and interaction

Correlations DAB IP "Encouragement" - Crew variables
Table 11.
Correlations (Pearson) in bold, Significance (p) within parenthesis

	Crew Part%	Crew Sequential#
DAB IP "Encouragement"	**.617** (.103) ns	**.755*** (.030)

ns=not significant
* Correlation is significant at the 0.05 level.

Regarding the IP score for "Encouragement", a strong and significant positive correlation exists with the level of crew interaction measured using the number of Crew Sequential utterances. The analysis also shows a similar but non-significant correlation with the degree of Crew Participation.

Questionnaire

In Table 12 (below), the resulting modal values for the answers to the questionnaires are shown. See Appendix 2 for a comprehensive list of answers.

Table 12.

Crew Questionnaire

I understand why facilitation is used as a debriefing tool	The instructor gave us a sufficient introduction and explained how the debriefing should be conducted	My partner and I did most of the talking	Our analysis went deeper than simply "what happened". We explored the related CRM issues and how they affected our performance	After the debriefing was over, we had identified why things went well (or not so well) and what we might do differently next time we are in a similar situation
Strongly agree	Strongly agree	Slightly agree	Strongly agree	Strongly agree

IP Questionnaire

I understand why facilitation is used as a debriefing tool	I am confident in my abilities as a facilitator	I gave the crew a sufficient introduction and explained how the debriefing should be conducted	The crew did most of the talking	The crew's analysis went deeper than simply "what happened". They explored the related CRM issues and how their performance was affected	After the debriefing was over, we had identified why things went well (or not so well) and what the crew might do differently next time they are in a similar situation	I would like to get more training / coaching as a facilitator	Having a basic guide or structure to base my facilitated debriefings on would be useful for me
Strongly agree	Slightly agree	Slightly agree	Slightly agree	Slightly agree	Slightly agree	Strongly agree	Strongly agree

Discussion

The aim of this study was to answer the following questions: *How does the Instructor Pilot's (IP) behaviour and use of facilitation techniques affect the quality and outcome of the debriefing? Are there measurable and quantifiable aspects of IP behaviour which are positively or negatively correlated with "desirable" Crew debriefing behaviour such as high levels of Crew participation, interaction, analysis and self-reflection?*

The analysis results show several significant correlations between aspects of IP behaviour, and debriefing quality and outcome. These correlations translate as:

- The more the IP talks, the less the crew members interact with each other
- The IP level of facilitation skills is predictive of the degree of Crew analysis and evaluation

- The amount of probing/follow-up questions by the IP is predictive of the amount of crew analysis
- IP encouragement leads to more interaction between the crew members
- The degree of crew interaction is predictive of the depth of crew activity during debriefing

H1: The more the IP talks, the less active the crew will be. This will have a negative impact on crew self-reflection and analysis. The analysis indicates a significant negative correlation between the amount of IP "talk time" and the number of Crew Sequential utterances. There is a large but non-significant negative correlation between IP Participation% and DAB score for Crew Analysis & Evaluation. No relationship is found between IP Participation and the total amount of Crew Analysis utterances. Crew Sequential utterances are indicative of crew interacting with each other, rather than just responding to IP questions. The Crew Sequential variable is positively correlated with DAB Depth of Crew Activity. This supports Hypothesis 1 and suggests that the IP should aim to let the crew members do most of the talking. The lack of statistical significance regarding the negative correlation between IP Participation and Crew Analysis&Evaluation, as well as with the amount of Crew Analysis utterances, may be an effect of the low sample size. If, for instance, a crew with a high ability for self-analysis and evaluation should be paired with an instructor with a propensity for talking a lot, this would create data that contradicts the assumption that a high degree of IP Participation has a negative impact on the resulting analysis and evaluation.

H2: IP facilitation skills are conducive to crew self-reflection and in-depth self-analysis of crew performance. The analysis data indicates that a higher overall IP Facilitation score correlates with higher crew scores for Analysis and Evaluation. A similar but less strong and statistically non-significant positive correlation exists for Depth of Crew Activity. One possible reason for the lack of significant correlation with Depth of Crew Activity is the distribution of DAB IP scores that make up the Overall score. The average score for the category "Encouragement" was 4.3, and for "Introduction" it was only 2.7. Without a comprehensive introduction specifically stating that the crew should "dig deep" and critically analyse their performance", it is less likely that the crew will do so. Therefore, the IP overall score may be "blunt" when it comes to the precise prediction of the depth of crew activity but may still be a general measure and predictive variable for Analysis and Evaluation scores. A similar difference in the "impact" of IP DAB variables on Crew variables is seen in the Dismukes et al. study (1997, p.41). Four out of five Instructor variables correlate stronger with Crew Analysis & Evaluations than with Depth of Crew Activity, the fifth shows an identical correlation. Thus, the data in this study partially supports Hypothesis 2. It is likely that improved facilitation skills will enable the IP to create more learning value for the crew.

H3: IP use of Probing/Follow-up questions result in more and better crew analysis. The measured number of IP Probing/Follow-up questions has a strong positive correlation with the number of Crew Analysis utterances. Non-significant but positive correlations exist with Crew AnalysisTime and DAB Crew Analysis & Evaluation. The relationship between IP PF questions and crew analysis statements is expected. As the purpose of probing and follow-up questions is to get the crew to analyse and deepen the discussion beyond simply "what happened", it is no surprise that the number of crew analysis utterances increase with the number of PF questions. Depending on the IP questions, it is possible that they have resulted in short analysing answers from the crew which would explain the stronger correlation with the number of crew analysis utterances than with Crew AnalysisTime. No significant correlation was detected between the assessed IP performance regarding "Questions" and the crew variables. One reason for this may be the relatively low scores for

IP performance in this category: on average 3.3 (on a scale of 1-7). The quantified analysis data does not differ between Probing and Follow-up questions. Based on the source data, the author's subjective opinion is that most of the questions coded as PF are "Probing" rather than "Follow-up". This is reflected in the DAB scores. Quite often, the facilitator receives replies from the crew that only describe what happened. To get the crew to analyse the "root cause" of an event and be able to create a "lesson learned" that can be applied during a multitude of similar situations during Line Flying, it may be beneficial to get the crew to "dig deeper" by using suitable follow-up questions. For future research, separating "Probing" from "Follow-up" questions may provide useful information about the effects of follow-up questions. Overall, the analysis data only supports the positive relationship between the number of probing/follow up questions, and the number of crew analysis utterances, i.e. "more analysis". Therefore, the answer to Hypothesis 3 is only confirmed regarding "more analysis", and not regarding "better analysis".

H4: A good introduction to facilitated debriefing will increase Crew Interaction and Crew Analysis. Again, only a small positive and non-significant correlation is found between the DAB Introduction score and the parameters for Crew Interaction and Crew Analysis. One possible reason for the lack of significant correlation is the generally low scores for IP Introductions. Three out of eight debriefings were performed without any introduction to facilitated debriefing whatsoever. To be assessed as "adequate" (Grade=4) according to DAB, the IP behavioural markers are: "Conveys that his/her role is guide/facilitator and that crew should do most of the talking, but does not emphasize strongly", "Conveys that crew should take an active role and initiate discussion", "Conveys that crew should analyze the LOFT and their performance" and "Gives a clear, though implicit rationale for the crew to participate actively and make their own analysis". The average score was 2,7 and the scores ranged from 1 to 5, so it is quite possible that only a small percentage of the introductions were sufficiently effective to create a measurable effect on the crew variables.

H5: IP encouragement will result in more crew activity and interaction. This hypothesis is partly supported by the analysis. There is a strong positive correlation between IP Encouragement and Crew Sequential utterings, indicating that IP encouragement has a positive effect on the amount of crew interaction, which in turn is positively correlated with the assessed scores for Depth of Crew Activity. Examples of IP Encouragement are: IP communicates an interest in crew views and actively strives to get them to do most of the talking and lead their own discussion, uses active listening and pauses, avoids interrupting, follows up on crew topics, consistently encourages all members to participate and draws out quiet members as necessary, consistently refrains from lecturing and giving own analysis before crew (McDonnell, 1996). The positive effects of this behaviour is indicated by the correlation with the variable for crew interaction. Concerning the relationship between IP Encouragement and Crew Participation, the analysis resulted in a positive but not sufficiently significant correlation. It is possible that the combined IP score for Encouragement doesn't fully reflect to which degree the IP strives to get the crew to do most of the talking, which is only one aspect of behaviour in this category. For a higher sample size, a more consistent positive effect should be evident. Additional separate analyses of all four components of the Encouragement grade are possible but have not been performed in this study.

Descriptive data

The average duration of the debriefings (34.1 minutes) in this study corresponds relatively closely to the average duration of the debriefings studied by Dismukes et al. (1997), which was 31 minutes. The same can be said for IP Participation. In this study, the IP spoke (on average) 58.1% of the time spent talking by all participants. In the 1997 study, 61% of the

words in debriefings of two-member crews were spoken by the IP. Regarding the suggestion from Dismukes et al. (1997) regarding IP Participation: "too much talking by the instructor discourages participation by the crew members", the statement is supported by the results in this study.

One significant difference between the findings of this study and Dismukes et al. (1997) is the amount of time spent (or words spoken) on the topic of CRM. During this study, discussion of CRM occurred 16.2% of the time. On the other hand, "Technical" topics were discussed 43.9% of the time. While this study was based on "time" as a measurement, and the Dismukes et al. (1997) study used "word count", the relative emphasis of the respective topics should be comparable between the two studies. In Dismukes et al. (1997), 45% (instructors) and 49% (crew) of the total word count were CRM-related. 16% (IP) and 12% (crew) were in the "Technical" category. It appears that CRM-related issues were discussed to a significantly higher degree during the 1997 study whereas technical subjects (pertaining to specific techniques of flying and navigating the airplane and/or managing the systems, without reference to coordination, planning, communication, judgement, or decision making among crew members) were much more in focus during the observed debriefings in this study. There can be many different reasons for this. For instance, it is possible that different airlines have different focus areas where their training efforts are concentrated. It is also possible that the average pilot performance in the respective airlines differs, which in turn affects the focus of the debriefings. Company culture, training culture and national culture as well as "experience on current aircraft type" are other possible reasons for the differences found. This could be a suitable topic for further studies.

Questionnaires

There are several reasons why the results of the post-debriefing questionnaires should be treated cautiously. The participants are usually tired due to the challenging nature of the training session and the debriefing. The simulator sessions are performed 24/7, some of the debriefings are conducted during the night. It is possible that the replies are subject to bias. The participants may have preconceptions regarding what replies are expected, what option their partner is selecting or what replies would please the instructor or observer.

With that in mind, a general comparison between the mode values of the Crew questionnaire results and the corresponding IP questionnaire results shows that the respective assessments of the quality of crew analysis differs: The crew generally "strongly agreed" that their analysis went deeper than simply "what happened", including exploration of CRM issues and how they affected their performance. The results for the Instructor Pilots regarding the crew's analysis were more cautious: "Slightly agree". This could for example indicate that the IP thinks the analysis didn't go deep enough or didn't involve CRM issues (which is quite likely, given the average session only discussed CRM topics for 16,2% of the time).

The questionnaires also reveal how difficult it can be for a participant to correctly assess quantifiable parameters from a completed training session. Only one out of eight crews and two out of eight instructors were able to correctly determine whether the crew or the instructor did most of the talking during the session.

Regarding Introductions, the general crew opinion seems to "strongly agree" to the statement "The instructor gave us a sufficient introduction and explained how the debriefing should be conducted". This may seem surprising given the generally low DAB scores for "Introduction". Bearing in mind that the crews are unlikely to have any facilitation experience, knowledge or training, the replies may indicate that they are unaware of how a facilitated debriefing should be conducted and therefore, any introduction (or no introduction) is deemed sufficient. The IP results regarding the introduction indicate a somewhat higher

degree of awareness, and the answers range from "Slightly disagree" to "Strongly agree" with the mode value = "Slightly agree".

Concerning the statement "I would like to get more training/coaching as a facilitator", the most frequent reply was "Strongly agree", with answers ranging from "Slightly disagree" to "Strongly agree".

Additional findings

Considering the recent introduction of facilitated debriefings and the lack of specific instructor facilitation training in the airline where the observations were conducted, the relatively low DAB scores for Instructor Pilot effectiveness at facilitation are not surprising. It is important to bear in mind that facilitated debriefings in no way eliminate the need for instruction. Instructional skills and facilitation skills are two different things, and a highly skilled facilitator and instructor will find the most suitable balance between facilitation and instructing based on the abilities and performance of the participating crew, the nature of the training event and the debriefing objectives. So, despite the high experience level of the participating instructors, there is a potential for development concerning aspects of facilitation. "Facilitation requires skills beyond those of a traditional instructor who is proficient in conveying information and assessing performance" (Smith et al., 2007, p.664).

An additional consequence of the recent introduction of facilitation as a debriefing method is that the participating crew members are unfamiliar with this concept. According to Pruchnicki (2018): "practitioner performance in facilitated debriefs is typically insufficient as practitioners have never been trained to self-reflect". It is likely that crew performance in debriefing-related aspects will improve significantly over time, as the crews get more practice at guided self-reflection and self-analysis.

Another observation (which is not quantified in this study) is the amount of positive outcomes and successfully handled dilemmas/malfunctions discussed in the analysed material. This may create problems for the Instructor Pilot to determine "debriefing objectives", for example: what events do I want the crew to discuss and learn from? The challenge for the Instructor Pilot is to enable the crew to create as much learning value as possible from the events that occurred. In some cases, the crew have performed exceedingly well during the session and there are hardly any negative aspects of their performance to discuss. These crews are equally deserving of a high-quality facilitated debriefing as an average or low-performing crew. Instead of looking for minor deviations or errors, the IP can ask the crew to analyse why things went so well? Preferably anchoring their self-analysis to the airline's established CRM concepts. The IP may for instance ask, "What options were considered?" or "Please tell me more about your workload/decision-making process/communication at the time you discovered the malfunction?". Alternatively, the IP may introduce a hypothetical element to stimulate the self-analysis process regarding an event that was performed well. For example: "What if your partner had been less proactive during the non-precision approach? How would you have dealt with that?" or "Let's say the fuel leak hadn't stopped. What would you have done?".

Limitations

In addition to the limited number of observations (n=8), there are other possible sources of error. Among these are: transcription errors, misunderstandings, language barriers, bias, risk of method error and insufficient span/differences between the observed Instructor Pilots in the level of competence at facilitation techniques. As the author of this report personally conducted all observations, transcriptions and assessments, any bias or

misunderstandings are likely to apply to all observations to a similar degree. The author/observer has >20 years of experience from the participating airlines' operational environment. This is likely to reduce the number of misunderstandings. Regarding bias, the design of the study should limit any such effects by the use of coding rules, DAB behavioural markers and the analyses that compare the observation data with the measured data. Method errors in this study may for instance be potentially incorrect assumptions about what the variables used truly represent, and whether a Likert (1932) scale can be applied to the DAB variables - is the difference between each grade linear? A mitigating fact concerning method error is the degree of replication between this study and Dismukes et al. (1997), as well as the general alignment of the respective results between the two studies. The difference in methods concerning *word count* versus *utterance time/ number of utterances* appears to be of minor significance, given the similarity of results between the studies.

Concerning the validity of this study, it rests on several assumptions such as: *LOS can be characterised as "Experiential Learning". Self-reflection is conducive to learning from an experience. The author/observer is proficient in using the DAB to assess the facilitated debriefing performance of instructors and crews. The DAB is suitable for determining the characteristics of instructor and crew performance. The crew and instructor performance are transcribed consistently and correctly. The calculations associated with the analysis are correct.* If the above assumptions were incorrect, the results obtained would be invalid. However, replicating a previous study (Dismukes et al. 1997) by using the same (or similar) quantitative measures and procedures, as well as the DAB, and achieving corresponding results despite the significantly smaller sample size, supports the validity and reliability of the findings of this study. Regarding generalisation, the 1997 study was performed with participants from five different airlines. Therefore, even though this study only included participants from one airline, the similarity with the 1997 study suggests that the findings of this study may be valid also for other airlines. Whether the results are applicable to other facilitated debriefing situations may be a suitable topic for further study.

Conclusions

The analyses and descriptive data of this study illustrate an early stage of implementation of facilitation as a debriefing tool in a well-established airline with very highly experienced pilots. The findings suggest that the level of Instructor Pilot facilitation skills is predictive of several characteristics of crew behaviour during debriefing, including the quality of their self-analysis and evaluation. Due to the relatively small number of observations, the statistical analysis results of this study should be interpreted with caution. However, the findings support the conclusions of Dismukes, Jobe and McDonnell (1997).

Recommendations

Suggestions for instructor development (based on the observations made during this study):

- A good introduction, outlining roles and expectations, is likely to improve the chances of a successful debriefing. Especially when the participating crew has little or no previous experience of facilitation. Explaining the rationale for facilitated debriefing will increase the "buy-in" of the crew.
- Let the crew do the talking, as much as possible. Avoid giving long speeches. Be an active listener. You want the crew to feel that *they* are the focus of the debriefing, not you.

- If the crew are struggling to self-analyse, try to guide them "back on track" with the smallest possible means. For instance, use silence. Allow them to think. Rephrase the question, and/or use building questions. Avoid interrupting and avoid giving your own analysis too soon. Only resort to "lecturing" if the crew is truly unresponsive or incapable of self-reflection and self-analysis.
- Summarise the main "lessons learned" and how they can be useful during line flying. Or even better - ask the crew to summarise! You can always fill in any missing details, if needed.

Suggested topics for further research

- Comparative study of debriefing topics in a setting where signs or cards depicting CRM aspects and tools as well as pilot competencies are visible during the debriefing. Will the discussions include more CRM topics as a result?
- Comparative study of Structured self-debriefing under the supervision of a facilitator vs facilitated debriefing.
- Comparative study of video-assisted facilitated debriefings vs non-video-assisted. Many Full Flight Simulators have the technical means to video-record the training sessions. Some airlines use this feature, some don't. What are the possible benefits and disadvantages of using video recordings for debriefing purposes? Is there a measurable difference in the quality of the debriefing outcome?

References

Boud, D., Keogh, R., & Walker, D. (1985). Debriefing in Experience-Based Learning. *Reflection: Turning experience into learning*. Routledge.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd Ed). L. Erlbaum Associates.

Dismukes, R.K., Jobe, K., & McDonnell, R.K. (1997). *LOFT Debriefings: An Analysis of Instructor Techniques and Crew Participation*. NASA Technical Memorandum 110442. DOT/FAA/AR-96/122

Dismukes, R.K., & Smith, G.M. (2000). *Facilitation and Debriefing in Aviation Training and Operations.* Ashgate Publishing Ltd.

European Union Aviation Safety Agency. (2012). *Commission Regulation (EU) No 965/2012*.

European Union Aviation Safety Agency. (2019). *Appendix to Opinion No 08/2019 (A) (RMT.0599).* https://www.easa.europa.eu/downloads/116001/en

European Union Aviation Safety Agency. (2021). *AMC and GM to Part-ORO of Regulation (EU) No 965/2012* Annex III to ED Decision 2021/002/R

Federal Aviation Administration. (2015). *Subject: Flightcrew Member Line-Operational Simulations: Line-Oriented Flight Training, Special Purpose Operational Training, Line Operational Evaluation.* Advisory Circular 120-35D.

Foushee H. C., & Helmreich, R. L. (1988). Group interaction and flight crew performance. In E. L. Wiener, & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 189-228). New York: Academic Press.

Gantt, L. T., Overton, S. H., Avery, J., Swanson, M., & Elhammoumi, C. V. (2018). Comparison of debriefing methods and learning outcomes in human patient simulation. *Clinical Simulation in Nursing*, 17, pp. 7-13. https://doi.org/10.1016/j.ecns.2017.11.012.

Helmreich, R. L., & Foushee, H. C. (2010). Why crew resource management? Empirical and theoretical bases of human factors training. *Crew resource management*. Elsevier.inc. https://doi.org.10.1016/B978-0-12-374946-8.10001-9

International Civil Aviation Organization. (2013). *Manual of Evidence-Based Training. Doc 9995.* (1st Ed). Downloaded from https://skybrary.aero/sites/default/files/bookshelf/3177.pdf

Jacobson, M., & Ruddy, M. (2004). *Open to outcome* (p. 2). Oklahoma City, OK. Wood'N'Barnes.

Jamieson, S. (2004). Likert scales: how to (ab) use them. *Medical Education, 38(12)*, 1217-1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

Kolb, D. (1984). *Experiential learning: experience as the source of learning and development.* Prentice-Hall.

Lederman, L.C. (1992). Debriefing: Toward a systematic assessment of theory and practice. *Simulation & Gaming*. Jun92, Vol. 23 Issue 2. https://doi.org/10.1177/1046878192232003

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology,* 140, 1–55.

McDonnell, L.K. (1996). *Facilitation Techniques as Predictors of Crew Participation in LOFT Debriefings.* NASA Contractor Report 196701. Downloaded 2022-04-19 from https://ntrs.nasa.gov/api/citations/19960042887/downloads/19960042887.pdf

McDonnell, L.K., Jobe, K.K., & Dismukes, R. K. (1997). *Facilitating LOS Debriefings: A Training Manual.* NASA Technical Memorandum 112192 DOT/FAA/AR-97/6.

McNaughton, D., Hamlin, D., McCarthy, J., Head-Reeves, D., & Schreiner, M. (2008). Learning to listen: Teaching an active listening strategy to preservice education professionals. *Topics in Early Childhood Special Education*, 27, 223- 231. https://doi.org/10.1177/0271121407311241

Pruchnicki, S. A. (2018). *Improving Facilitated Debriefings - How are barriers to learning recognized by instructors and mitigated during post-simulator debriefings?* The Ohio State University. https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu153977469273507&disposition=inline

Richie, J., & Lewis, J. (2003). *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. SAGE Publications, London.

Rogers, C. (1969). *Freedom to Learn: A View of what Education Might Become*. C.E. Merrill Publishing Company.

Smith, M. O., Smith, G. M., Budenberg, C., Willats, C., & Griffiths, C. (2007). *An Airline Evaluation of the Facilitation Skills Assessment Tool [FSAT] as a Device for Training and Assessing Facilitators.* 2007 International Symposium on Aviation Psychology, 664-669. https://corescholar.libraries.wright.edu/isap_2007/22

Wain, A. (2017). Learning through reflection. *British Journal of Midwifery*. October 2017, Vol 25, No 10. https://doi.org/10.12968/bjom.2017.25.10.662

Weger, H. Jr. , Bell, G.C., Minei, E.M., & Robinson, M.C. (2014) The Relative Effectiveness of Active Listening in Initial Interactions. *International Journal of Listening,* 28:1, 13-31. https://doi.org.10.1080/10904018.2013.813234

Appendix 1: Anchoring of the Debriefing Assessment Battery
(McDonnell, Jobe & Dismukes, 1997)

Instructor Profile

The Instructor Profile is a summary of the strategies and techniques IP's use to assist crews in conducting their own debriefings while giving direction and focus as necessary. The two main goals of the debriefing 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. These goals are based on the assumption that active participation by the crew will result in a higher level of learning and increased likelihood of transfer to the line.

Directions:
Use the scale below to rate the instructors on each of the following elements:
Poor (1), Marginal (2), Needs Improvement (3), Adequate (4), Good (5) Very Good (6), Outstanding (7).

IP Introduction

Outstanding:
- Very specifically and thoroughly explains that the role of the IP is guide/facilitator and that crew should do most of the talking and lead the discussion
- Sets strong expectations for proactive crew participation, explicitly stating they should initiate discussion rather than just responding to IP questions
- Explicitly and emphatically states that crew should dig deep, critically analyzing the LOFT and their performance
- Gives a persuasive rationale for the crew to participate actively and make their own analysis and makes a strong case for why it is important to do it this way.

Very Good:
- Clearly conveys that the role of the IP is guide/facilitator and that crew should do most of the talking and lead the discussion
- Clearly conveys that crew should take an active role, initiating discussion rather than just responding to IP
- Clearly conveys that crew should dig deep, critically analyzing the LOFT and their performance
- Clearly conveys the general rationale for the crew to participate actively and make their own analysis

Good:
- Conveys that the role of the IP is guide/facilitator and that crew should do most of the talking, but not specifically that they should lead their own discussion.
- Conveys that crew should take an active role, initiating discussion rather than just responding to IP
- Conveys that crew should dig deep, critically analyzing the LOFT and their performance
- Makes a general statement of the rationale for the crew to participate actively and make their own analysis

Adequate:
- Conveys that the role of the IP is guide/facilitator and that crew should do most of the talking, but does not emphasize strongly
- Conveys that crew should take an active role and initiate discussion
- Conveys that crew' should analyze the LOFT and their performance
- Gives a clear, though implicit rationale for the crew to participate actively and make their own analysis

Needs Improvement:
- Implies that the role of the IP is guide/facilitator and that crew should do most of the talking, but does not emphasize strongly
- Implies that crew should take an active role and initiate discussion
- Implies that crew should analyze the LOFT and their performance
- Gives a vague, implicit rationale for the crew to participate actively and make their own analysis

Marginal:
- Implies that the role of the IP is guide/facilitator and that the crew should talk, but does not emphasize
- Implies that crew should take an active role, but does not specify what they should do.
- Implies that crew should discuss the LOFT and their performance
- Gives vague impression of why crew should participate actively

Poor
- Does not make clear that the role of the IP is guide/facilitator or that crew should do most of the talking
- Does not make clear that crew should take an active role or initiating discussion
- Does not make clear that crew should dig deep or critically analyze the LOFT and their performance
- Does not give rationale for the crew to participate actively and make their own analysis

IP Questions

Outstanding:
- Consistently asks questions as appropriate to get crew talking & lead them to issues
- Consistently rewords questions or otherwise avoids answering for the crew when they do not respond immediately or correctly, and consistently uses a pattern of questioning that keeps the focus on the crew
- Consistently uses probing and follow-up questions as a tool to evoke in-depth discussion and optimize crew self-discovery, while forcing crew to go beyond yes/no and brief factual answers
- Consistently uses questioning techniques to encourage substantial interaction and sharing of perspectives among crew members

Very Good:
- Frequently asks questions when appropriate to get crew talking & lead them to issues
- Predominantly rewords questions or otherwise avoids answering for the crew when they do not respond immediately or correctly and predominantly uses a pattern of questioning that keeps the focus on the crew
- Frequently uses probing and follow-up questions as a tool to evoke in-depth discussion and optimize crew self-discovery, pushing crew to go beyond yes/no and brief factual answers

- Frequently uses questioning techniques to encourage interaction and sharing of perspectives among crew members

Good:
- Generally asks questions as necessary to get crew talking & lead them to issues
- Generally rewords questions or otherwise avoids answering for the crew when they do not respond immediately or correctly and generally uses a pattern of questioning that keeps the focus on the crew
- Generally uses probing and follow-up questions to get crew to analyze in depth and to go beyond yes/no and brief factual answers but may steer crew to predetermined answers while emphasizing self-discovery.
- Generally uses questioning techniques to encourage interaction and sharing of perspectives among crew members

Adequate:
- About half of the time asks questions when necessary to get crew talking & lead them to issues
- Generally avoids answering for the crew when they do not respond immediately or correctly, but may not reword the questions. On average uses a pattern of questioning that keeps the focus on the crew
- On average uses probing and follow-up questions to get crew to analyze in depth and to go beyond yes/no and brief factual answers but steers crew to predetermined answers as much as emphasizes self-discovery.
- On average uses questioning techniques to encourage interaction among crew members

Needs Improvement:
- Sometimes asks questions when necessary to get crew talking & lead them to issues
- To some extent avoids answering for the crew when they do not respond immediately or correctly and uses a pattern of questioning that keeps the focus on the crew
- Sometimes uses probing and follow-up questions to get crew to analyze in depth and to go beyond yes/no and brief factual answers but steers crew to predetermined answers more than emphasizes self-discovery.
- Sometimes uses questioning techniques to encourage interaction among crew members

Marginal:
- Occasionally asks questions to get crew talking & lead them to issues
- Occasionally avoids answering for the crew when they do not respond immediately or correctly but generally answers for them rather than keeping focus on the crew.
- Occasionally uses probing and follow-up questions to get crew to analyze in depth but generally settles for yes/no and brief factual answers
- Occasionally uses questioning techniques to encourage interaction among crew members

Poor:
- Rarely asks questions to get crew talking or lead them to issues
- Usually answers for the crew when they do not respond immediately or correctly.
- Rarely uses probing and follow-up questions to get crew to analyze in depth. Usually settles for yes/no and brief factual answers
- Rarely uses questioning techniques to encourage interaction among crew members

IP Encouragement

Outstanding:
- Consistently communicates an interest in crew views and actively strives to get them to do most of the talking and lead their own discussion.
- Consistently uses active listening and pauses, avoids interrupting, and follows up on crew topics.
- Consistently encourages all members to participate and draws out quiet members as necessary.
- Consistently refrains from lecturing and giving own analysis before crew.

Very Good:
- Clearly communicates to the crew that their views are important and works to get them to do most of the talking and to lead their own discussion.
- Frequently uses techniques such as active listening and pauses, avoids interrupting, and follows up on crew topics to encourage continued discussion.
- Frequently encourages all members to participate and attempts to draw out quiet members as necessary.
- Usually refrains from lecturing and giving own analysis before crew.

Good:
- Shows a clear interest in crew views and attempts to get them to do most of the talking. Makes an effort to get crew to lead their own discussion.
- Often uses active listening and pauses, avoids interrupting, and follows up on crew topics.
- Generally encourages all members to participate, drawing out quiet members as necessary.
- Sometimes lectures, but generally gets crew to analyze situation before giving own analysis.

Adequate:
- On average demonstrates a desire to have crew participate and discuss their views.
- Uses some facilitation techniques to encourage crew discussion and generally avoids interrupting them. Acknowledges crew topics but may not follow up on them thoroughly.
- Attempts to get all crew members involved.
- On average gets the crew to analyze the situation themselves before evaluating and lecturing to them.

Needs Improvement:
- Shows interest in crew views but does not push them to do most of the talking.
- Sometimes uses active listening and pauses, and follows up on crew topics, but also sometimes interrupts.
- Expresses a desire for crew to participate but does not put a lot of effort into getting all members actively involved.
- Sometimes lectures rather than letting crew do the talking.

Marginal:
- Exhibits only modest interest in crew views.
- Only occasionally uses active listening, pauses, and/or follows up on crew topics, and often interrupts.
- Expresses a desire for crew to participate but puts minimal effort into actively encouraging them to do so.
- Tends to lecture and analyze for crew without encouraging them to discuss what happened themselves.

Poor:
- Gives the impression that crew views are not valued.
- Frequently hinders rather than encourages crew talk and does not follow up on topics initiated by crew.

- Makes little attempt to get crew members to participate.
- Frequently lectures to crew about what they did and how to improve.

IP Focus on Crew Analysis and Evaluation

Outstanding:
- Continually encourages and pushes crew to analyze along CRM dimensions the situation that confronted them, what they did to manage the situation, and why they did it.
- Consistently encourages and pushes crew to evaluate their performance and/or ways they might improve.
- Consistently encourages crew to explore CRM issues and how they specifically affect LOFT performance and line operations.
- Continually encourages crew to analyze issues, factors, and outcomes in depth, going beyond simply describing what happened and what they did.

Very Good:
- Frequently encourages and pushes crew to analyze along CRM dimensions the situation that confronted them, what they did to manage the situation, and why they did it.
- Frequently encourages crew to evaluate their performance and/or ways they might improve.
- Frequently encourages crew to explore CRM issues and how they specifically affect LOFT performance and line operations.
- Frequently encourages crew to analyze issues, factors, and outcomes in depth, going beyond simply describing what happened and what they did

Good:
- Generally encourages crew to analyze along CRM dimensions the situation that confronted them, what they did to manage the situation, and why they did what they did, but may settle for less than extensive discussion.
- Generally encourages crew to evaluate their performance and/or ways they might improve.
- Generally encourages crew to explore CRM issues, and attempts to get crew to discuss how they specifically affect LOFT performance and line operations.
- Generally encourages crew to analyze issues, factors, and outcomes in depth. Generally encourages crew to go beyond simply describing what happened and what they did.

Adequate:
- On average encourages crew to analyze along CRM dimensions the situation that confronted them and what they did to manage the situation. Encourages but does not push crew to analyze why they did what they did.
- Tends to encourage crew to evaluate their performance and/or ways they might improve, but may not pursue thoroughly.
- On average encourages crew to explore CRM issues but tends not to get crew to discuss how they specifically affect both LOFT performance and line operations.
- Generally encourages crew to analyze issues, factors, and outcomes, but settles for moderate depth, sometimes letting crew simply describe what happened and what they did.

Needs Improvement:
- Sometimes encourages crew to analyze along CRM dimensions the situation that confronted them and what they did to manage the situation but does not push crew to discuss why they did what they did.
- Verbally requests but does not pursue getting the crew to evaluate their performance and/or ways they might improve.
- Encourages crew to explore CRM issues but does not ask crew to discuss how they specifically affect LOFT performance and line operations.

- Tends not to push crew to analyze issues, factors, and outcomes in depth. Often settles for letting the crew simply describe what happened and what they did.

Marginal:
- Only minimally encourages crew to analyze along CRM dimensions the situation that confronted them and/or what they did to manage it. Does not push crew to discuss why they did what they did.
- Only occasionally encourages crew to evaluate their performance and/or ways they might improve.
- Occasionally encourages crew to explore CRM issues, and does not encourage crew to discuss how they affect LOFT performance or line operations.
- Only occasionally encourages crew to analyze issues, factors, and outcomes in depth. Content for crew to describe what happened and what they did.

Poor:
- Does not encourages crew to analyze along CRM dimensions the situation that confronted them, what they did to manage the situation, or why they did it.
- Rarely encourages crew to evaluate their performance or ways they might improve.
- Rarely encourages crew to explore CRM issues.
- Rarely encourages crew to analyze issues, factors, and outcomes in depth

Crew profile

The crew profile measures the degree and depth of participation by the crew.

Directions:
Use the scale below to rate the crew on each of the following elements, then average the scores to get the overall rating for each category.
Poor (1), Marginal (2), Needs Improvement (3), Adequate (4), Good (5), Very Good (6), Outstanding (7).

Crew Analysis and Evaluation

Outstanding*:*
- Consistently analyze along CRM dimensions the situation that confronted them, what they did to manage the situation, and why they did it.
- Consistently evaluate their performance and ways they might improve.
- Consistently explore CRM issues and how they affect LOFT performance and line operations.
- Consistently analyze issues, factors, and outcomes in depth, going beyond simply describing what happened and what they did.

Very Good:
- Frequently analyze along CRM dimensions the situation that confronted them, what they did to manage the situation, and why they did it.
- Frequently evaluate their performance and ways they might improve.
- Often explore CRM issues and how they affect LOFT performance and line operations.
- Frequently analyze issues, factors, and outcomes in depth, going beyond simply describing what happened and what they did.

Good:
- Generally analyze along CRM dimensions the situation that confronted them and what they did to manage the situation.

- Briefly discuss why they did what they did.
- Generally evaluate their performance and ways they might improve.
- Generally explore CRM issues and how they affect LOFT performance and/or line operations.
- Generally analyze issues, factors, and outcomes in moderate depth, usually going beyond simply describing what happened and what they did.

Adequate:
- On average analyze along CRM dimensions the situation that confronted them and what they did to manage the situation.
- Briefly discuss why they did what they did.
- On average evaluate their performance and/or ways they might improve.
- On average explore CRM issues and how they affect LOFT performance and/or line operations.
- Analyze some issues, factors, and outcomes in some depth, often going beyond simply describing what happened and what they did.

Needs Improvement:
- Only part of the time analyze along CRM dimensions the situation that confronted them, what they did to manage the situation, or why they did it.
- Only sometimes evaluate their performance and ways they might improve.
- Sometimes explore CRM issues but give little discussion of how they affect LOFT performance or line operations.
- Analyze only a few issues, factors, and outcomes in any depth, sometimes going beyond simply describing what happened and what they did.

Marginal:
- Occasionally analyze along CRM dimensions the situation that confronted them. Occasionally discuss what they did to manage the situation or why they did it.
- Only occasionally evaluate their performance and do not discuss ways they might improve.
- Only occasionally explore CRM issues and do not discuss how they affect LOFT performance and line operations.
- Analyze issues, factors, and outcomes in very little depth, rarely going beyond simply describing what happened and what they did.

Poor:
- Do little to analyze along CRM dimensions the situation that confronted them, what they did to manage the situation, or why they did it.
- Rarely evaluate their performance or ways they might improve.
- Rarely explore CRM issues and how they affect LOFT performance and line operations.
- Do not analyze issues, factors, and outcomes in depth; only briefly describe what happened.

Depth Of Crew Activity

Outstanding:
- Consistently go substantially beyond minimal responses to IP questions.
- Consistently participate deeply and thoughtfully.
- Continually initiate dialogue and pursue issues to completion rather than just responding to questions, and consistently interact with each other rather than only with the IP.
- Behave in a consistently proactive rather than reactive manner, being actively involved rather than just passing through the training.

Very Good:
- Frequently go substantially beyond minimal responses to IP questions.
- Usually participate deeply and thoughtfully.

- Frequently initiate dialogue rather than just responding to questions, and often interact with each other rather than only with the IP.
- Usually behave in a proactive rather than reactive manner, being actively involved rather than just passing through the training.

Good:
- Generally go well beyond minimal responses to IP questions.
- Generally participate deeply and thoughtfully.
- Tend to initiate dialogue rather than just responding to questions and generally interact with each other rather than only with the IP.
- Generally behave in a proactive rather than reactive manner, being actively involved rather than just passing through the training.

Adequate:
- On average go somewhat beyond minimal responses to IP questions.
- On average participate somewhat deeply and thoughtfully.
- On average initiate dialogue rather than just responding to questions and interact with each other rather than only with the IP.
- On average behave in a proactive rather than reactive manner, being actively involved rather than just passing through the training.

Needs Improvement:
- Tend to give slightly more than minimal responses to IP questions.
- Sometimes participate deeply and thoughtfully.
- Tend to just respond to questions rather than initiate dialogue. Tend to interact with the IP more than with each other.
- Sometimes behave in a more reactive than proactive manner.

Marginal:
- Frequently give only minimal responses to IP questions.
- Only occasionally participates deeply or thoughtfully.
- Tend to just respond to questions rather than initiate dialogue. Only occasionally interact with each other; tend to interact only with IP.
- Behave in a generally reactive rather than proactive manner.

Poor:
- Consistently gives only minimal responses to IP questions.
- Rarely participate deeply or thoughtfully.
- Rarely initiate dialogue; usually just respond to IP. Rarely interact with each other.
- Behave in a consistently reactive rather than proactive manner. Appear to just pass through the training rather than being actively involved.

Appendix 2: Questionnaires

The numbers in each answer box represent the total number of replies received for that answer option.

Crew Questionnaire

I understand why facilitation is used as a debriefing tool.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
		7	9

The instructor gave us a sufficient introduction and explained how the debriefing should be conducted.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
		1	15

My partner and I did most of the talking.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
		11	5

Our analysis went deeper than simply "what happened". We explored the related CRM issues and how they affected our performance.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
		4	12

After the debriefing was over, we had identified why things went well (or not so well) and what we might do differently next time we are in a similar situation.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
		1	15

Instructor Questionnaire

I understand why facilitation is used as a debriefing tool.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
			8

I am confident in my abilities as a facilitator.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
	2	5	1

I gave the crew a sufficient introduction and explained how the debriefing should be conducted.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
	2	4	2

The crew did most of the talking.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
	1	4	3

The crew's analysis went deeper than simply "what happened". They explored the related CRM issues and how their performance was affected.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
	3	3	2

After the debriefing was over, we had identified why things went well (or not so well) and what the crew might do differently next time they are in a similar situation.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
		5	3

I would like to get more training/coaching as a facilitator

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
	2	1	5

Having a basic guide or structure to base my facilitated briefings on would be useful for me.

STRONGLY DISAGREE	SLIGHTLY DISAGREE	SLIGHTLY AGREE	STRONGLY AGREE
		4	4

Appendix 3: Coding rules

The coding rules defined below are based on the coding rules used by Dismukes, Jobe & McDonnell in "LOFT Debriefings: An Analysis of Instructor Techniques and Crew Participation" (1997). Some adaptations have been made to reduce the amount of analysis time, mainly the use of the variable *time* instead of *word count* as a quantitative measure and the omission of some coding parameters such as Completed/Unfinished sentences, Question Target and Video Analysis related parameters.

Utterance time (TIME). Record the length of each utterance, in seconds.

Speaker (SPKR) Identify the speaker of each utterance using one of the following; IP, CA or FO.

Transcribing Utterances (UTTERANCE) 1. Transcribe the audiotape verbatim. 2. Record all pauses 3 seconds or longer in bold type. . . Type titles in parentheses [e.g., (CA) or (FO)] in place of spoken names and type (XX) in place of spoken name of airline. If an utterance is phrased as a statement but is intended to evoke a response, end the utterance with a "(?)" so it can be coded as a command. If a speaker is interrupted (interjections of active listening or brief interruptions which do not change the flow of the original speaker's utterance) or is talked over but clearly continues on to complete the sentence or thought, transcribe and code the continuation(s) as part of the initial utterance with ".." where the interruption or interjection occurs, and type and code each interrupting utterance separately below ('T' in the INT column). 7. If a speaker makes a statement and then asks a question during a single speaker turn, break it into two separate utterances where the question begins. 8. If a speaker clearly changes topics in the middle of a single speaker turn, transcribe and code the topic change as a separate utterance.

Interruptions / Interjections (INT) 1. Code all utterances that interrupt or interject the preceding speaker as "I".

Utterance Type (TYPE) Question = Any utterance that explicitly asks a question. Command - - Any IP utterance that commands a response but is not phrased in question form. Response = First utterance by any or all crew members following a Question or Command, unless content of utterance makes it obvious that it is non-responsive. S1 (crew) = All self-initiated, substantive crew statements that raise issues, introduce topics, or add information to an existing topic. Statement = All utterances that do not fit the criteria for Q, C, R, or S 1, unless content makes it obvious that the utterance is responsive (R) to the preceding Q or C (e.g., when separated by an intervening utterance).

Crew Proactive Questions (PAQ) 1. Record a "P" in the crew PAQ column if crew question is proactive, or an "O" (other) if the question is not proactive (i.e., reactive or misc.)

Proactive questions include clarification/verification questions used to raise new issues or bring new information into the conversation (e.g., "You wanted help?") and questions designed to gather information (e.g., "Did we have runway three?")

Topic Type (TYPE) CRM = Pertains to the coordination and interaction of the crew and specifically relates to one or more CRM issues or topics. Technical = Pertains to specific techniques of flying and navigating the airplane and/or managing the systems, without reference to coordination, planning, communication, judgment, or decision making among crew members. Mixed = Has between 1/3 and 2/3 of both CRM and technical. Non-Specific = Does not refer specifically to either CRM or technical topics. Includes undetermined, extraneous, procedural, and maintenance of discourse.

Analysis (ANALYSIS) Code all utterances that indicate the speakers are analyzing the situation &/or their performance in the LOFT by considering any of the following issues (both explicit and implicit) as A (Analyzes). Code all utterances which are not analytical as 0 (Other). Generally, analyzing utterances are those that go beyond just describing what happened to discussing why it happened and identifying what factors contributed to the situation and/or how these factors influenced the outcome. • explanations of why something was done and/or done a certain way, or what could have been done differently. Key words include: because, should have, could have, and might have (e.g. "I think we could have performed faster in holding because we had to take a couple of turns in holding just to make sure we got set up" and "I felt a little disorganized pushing off and taxiing out and doing all of that and then having to de-ice; that breaks your flow because you don't put the flaps down") • how & why factors influenced decisions, actions, and outcomes (e.g. "The reason this influenced my decision/actions was ..." and "I was thinking this, so I did this"). • contingencies (e.g. "It might have been a lot different/if we had asked for more time before we took that turn. Maybe I should have asked for one more minute.")

Probing and Follow-up questions (PFQUEST) Code all utterances that indicate the IP is using probing and follow-up questions to get crew to analyze in depth and to go beyond yes/no and brief factual answers as "PF"

Crew Sequential Utterances (CREWSEQUENTIAL) Code all utterances where a crew member speaks following an utterance of the other crew member (as opposed to an utterance by a crew member following an utterance by the IP).