Beaubien, J. M, & Baker, D. P. (in press). Post-training feedback: The relative effectiveness of team- versus instructor-led debriefs. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*, Denver, CO.

POST-TRAINING FEEDBACK: THE RELATIVE EFFECTIVENESS OF TEAM- VERSUS INSTRUCTOR-LED DEBRIEFS

J. Matthew Beaubien
David P. Baker
American Institutes for Research
Washington, DC

In many high-risk domains, simulators are used for training and evaluating team performance under realistic conditions. Once the simulation is complete, the teams review their performance to identify the lessons that they have learned. These post-training debrief sessions may be either instructor- or team-led. Unfortunately, the relative effectiveness of instructor- versus team-led debriefs remains unclear. To address this question, we surveyed a nationwide, representative sample of over 30,000 pilots from 24 U.S. airlines. Despite having a high degree of statistical power and a reliable scale, we found no statistically or practically significant differences among the four most common approaches to post-training feedback: team debrief with videotape, team debrief without videotape, instructor debrief with videotape, and instructor debrief without videotape. The results suggest that all four approaches may be equally effective.

INTRODUCTION

In many high-risk domains – such as aviation, medicine, and nuclear power – high-fidelity simulators are used for training and evaluating team performance under realistic conditions (Butler, 1993; Gaba, Howard, Fish, Smith, & Sowb, 2001). During the simulation, the team members practice their trained skills. Following the simulation, the teams review their performance to identify the lessons that they have learned. Both components of training – skills practice and feedback – are essential for improving team performance (Butler, 1993).

Such post-training feedback sessions are often facilitated by a trained instructor (Dismukes, McDonnell, & Jobe, 2000). Instructor-led debriefs have the advantage of an impartial observer who can objectively comment on the team's strengths and weaknesses. Moreover, because they receive formal training in group facilitation techniques, instructors are highly skilled at diagnosing team performance deficiencies, probing for additional information, using silence to elicit thoughtful responses, and ensuring that feedback does not erode team morale (Dismukes, McDonnell, Jobe, & Smith, 2000).

Alternatively, the teams may debrief themselves (Butler, 1993). Team-led debriefs require the team members to actively participate in the learning process by identifying their deficiencies, diagnosing the underlying causes, synthesizing the lessons that they have learned, and developing specific plans for improving their performance. Team-led debriefs also provide team members with the skills and confidence to debrief their performance in the post-training environment (Butler, 1993).

All debriefings, whether they are instructor- or team-led, work best when they provide specific examples of good and bad performance. As a result, audiovisual recording equipment is currently used in many simulators (Dismukes, et al., 2000; Gaba, et al., 2001). Audiovisual equipment is best suited for teamwork-related skills such as communication and decision-making that are easily captured on tape. It is somewhat less useful for technical skills such as interacting with automation, because the instrument displays may be difficult to capture on videotape (O'Neill, personal communication, 2003).

The purpose of this study was to assess the relative effectiveness of different approaches to debriefing team performance: team debrief with videotape, team debrief without videotape, instructor debrief with videotape, and instructor debrief without videotape. We hypothesized that the four approaches would not be equally effective. However, the lack of consensus in the literature made it impossible to hypothesize whether team- vs. instructor-led debriefs would be more effective. Based on our personal experience, we hypothesized that debriefings which incorporate videotape would be perceived as more effective than those which do not.

METHOD

Participants

The survey participants included 10,166 line-qualified pilots from 24 U.S. airlines. There were roughly equal numbers of Captains (49.1%) and First

Officers (44.4%). The participants included both highly seasoned veterans and relative novices. A sizeable number reported that they had logged over 14,000 hours in commercial and military aircraft (25.8%). However, most reported having logged between 2,000 and 14,000 hours (72.9%). A handful reported having flown fewer than 2,000 hours (1.3%). Prior to joining their current airline, many had previously flown for regional carriers (30.5%), supplemental or cargo carriers (13.2%), the military (52.2%), private companies or charter carriers (34.1%), or for other types of flight operations (33.4%). Fewer had flown for other major (11.3%) or national carriers (7.6%).

Although the survey was designed to address a variety of training-related issues, the focus of this paper is on team training in a simulator. In the aviation domain, this type of training is referred to as Line-Oriented Flight Training (LOFT). It is a critical phase of pilot crew training where crewmembers practice and receive feedback on their technical and CRM skills (Federal Aviation Administration, 1990). LOFT is not required in every pilot crew training event. Therefore, all subsequent analyses are based on those pilots who had received a LOFT, who identified the type of debriefing they received, and who had valid data for each of the covariates (n=6,491 pilots or 64% of the respondents who provided usable data). The demographic composition of the LOFT participants was similar to that of the total respondent pool.

Materials

The materials included a cover letter, a survey, and a follow-up postcard. The cover letter was printed on union letterhead, signed by the union president, and personally addressed to each pilot. The survey addressed a variety of issues, including Line-Oriented Flight Training (LOFT), which is a critical phase of pilot crew training that integrates technical and Crew Resource Management (CRM) skills training in a full-motion simulator (Federal Aviation Administration, 1990). All of the survey questions used a 5-point rating scale with anchors that ranged from 1 (strongly disagree) to 5 (strongly agree).

Measures

The independent variable was the type of LOFT debrief (team debrief with videotape, team debrief without videotape, instructor debrief with videotape, or instructor debrief without videotape). The dependent variable was an 8-item measure of the perceived usefulness of LOFT training. Example items include "The instructor demonstrated a thorough understanding of the LOFT scenario," "The LOFT scenario realistically represented line operations," "LOFT was an effective use of training time," and "The LOFT debrief provided valuable feedback about my performance" (α=.85).

Covariates included the type of training curriculum (initial vs. continuing qualification training), the type of training program (14 CFR Part 121, Single Visit Exemption, Advanced Qualification Program), pilots' perceptions of their instructors' credibility (the instructors' perceived familiarity with and experience in line operations), and pilots' reactions to Crew Resource Management (CRM) training. Because the procedures for developing and administering LOFT are standardized (Federal Aviation Administration, 1990), we did not expect the results to vary as a function of training curriculum or training program. However, we did expect that results to vary as a function of the instructors' perceived credibility, as well as the pilots' reactions to Crew Resource Management (CRM) training.

Procedure

Survey development involved several steps: document review, focus groups, item development, and pre-testing. We began by reviewing the aviation psychology, training, and human factors research literatures to identify important issues in pilot crew training and debriefing. We then conducted a series of focus groups with pilots to better understand which issues personally affect them. Based on our findings, we developed a series of survey items. We then pre-tested the items with four samples of airline pilots. After each pre-test, we revised the instrument as necessary. Pre-testing continued until no more substantive changes were required to either the survey content or layout.

Membership lists from 3 major pilot unions served as our sampling frame. We used stratified random sampling to select one-half of the pilots (n=30,732) from 24 U.S. carriers that was representative in terms of airline, aircraft type, and pilot rank. The survey was administered via U.S. mail to the pilots' home addresses. To increase the response rate, we sent a follow-up postcard 2 weeks after the initial survey mailing.

RESULTS

Prior to conducting any statistical analyses, we performed a series of data screening and checking procedures to ensure the quality of the data. Because the survey was machine scored, we encountered relatively few problems. Next, we calculated the respondent representativenes vis-à-vis the intended population. The differences between the sample and usable response proportions were generally less than 1%, and never exceeded 4%. These differences were extremely small, and obviated the need for weighting the survey results.

The results were tested using hierarchical multiple regression using dummy codes to represent categorical variables such as training program and debrief type (see Table 1). Regression was chosen instead of ANOVA because it allows the researcher to specify the covariates' order of entry in the analysis. Training curriculum (initial vs. continuing qualification training) and training program (14 CFR Part 121, Single Visit Exemption, Advanced Qualification Program) were entered in Step 1. Together, they explained only 1.6% of the variance in pilots' reactions to LOFT. This was not unexpected, because LOFT development and administration are strictly proscribed by the FAA, thereby making LOFT highly consistent across training curricula and programs.

Pilots' perceptions of their instructors' credibility were entered during Step 2. As expected, perceived instructor credibility explained both statistically and practically significant amounts of incremental variance (16.6%) in pilots' reactions to LOFT.

Next, pilots' reactions to CRM training were entered in Step 3. Pilot reactions to CRM training explained both statistically and practically significant amounts of incremental variance (14.4%) in their reactions to LOFT. This was expected, because LOFT provides trainees with the opportunity to practice their CRM skills in a realistic environment.

LOFT debrief type was entered during Step 4. After controlling for the covariates that were entered during steps 1-3, LOFT debrief type was unrelated to pilots' reactions to LOFT. Specifically, the estimated marginal means for the four categories of LOFT debriefing – team debrief with videotape, team debrief without videotape, instructor debrief with videotape, and instructor debrief without videotape – were 3.67, 3.59, 3.67, and 3.63, respectively. These values are virtually identical to the actual cell means that were calculated without correcting for the covariates. Taken together, the data suggest that LOFT debrief type is unrelated to pilots' perceptions concerning the usefulness LOFT. In other words, the four debriefing approaches appear to be equally effective.

DISCUSSION

Despite having a large sample, a highly representative pool of respondents, a high degree of statistical power, and a reliable scale, the data suggest that there is no best way to de-brief team performance. Rather, the four major approaches – team debrief with videotape, team debrief without videotape, instructor debrief with videotape, and instructor debrief without videotape – appear to be equally effective. These results are encouraging. Because all of the debriefing techniques were equally effective, carriers can rest assured in knowing that they do not need to abandon their current debriefing approach in favor one that might be at odds with their organizational culture (Helmreich & Merritt, 1998). Given the recent downturn in the airline industry, this should come as welcome news to training personnel who have already been forced to streamline their budgets.

As mentioned earlier, we were unable to predict *a priori* whether team- or instructor-led debriefs would be perceived as more effective by the pilot crewmembers. Because the instructors received formal training in group facilitation techniques, we initially thought that they might structure the debrief session more effectively than the crews themselves (Dismukes, et al. 2000). On the other hand, because team-led debriefs require the team members to actively participate in the learning process, we thought that the team-led debriefs might lead to greater learning (Butler, 1993). However, given the absence of previous empirical research, we were unable to determine which component – more efficiently structured debrief vs. active participation – would exert a greater effect. Based on the results from the current study, it appears that these two components may cancel each other out.

Contrary to our hypothesis, the participants did not perceive debriefs which incorporate videotape as being more effective than those which did not incorporate videotape. Based on our personal experience, videotape can be an effective tool in post-training debriefings. Because this is only a first look at the effectiveness of videotape in post-training debriefings, we caution the reader to not summarily dismiss videotape as a potential training tool. We believe that these results need to be replicated in other domains before drawing firm conclusions.

As with every study, this one has its limitations. First, we recognize that utility reactions are only one measure of training effectiveness. However, because

trainees are the consumers of training, they are a valuable source of information regarding its effectiveness. Moreover, research by Kraiger and colleagues (1993) has shown that trainee satisfaction is an important outcome of training, while Alliger and colleagues (1997) have shown that utility reactions are positively correlated with learning and training transfer. Therefore, the more that trainees are satisfied with their training and find it useful, the more likely that training will be effective. Second, we recognize that these results need to be replicated in other industries – such as medicine or nuclear power – to assess their generalizability.

In light of these limitations, we refrain from making recommendations for practice. However, we do recommend that researchers continue to explore techniques for assessing and improving the effectiveness of post-training debriefs. Because the debrief is an essential component of team training (Butler, 1993), it plays a major role in ensuring the margin of safety. We believe that this line of research deserves additional attention by both researchers and practitioners.

REFERENCES

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*, 341-358.

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

Dismukes, R. K., McDonnell, L. K., & Jobe, K. K. (2000). Facilitating LOFT debriefings: Instructor Techniques and Crew Participation. *International Journal of Aviation Psychology, 10*, 35-58.

Dismukes, R. K., McDonnell, L. K., Jobe, K. K., & Smith, G. M. (2000). What is facilitation and why use it? In R. K. Dismukes and G. M. Smith (Eds.), *Facilitation and debriefing in aviation training and operations* (pp. 1-12). Aldershot, UK: Ashgate.

Federal Aviation Administration. (1990). *Line operational simulations: Line oriented flight training, special purpose operational training, line operational evaluation* (Advisory Circular 120-35B). Washington, DC: U.S. Department of Transportation.

Gaba, D. M., Howard, S. K., Fish, K. J., Smith, B. E., & Sowb, Y. A. (2001). Simulation-based training in anesthesia crisis resource management (ACRM): A decade of experience. *Simulation & Gaming, 32*, 175-193.

Helmreich, R. L., & Merritt, A. C. (1998). *Culture at work in aviation and medicine: National, organizational, and professional influences*. Aldershot, UK: Ashgate.

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

Table 1.
Dependent Variable: Pilots' Perceptions of Line-Oriented Flight Training (LOFT)

Step	Variable	b	Std. Error	β	Sig.	R^2 Change	Power
1	Training Curriculum	0.02	0.02	0.01	0.38	.016	1.00
	Training Program (Dummy 1)	0.12	0.02	0.09	0.00		
	Training Program (Dummy 2)	-0.09	0.03	-0.06	0.00		
2	Instructors' Understanding of Line Ops	0.15	0.01	0.22	0.00	.166	1.00
	Instructors' Experience in Line Ops	0.13	0.01	0.21	0.00		
3	Pilots' Experiences in CRM Training	0.33	0.01	0.41	0.00	.144	1.00
4	LOFT Debrief Type (Dummy 1)	-0.05	0.04	-0.04	0.19	.002	1.00
	LOFT Debrief Type (Dummy 2)	0.00	0.04	0.00	0.97		
	LOFT Debrief Type (Dummy 3)	-0.09	0.05	-0.03	0.08		

Note 1: Power was computed using N=6491, medium effect size (R^2 = .15), 9 predictor variables, α = .05
Note 2: Actual R^2 = .327 (Adjusted R^2 = .326)