# Title: Performance Measurement for Diagnosing and Debriefing Distributed Command and Control Teams

**Track 6: C2 Assessment Tools & Metrics**

**Joan H. Johnston, Ph.D.**
NAVAIR Orlando Training Systems Division
CODE AIR 4961, 12350 Research Parkway, Orlando, FL 32826
Joan.Johnston@navy.mil

**Daniel Serfaty & Jared T. Freeman, Ph.D.**
Aptima, Inc.
12 Gill St. Ste 1400, Woburn, MA 01801
freeman@aptima.com, serfaty@aptima.com,

# Performance Measurement for Diagnosing and Debriefing Distributed Command and Control Teams

**Joan H. Johnston, Ph.D.**
NAVAIR Orlando Training Systems Division

**Daniel Serfaty & Jared T. Freeman, Ph.D.**
Aptima, Inc.

**Abstract**

Distributed simulation-based training is a challenging and resource intensive effort that the services must perform on a day-to-day basis--within services, joint service, and in joint coalition forces. Research gains have been made that demonstrate advanced training assessment and diagnosis, however, there remains a need for research and development of automated capabilities for assessing team performance, diagnosing root causes of team failure, and debriefing the results across distributed platforms in exercises. A major thrust of this research needs to be the development of measurement technologies for distributed team training that advances the efficiency and speed of delivery to the war fighter in training. In this paper we draw from the recent research on adaptive team architectures for command and control, team decision-making and training, and examples of team measurement tools to derive assessment requirements that support the instructional processes required for distributed team training debrief and after action review.

## Introduction

Distributed simulation-based training is a challenging and resource intensive effort that the military services must perform on a day-to-day basis--within services, joint service, and in coalition forces. The services have strived to replicate the dynamic command and control environment to achieve the "train like we fight" objective (Acton & Stevens, 2001). Major investments have been made in technologies that: simulate real world platforms and related sensors; bolster connectivity across distributed teams through advanced networks; and enable interoperability demonstrations within and across services, and with coalition forces (Clark et al., 2001). As a result, distributed simulation-based exercises significantly increase the dynamic nature of training and the job of conducting them. For example, the impact of one team's adaptation to a mission can quickly impact the performance of other teams, initiating a ripple effect on overall mission outcomes. Instructors typically respond to team adaptation by modifying the scenario to ensure it remains engaging. The changes pose a major challenge for reconstruction of "what happened" during a scenario. Any "course corrections" that were made by teams and instructors must be addressed in the debrief/After Action Review (AAR) process. Modification and tracking of exercise changes in a distributed training environment rapidly becomes unmanageable without many people in the loop monitoring the exercise (Oser et al., 1997). Some training programs have invested in automated reconstruction of simulated platforms for replay and review to reduce the instructor workload, increase fidelity of debrief/AAR information, and reduce training costs. For example, the Powerstripes system, under development by the US Army, and the Navy Battle Force Tactical Training (BFTT) system enable rapid production of scenario replay and snapshots, illustrated maps for simulated track history over time, and charts and graphs of specific results of scenario events (AcuSoft, 2001; McGaughey, 2001).

Overall, however, the time needed to sift through simulation data collected from exercises can take several hours to many weeks, using up precious resources. This also creates a major stumbling block for training a team of experts to learn how to be an expert team through rapid debrief and AAR (Oser et al., 1995). The services have recognized the human performance technologies gap by drafting simulation master plans that include specifications for automating performance assessment, diagnosis, and debrief/AAR. In addition, the 6.4 research programs for NAVAIR Advanced Warfare Training Development and NAVSEA BFTT are sponsoring an Office of Naval Research (ONR) Future Naval Capabilities program for automating the process of assessment and diagnosis of root causes of team performance, and rapidly debriefing and conducting AAR across distributed simulation-based exercises.

As a core requirement for such systems, a principled strategy is needed for developing and using team performance assessment to address the challenge of partitioning large numbers of participants into teams that can conduct effective debrief/AAR, and should do so because their interactions had an impact on team performance. In this paper we propose a team measurement strategy that is based on a model of organizational structure and behavior, and provide an illustrative example of how the strategy would be implemented for partitioning debrief/AAR. Finally, we identify recommendations for future research.

**Model of Organizational Behavior**

Adaptability is a critical team skill in the highly dynamic, uncertain environments that are experienced by military teams. It is even more important in distributed teams given the highly changing nature of teamwork and team membership. Research has shown that expert teams adapt to stressors in the tactical environment by changing the processes by which they communicate and coordinate, and – in some cases – by restructuring themselves (Serfaty, 1999; Serfaty & Entin, 1997; Serfaty et al., 1998). In a manner analogous to individual adaptation, yet enriched by the additional communication and coordination a team has at its disposal, teams are able to modify their work strategy to adapt to the demands and constraints of their environments. Adaptation can take the form of process adaptation, such as alternative decision-making procedures and coordination strategies, or structure adaptation, such as dynamic reconfiguration. The research findings from adaptive architectures for command and control and team training (Macmillan et al., in press) have inspired defining team member interactions based on two types of information. The first type is the static or official structure of the organization. Well-defined teams are characterized by the crisp structural constraints of decision hierarchy and task responsibilities, they include: who is subordinate to whom through command authority; who has control of specific assets; who has access to specific information resources; communication rules for passing information; and distributed expertise. During mission execution, team members command, communicate, control assets, access information, and exercise expertise, which can be dynamic and context dependent. Therefore, the second type of information is the dynamic or actual behaviors of the organization. To be adaptive in a dynamic, changing military environment, teams modify their processes through changing assets and communications. The dynamic aspect of team communication is defined by the content and timing of communication exchanges among team members during mission execution.

Figure 1 presents a model of the static and dynamic features of a 3-person team. The static or official communication transactions are represented as thin black lines that link team members #1 and #2, and team members #1 and #3. The actual communication transactions are represented as wide white lines. Team members communicate with others to conform to the organization's defined constraints by taking some of the opportunities afforded by the architecture, while foregoing other opportunities provided by it. In addition, they may depart from the prescribed architecture and violate some constraints. These events are represented in Figure 1 by large arrows indicating (1A) "legal" communication transactions between team members #1 and #2, and (2) "illegal" communication transactions between team members #2 and #3. The lack of legal communication transactions between team members #1 and #3 is indicated by the absence of a large arrow.

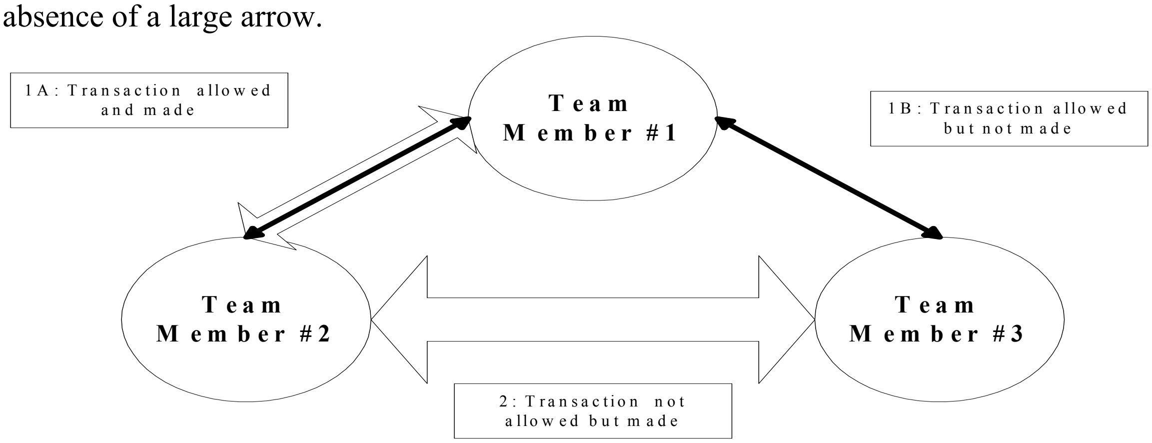


Figure 1: Three-person team with static (official) communications transactions represented by thin black lines and dynamic (actual) communication transactions represented by wide white lines.

Considering both the static architecture and dynamic performance of organizations enables representing team performance normatively and descriptively, making decisions concerning who should debrief with whom, and setting priorities in answering the related questions of when and where to debrief distributed teams. Specifically, the additional challenge of distributed debriefing/AAR is to identify individuals within the organization whose interactions influenced team performance, and to bring those individuals together to critique the interactions, and recommend improvements in them. The challenge of partitioning the organization is not trivial. The partitions may not correspond to existing boundaries between teams or functions. They may be quite novel. There may be more interacting groups than there are venues (e.g., videoconferencing facilities) and time to debrief them. Thus, some practical balance must be struck between debriefing every interacting sub-team about every failure, and the common solution, which is to debrief everyone in intact groups (e.g., each flight element) about everything. This solution, which we have witnessed after many military exercises, variously overwhelms participants in detail that is irrelevant to most of them with respect to any one topic (e.g., "Darkstar failed to respond to a request for information from Eagle 1 at time 08:10:30"), and burdens them with diagnosing their role in organizational failures that are too abstractly described (e.g., "The battlegroup was not well synchronized in day four operations").

Therefore, we propose that partitioning be determined by assessing the transactions made among team members in response to the organizational structure. Behaviors that conform to architectural constraints (1A) may warrant feedback that reinforces the behavior, particularly if that behavior is spotty. Behaviors that fail to exploit the allowances of the architecture (1B) may warrant instruction that reminds team members of this aspect of the architecture (e.g., the available communication channel or control over some asset). It may also prompt the organization to reconsider this aspect of the structure, if it adds cost or complexity without perceived value to the team. Finally, behaviors that violate architectural designs (2) may warrant corrective instruction or organizational redesign. The volume and criticality of these behaviors (indicated by the breadth of lines in Figure 1) can guide instructors to prioritize feedback so that they give the most weight to most frequent or critical errors – whether of commission (1B) or omission (2). The identity of the individuals joined by the arcs indicates who should be engaged in debriefs. In this case, the large volume of illegal communications between team members #2 and #3 suggests that they be brought together in a debrief/AAR. In a realistically complex multi-team organization as depicted in Figure 2, larger groups would be assembled for debriefs because they might be involved in violations of the static structure of the organization, and might require the attendance of some individuals (such as a commander) in each debrief/AAR.

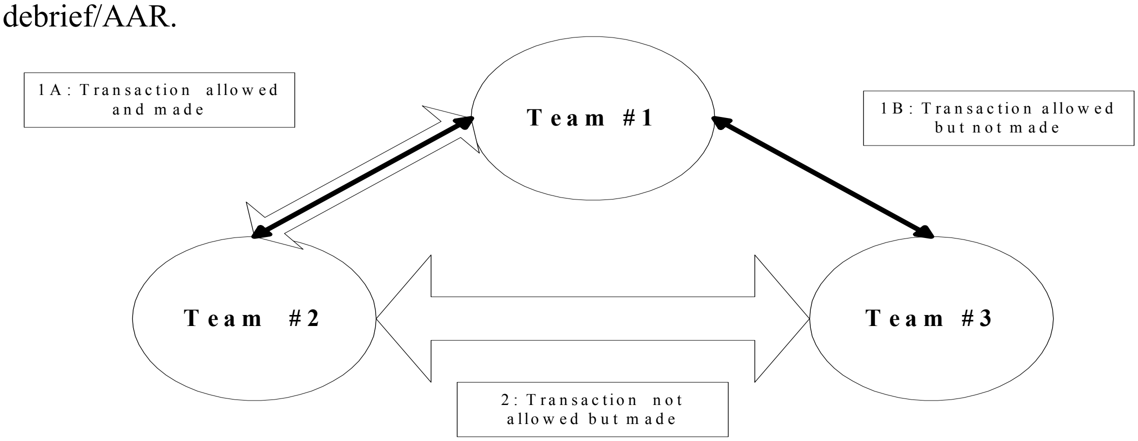


Figure 2: Three distributed teams with static (official) communications transactions represented by thin black lines and dynamic (actual) communication transactions represented by wide white lines.

## Illustrative Example

The approach discussed in the previous section is appealing because it enables prescription of measurement strategies that support partitioning decisions. Team performance measurement consists of local measures of performance (e.g., the timeliness and accuracy of one team member's actions) and global measures (e.g., the synchronization of actions by many teams, as well as the effects of these actions on the enemy) (Alberts, et al., 2001). Freeman & Serfaty (2002) present a comprehensive table of team cognition and performance measures developed over the past decade that supports the idea of using a combination of assessments to address the types of transactions among team members and across teams presented in Figures 1 and 2. In particular, the numerous process measures noted in the table are designed to enable recording of the transactions of team members, and produce diagnostic information about the dynamic interchange. To illustrate, we describe how a process measure for team tactical decision-making could be used to assess dynamic transactions: behaviors that conform to architectural constraints, but are intermittent; and behaviors that violate architectural designs. We then describe how these measures support decisions about partitioning a debrief/AAR for two teams.

***Air Warfare Team Performance Index (ATPI).*** A paper-based and hand-held electronic ATPI was developed through ONR sponsored research for TADMUS and Manning and Affordability (Lyons & Allen, 2000; Paris et al., 2000; Pharmer et al., 2000). It is designed to assess and diagnose tactical decision-making in a ship's air defense warfare (ADW) team during execution of a combat team simulation scenario. The ADW team performs the detect-to-engage (DTE) sequence on aircraft in the vicinity of the battle group, and is comprised of several sub-teams reporting to the Tactical Action Officer and Commanding Officer. The DTE sequence is an established static architecture for the team; team members are required to perform the task according to Navy doctrine. The objective is to evaluate and monitor critical tracks of interest, and determine whether communications with them are warranted based on rules of engagement. Tracks determined to be hostile may be engaged if they meet the rules of engagement. Team members use dynamic transactions through electronic and verbal channels to manage and share the workload. The ATPI allows subject matter experts to note whether a DTE action was taken, and the team member that performed the action. A key feature is that the DTE sequence is categorized into four dimensions identified as required competencies for team decision-making. For ADW teams, the identification competency involves timely and accurate detection, reporting, and monitoring of potentially important aircraft. Elaboration involves timely and proper categorization of the track as hostile, friendly, or unknown, and giving it a priority for further actions. Planning and Execution competencies involve the team making timely and accurate plans for further actions, and then executing those actions in a timely and accurate manner, if approved (e.g., queries, final warnings, Friend or Foe challenges, and Engaging the track with weapons).

Tables 1 and 2 present a simple example of an ATPI developed for the ONR Tactical Decision Making Under Stress program (Paris et al., 2000). Table 1 highlights the identification and elaboration phase for the first critical event that begins at one minute and runs until five minutes into the scenario. Three aircraft appear in a four-minute time window. Table 2 highlights the planning and execution phase for the same aircraft during the second event in the scenario, beginning at six minutes and ending at 15 minutes. Active and diverse DTE sequences on several tracks per event are required. Subject matter experts establish the windows of opportunity for each task on each track. Evaluators note whether teams exceeded the window of opportunity for a task, made an error in performing the task, or did not perform it at all. Poor results can range from increasing workload among team members to engaging tracks that were not violating rules of engagement.

Table 1: Example of an Air Warfare Team Performance Index for Identification/Elaboration Actions on Detect-to-Engage. Evaluators Note On-Time, Late, And Incorrect Actions.

| Event 1: Begins 1 minute after scenario start and ends at 5 minutes | Made Detection<br><br>Minutes:Seconds | Made Platform Identification | Made Platform Threat Id And Priority |
|---|---|---|---|
| Commercial Aircraft #1 Wanders Off of Comair Route | 1:30 | 2:30 | 3:00 |
| Potentially Hostile Helicopter | 2:30 | 4:30 | 5:00 |
| Potentially Hostile P3 Aircraft | 3:30 | 4:30 | 5:00 |

Table 2: Example of an Air Warfare Team Performance Index for Planning and Execution Actions on Detect-to-Engage. Evaluators Note On-Time, Late, And Incorrect Actions.

| Event 2: Begins 5 min after scenario start and ends at 15 minutes | Plan/Execute Query<br><br>Min:Sec | Plan/Execute Final Warning | Plan/Execute Illuminate | Plan/Execute Cover With Weapons |
|---|---|---|---|---|
| Commair #1 | 6:00/7:00 | N/a | N/a | N/a |
| Helicopter | 5:30/5:30 | 7:30/7:30 | 8:00/8:30 | 8:00/8:30 |
| Potentially Hostile P3 Aircraft | 6:00/7:00 | 7:30/8:30 | 8:30/8:30 | 14:30/15:00 |

DTE performance results can be summarized for each dimension across each scenario event. For example, Table 3 presents ATPI results for hypothetical Team A. Team A performed just 60 percent of the identification and elaboration processes across each of the events in the scenario. The team behaviors conformed to architectural constraints, but they were spotty. Team A's performance on planning and execution dropped from 40 percent to no actions in the third critical event. Table 4 presents ATPI results for hypothetical Team B. In all three events, they performed less than half the required identification and elaboration actions by the third event (violating architectural constraints), but they achieved almost all planning and execution actions. The two examples illustrate how focusing on the results of decision making dimensions can increase the efficiency of meeting instructional needs by specifying how debrief/AAR should be partitioned. Debrief/AAR for Team A should focus on critiquing their planning and execution processes, whereas, Team B should focus on critiquing and improving their identification and elaboration processes. In the next section, we describe how the ATPI could be used in conjunction with a teamwork process measure to address the third transaction type-behaviors that fail to exploit the allowances of the architecture-to further refine the partitioning process.

Table 3: Percent of Scenario Event Actions Performed by Team A for Each Decision Making Dimension.

| Scenario | Identification | Elaboration | Planning | Execution |
|---|---|---|---|---|
| Event 1 | 90 | 90 | 40 | 40 |
| Event 2 | 80 | 80 | 20 | 20 |
| Event 3 | 80 | 80 | 0 | 0 |

Table 4: Percent of Scenario Event Actions Performed by Team B for Each Decision Making Dimension.

| Scenario | Identification | Elaboration | Planning | Execution |
|---|---|---|---|---|
| Event 1 | 60 | 40 | 99 | 99 |
| Event 2 | 60 | 40 | 99 | 99 |
| Event 3 | 50 | 30 | 99 | 99 |

***Anticipation ratio***. Research has shown that good teams can adapt their behaviors to changing, stressful conditions. For example, good ADW teams are able to adapt to increased stressors and workload because they anticipate the need for more information in these conditions (Serfaty et al., 1998). By sending information to each other and to higher authority without being asked, fewer requests for information are required, especially from higher authority, thus increasing the speed and efficiency of the team. The anticipation ratio tool enables diagnosis of the dynamic communication exchanges among team members (Serfaty et al., 1998). The tool is designed so that subject matter experts categorize and record team member exchanges. Categories include the type of communication (e.g., information exchange, situation updates, supporting behavior, error correction, and feedback), and the direction of the communication in the hierarchy (e.g., team member to team member and team member to higher authority). An important anticipation ratio is calculated by dividing the number of team member communications to higher authority by the number of communications from higher authority to team members. A good anticipation ratio indicates high levels of team adaptability. A poor anticipation ratio means higher authorities are increasing their workload by increasing their requests for critical information (“pulling information”) from team members.

Measurement of a team's dynamic communication patterns could help to assess the communications violations that failed to exploit the allowances of the organization's architecture. For example, the anticipation ratio for Team A could reveal that higher authority was asking for information, but team members were not providing it. In contrast, Team B's identification and elaboration problems could be the result of improper team member to team member communication patterns. Both situations warrant reminding team members of the requirements of organizational architecture, but the anticipation ratio helps to pinpoint the type of critiquing and feedback needed. During debrief/AAR Team A would focus on how to use available communication channels for pushing information to higher authority, whereas Team B would focus on using the proper sequence of communications among team members.

## Summary and Thoughts for Future Research

In this paper we presented an organizational model and a rationale for combining team performance assessments to derive a principled approach to decisions for distributed debriefing/AAR. It is important to note that we provided exemplars of small team membership. Over the next several years, distributed simulation-based training plans are in place to develop network connections to simulations across training communities. A number of analyses have already been conducted that identify the challenges of large-scale exercises (Bergondy et al., 1999; Bergondy & Salas, 1999; Neville et al., 2001). The demand for additional instructor and scenario controller workload will increase exponentially without automated strategies for measurement, diagnosis, and debrief/AAR (Oser et al., 1997). The approach we described in this paper is useful in small team training exercises with short 30-minute scenarios, but the scale of application does not readily support distributed simulation based training that can take hours to several days without requiring many people-in-the-loop. Therefore, implications for future research include developing:

- Automatic recording of high-level operator performance actions and deriving strategies for integrating them with voice communication patterns,
- Computational models to assess team performance patterns over the course of scenario implementation,
- Team cognition assessment that enables better diagnosis of team processes, and
- After-action review aids that automatically incorporate critiquing dialog based on the focus of the diagnosed team performance.

## Acknowledgements

Aspects of this work are funded by the Office of Naval Research. The opinions expressed here are the authors' and do not necessarily reflect the views of the U.S. Navy or the Department of Defense.

## References

Acton, B.J., & Stevens, B.J. (2001). Objective based training and the battle force tactical training system; focusing our fleet training processes. *Proceedings of the 23rd Annual Interservice/Industry Training, Simulation and Education Conference,* Orlando, FL, 1418-1428.

AcuSoft (2001). Powerstripes Users Guide Version 1.0. Orlando, FL: Acusoft.

Alberts, D.S., Garstka, J.J., Hayes, R.E., and Signori, D.A. (2001). *Understanding Information Age Warfare.* www.dodccrp.org.

Bergondy, M. L., Fowlkes, J. E., Merket, D. C., & Milham, L. M. (1999). Analysis of deployed training requirements in the F/A-18 community. *Proceedings of the 21st Annual Interservice/Industry Training Systems and Education Conference,* Orlando, FL, 640-648.

Bergondy, M. L., & Salas, E. (1999). Is it old wine in a new bottle?: The issues and challenges of distributed team training. *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society*, Houston, TX, 1143-1145.

Clark, P., Ryan, P., Zalcman, L., O'Neal, M., Brewer,J., & Beasley, D. (2001). Building towards coalition warfighter training. *Proceedings of the 23nd Annual Interservice/Industry Training, Simulation and Education Conference*, Orlando, FL, 1233-1242.

Freeman, J.T., & Serfaty, D. (2002). Team collaboration for command and control: A critical thinking model. In *Proceedings of the Command and Control Research and Technology Symposium*, Monterrey, CA. www.dodccrp.org.

Lyons, D., & Allen, R. C. (2000). Mobile aid for training and evaluation (MATE): A hand-held, configurable set of team performance measurement tools. *Proceedings of the 22nd Annual Interservice/Industry Training, Simulation and Education Conference,* Orlando, FL, 661-671.

Macmillan, J., Entin, E. E., & Serfaty, D. (in press). Communication overhead: The hidden cost of team cognition. To appear in E. Salas and S. M. Fiore (Eds.), *Team cognition: Process and performance at the inter- and intra-individual level.* Washington, DC: American Psychological Association Press.

McGaughey S. (2001). Battle force tactical training after action review. *Proceedings of the 23nd Annual Interservice/Industry Training, Simulation and Education Conference*, Orlando, FL, 1194-1202.

Neville, K., Fowlkes, J., Milham, L., Merket, D. C., Bergondy, M. L, Walwanis, M., & Strini, T. (2001). Training team integration in a large, distributed, tactical team: A cognitive approach. *Proceedings of the 23rd Annual Interservice/Industry Training, Simulation and Education Conference,* Orlando, FL, 1035-1044.

Oser, R. L., Cannon-Bowers, J. A., Dwyer, D. J., & Salas, E. (1997). Establishing a learning environment for JSIMS: Challenges and considerations. *Proceedings of the 19th Annual Interservice/Industry Training, Simulation and Education Conference,* Orlando, FL, 144-153.

Oser, R. L., Dwyer, D. J., & Fowlkes, J. E. (1995). Team performance in multi-service distributed interactive simulation exercises: Initial results. *Proceedings of the 17th Annual Interservice/Industry Training, Simulation and Education Conference,* Arlington, VA, 163-171.

Paris, C. R., Johnston, J. H., & Reeves, D. (2000). A schema-based approach to measuring team decision-making in a Navy combat information center. In C. McCann & R. Pigeau (Eds.), *The human in command: Exploring the Modern Military Experience* (pp. 263-278). New York: Kluwer Academic/Plenum Publishers.

Pharmer, J., Hildebrand, G. & Campbell, G. (2000). Measuring the impact of human-centric design on human performance in a combat information center. *Presented at* *the 45th Annual Meeting of SPIE International Symposium on Optical Science and Technology- Integrated Command Environments Session (4126),* San Diego, CA.

Serfaty, D. (1999). Lessons Learned From The A2C2 Experience. In *Proceedings of the 1999 Command and Control Research and Technology Symposium*, Newport, RI. www.dodccrp.org.

Serfaty D. & Entin, E. (1997). Team Adaptation and Coordination Training. In R. Flin, E. Salas, M. Strub, & L. Martin, Eds., *Decision Making Under Stress: Emerging Themes and Applications* (pp. 170-184). VT: Ashgate.

Serfaty, D., Entin, E.E., and Johnston, J.H. (1998). Team coordination training. In J.A. Cannon-Bowers and E. Salas (Eds.), *Making Decisions Under Stress: Implications for Training and Simulation* (pp. 221-245). Washington, DC: American Psychological Association.