Team Skills Assessment: A Test and Evaluation Component for Emerging Weapon Systems

Jennifer E. Fowlkes[1], Daniel J. Dwyer[2], Laura M. Milham[1], John J. Burns[3], & Linda G. Pierce[4]

[1]Team Performance Laboratory, University of Central Florida, Orlando, FL
[2]Naval Air Warfare Center Training Systems Division, Orlando, FL
[3]Sonalysts, Inc., Orlando, FL
[4]Army Research Laboratory, Human Research and Engineering Directorate, Fort Sill Field Element, Fort Sill, OK

ABSTRACT

The purpose of this paper is to report on the development of a team performance measurement system capable of supporting the Army's recent initiatives for testing systems and doctrine in synthetic environments. One mechanism for evaluating emerging Army systems and doctrine is the Advanced Warfighting Experiment in which a synthetic theater of war (STOW) environment is created. STOW environments provide a key function in the test and evaluation (T&E) process, yet they pose challenges for effective team performance measurement, an important component for providing feedback to the development process. Challenges include lack of control of task content, requirement to capture performance in near real-time, and the use of multiple observers who must be kept cognizant of ground truth in a complex, dynamic environment so that they can legitimately assess team performance. To address the challenges outlined above, the TRACTs (Tactically Relevant Assessment of Combat Teams) performance measurement system was developed. TRACTs capitalizes on recent work performed by the Navy and Army in the area of team performance measurement. It captures both task-based and team-based performance and is implemented on a computer-based, hand-held data collection device. TRACTs was recently used to evaluate team and task performance during an assessment of employment concepts for the Crusader system. Crusader is a revolutionary weapon system which will be fielded by the Army in 2005. This paper addresses TRACTs design, TRACTs data obtained during the Crusader experiment, and discussions of how these data can be used to augment the T&E process.

AUTHOR BIOGRAPHIES

Dr. Jennifer E. Fowlkes is Research Psychologist with the Team Performance Laboratory at the University of Central Florida. She has over twelve years of experience in areas of human factors and training which includes team training and performance, training effectiveness evaluations, simulator sickness research, and performance test battery development. Most recently, her research has focused on measuring team performance in distributed training environments. Dr. Fowlkes holds a Ph.D. in Experimental Psychology from the University of Georgia.

Dr. Daniel J. Dwyer is a Senior Research Psychologist at the Naval Air Warfare Center Training Systems Division where he has 20 years experience in military training research, development, and acquisition support. His MS degree from the University of Central Florida and his Ph.D. degree from the University of South Florida are both in Industrial/Organizational Psychology. Dr. Dwyer's most current research efforts are in the areas of multi-service training, team performance measurement, and distributed training.

Laura Martin Milham is a Graduate Research Assistant at the Team Performance Laboratory at the University of Central Florida where she has worked for 3 years. She is currently pursuing her Ph.D. in Human Factors Psychology at the University of Central Florida. She has worked on team situation assessment and computer based training. Her major research interests are the use of cognitive strategies to enhance training effectiveness.

Dr. John J. Burns is a senior scientist at Sonalysts, Inc. His work has focused on the development of shipboard training systems, with an emphasis on performance measurement methodologies and technologies. Most recently, Dr. Burns has been involved with the development of the Shipboard Mobile Aid for Training and Evaluation (ShipMATE), a prototype handheld PC designed to support instructors in all aspects of training. Dr. Burns received his Ph.D. in psychology from the University of Massachusetts in Amherst and has 6 years of experience in team training research and development.

Dr. Linda G. Pierce is Chief of the Army Research Laboratory, Human Research and Engineering Directorate, Fort Sill Field Element. She supports the U.S. Army Field Artillery School and the Depth and Simultaneous Attack Battle Lab as Chief of the Simulations Branch. Her programs include using simulations for training and system acquisition; assessing team performance; and defining decision making processes in highly automated environments. She holds a Ph.D. in Industrial/Organizational Psychology from Texas Tech University.

Team Skills Assessment: A Test and Evaluation Component for Emerging Weapon Systems

Jennifer E. Fowlkes[1], Daniel J. Dwyer[2], Laura M. Milham[1], John J. Burns[3], & Linda G. Pierce[4]

[1]Team Performance Laboratory, University of Central Florida, Orlando, FL
[2]Naval Air Warfare Center Training Systems Division, Orlando, FL
[3]Sonalysts, Inc., Orlando, FL
[4]Army Research Laboratory, Human Research and Engineering Directorate, Fort Sill Field Element, Fort Sill, OK

OBJECTIVES

The purpose of this paper is to describe a performance measurement approach, referred to as TRACTs (Tactically Relevant Assessment of Combat Teams). TRACTs was developed to support the Army's recent initiatives for testing systems and doctrine in synthetic theater of war (STOW) environments. Such initiatives are due in large part to the requirement to test emerging warfighting systems in realistic battlefield scenarios. In this context, team and individual performance measurement is needed in addition to outcome measures to provide realistic assessments of system capabilities. STOW environments provide a key function for test and evaluation (T&E), yet they pose significant challenges for effective team and individual performance measurement. Our objective for TRACTs was to develop a measurement system responsive to the challenges inherent to these complex testing environments.

TRACTs capitalizes on recent work performed by the Navy and Army in the area of team performance measurement. Its specific development was a result of two Army efforts. The first focused on an assessment of simulation systems that could be used to train division artillery staff (McCluskey, Fowlkes, Pierce, & Dwyer, 1998). For this effort, TRACTs was developed primarily to assess taskwork performance (i.e., to determine whether individual and collective tasks were performed in accordance with standard operating procedures). It was learned that TRACTs could be used to provide highly detailed assessments of performance strengths and weaknesses of operational tasks related to battlefield functions. However, it was also learned that assessment of teamwork functions was needed to complete the diagnostic picture. Thus in the second effort reported herein, TRACTs was utilized to capture both teamwork and taskwork performance measures in a STOW exercise designed to develop operational concepts for the Crusader weapon system. Prior to describing its implementation for the Crusader work, we provide a description of the approach underlying TRACTs design.

APPROACH

We began the development of TRACTs by identifying measurement requirements related to STOW testing environments. These are discussed below.

STOW-Related Measurement Requirements

Assess Performance of Teams, Subteams & Individuals. STOW training and testing environments may involve many, even hundreds, of system operators. Thus, one of the most apparent requirements for successful performance measurement in STOW testing environments is that diagnostic measurement systems will have to assess the performance of both individuals and teams. As Lane and Alluisi (1992) noted:

> The players in this simulated battlefield environment are not only the weapon system operators, but also the commanders, staffs, logisticians, support units, intelligence personnel, and decision makers at all levels—in short, all the combat, combat support, and combat-service support elements assigned to the battle force and its support. (p. 23)

Control Task Content. A key characteristic of any performance measurement system is the extent to which what is being practiced or trained can be determined. In STOW environments, there are a number of factors that adversely affect the ability to specify task content.

<u>Nature of warfighting skills</u>. STOW environments are created specifically to provide opportunities to utilize "warfighting skills" as opposed to "basic skills" or the skills required to operate individual platforms. By their nature, warfighting skills

are multidimensional with the consequence that the task content to be measured will be less certain.

Multiple goals. The problem of uncertain task content is further compounded by the multiple goals set for warfighting experiments, where often the intent is to impact doctrine, training, leader development, organizations, materiel, and soldiers (i.e., "DTLOMS"). The problem is that factors conducive to training, for example, may be at odds with those that are needed to develop doctrine. The cross-purposes and resultant lack of control will adversely impact task content.

Free-play nature of battle simulations. Finally, the free play nature of battle simulations presents one of the most significant threats to task content. The event flow is largely determined by the give-and-take, real-time interactions of players in the exercise. While free play exercises are probably more representative of the actual battlefield than scripted exercises, they are difficult to evaluate because task content is largely left to chance (Fowlkes, Lane, Dwyer, Willis, & Oser, 1995).

Facilitate Observation. An important component of measurement is ensuring that opportunities are provided for data collectors to acquire performance-related information. In STOW environments, there are a number of hurdles that may interfere with observation.

- Much of the information passed between large, distributed Army teams is digital in nature (i.e., passed over local and wide area networks). This can be expected to increase as the Army continues to incorporate information technologies. The consequence is that it is difficult for data collectors to observe information exchange among participants.
- Data collector workload may be high in STOW environments; they must observe the performance of many individuals, assess the integrity of the network, keep apprised of the battle and ferret out what information has been passed and to whom. Besides the burden of performance assessment, other tasks may compete with data collection efforts (Dwyer, Fowlkes, Oser, Salas, & Lane, 1997).
- The problems outlined above are likely to be exacerbated because a team of data collectors is required to adequately observe and judge performance in STOW environments. Each will have a different physical viewpoint, training, and responsibility. Moreover, for joint operations, different data collectors will be required from each of the services represented (Fowlkes et al., 1995). Thus for performance measurement, the input from multiple observers, who may have different training and backgrounds, will likely be required.

Near Real-Time Measurement. A final requirement in STOW-based testing environments will be capturing team performances in real-time or in near real-time. For the most part, team performance measurement requires a team of humans to observe a team of humans—it cannot yet be captured through automated, on-line measurement (Vreuls & Obermayer, 1985). In many simulation systems there is not a replay capability that incorporates team performance so that data collection can take place after an exercise. Thus, there is generally only one chance to capture relevant performances.

Besides being necessary, near real-time measurement is important to pursue because it may provide other benefits. Near real-time data capturing and reduction/analysis can be used to support data quality control, decisions to move from one experimental phase to the next during an exercise (Vreuls & Obermayer, 1985), and the identification of performance patterns or trends. Another highly desirable benefit is the ability to provide timely post-exercise feedback to participants.

TRACTs Measurement System Components

The key components of the TRACTs performance measurement approach are summarized in Table 1. The left hand column identifies the STOW-related measurement requirements discussed above. The remaining columns show the elements incorporated into TRACTs specifically to address them: process performance measurement, event-based measurement, and computer-based data collection.

Process Performance Measurement. To assess the performance of teams and individuals, TRACTs incorporates process performance measurement as opposed to outcome measurement. Outcome measures, which characterize end results, are usually given a high priority for data collection in STOW environments because they are objective, possess high face validity, and are usually easily obtained. However, as "end results" are affected by many variables—controlled and uncontrolled—they typically suffer from lack of reliability (Lane, 1986; Lane & Kennedy, 1994) and thus have limited utility in deriving answers to questions addressed in warfighting experiments. Also, while outcome measures can signal a problem (e.g., a decrease in the number of fire missions performed), they have restricted usefulness in

Table 1. STOW-related measurement requirements and TRACTs measurement system components

Measurement Requirement	Components of TRACTs Measurement System		
	Process Performance Measurement	Event-based Measurement	Computer-based Data Collection
Assess Performance of Teams, Subteams & Individuals	**X**	**X**	
Control Task Content		**X**	**X**
Facilitate Observation		**X**	**X**
Near Real-Time Measurement		**X**	**X**

diagnosing the cause of the problem (e.g., Cannon-Bowers & Salas, 1997; Dwyer et al., 1997).

In contrast, process measures characterize how teams and individuals perform tasks. They can be extremely diagnostic of performance deficiencies if the observation of team processes is driven by *a priori* constructs and expectations.

Taskwork assesssment using TRACTs entailed examining discrete, observable behaviors involved in task execution. Teamwork measurement centered around the assessment of team functions identified in Table 2. These had been developed for the assessment of Navy shipboard teams (Smith-Jentsch, Zeisig, Acton, & McPherson, 1998) and were modified slightly to accommodate Army fire support operations.

Event-based Measurement. As shown in Table 1, TRACTs incorporates "event-based measurement" to address many of the measurement difficulties inherent to STOW environments. Event-based techniques create measurement opportunities within an exercise by systematically identifying and introducing trigger events that provide known opportunities to observe specific behaviors of interest (Fowlkes, Dwyer, Oser & Salas, 1998; Fowlkes, Lane, Salas, Franz, & Oser, 1994; Pruitt, Burns, Wetteland, & Demestre, 1997; Salas & Cannon-Bowers, 1997; Smith-Jentsch, Johnston, & Payne, 1998). The use of trigger events ensures that observation, and the resulting information obtained, is directly related to targeted issues or warfighting competencies. It focuses observation on those behaviors or processes that are responses to the events.

<u>Assess performance of teams, subteams, and individuals</u>. To assess the performance of artillery teams with TRACTs, battlefield events were identified for critical command cells or elements. These events provided opportunities for teams to perform tasks related to fire support. Table 3 shows both the teams or command cells in which assessments were made and the battlefield events targeted for the Crusader effort.

Using an event-based approach in a scripted exercise, the battlefield events would be introduced at preplanned times that would be known to the data collectors, making it easy for data collectors to assess responses to them. In a free play exercise, such as in Crusader effort, these events will occur naturally as a result of the interactions between participants and simulated entities, however their timing cannot be known *a priori*. In this case, data collectors are instructed to sample some predetermined number of each type of event. Thus in this case, the task for the data collector is more difficult, and he or she must remain vigilant to identify the occurrence of key events. In either case, once a battle event occurs, using the TRACTs methodology, data collectors complete a checklist that lists the acceptable responses to the event. These are scored as either present or absent. Acceptable responses are determined through the application of standard operating procedures and incorporating subject matter expert input. Performance scores represent the percent of acceptable subtasks performed. Figure 1 provides an example of a checklist that would be used.

<u>Control task content</u>. Deliberately introducing realistic battlefield events as well as sampling events is

Table 2. Teamwork functions and associated behaviors

Information Exchange	**Communication**
• Analysis • Develop shared picture • Plot/log information • Utilize available sources of information • Pass information to the right persons • Provide big picture updates	• Proper phraseology • Completeness of standard reports • Brevity/avoiding excess chatter • Clarity/avoiding inaudible comms • Correct medium
Supporting Behavior • Cross check information • Monitor & correct errors • Provide & request backup or assistance to balance workload	**Initiative** • Provide guidance or suggestions • State priorities

Table 3. Battlefield events driving the development of checklists and command cells in which performance was assessed

Command Cells	Battlefield Events
• Task Force Fire Support Element (TF FSE) • Brigade Fire Support Element (BDE FSE) • Battalion Fire Direction Center (BN FDC) • Battalion Operations and Intelligence (BN O&I) • Battery Operations Center (BOC) • Platoon Operations Center (POC) • Crusader Workstations	• Intelligence summaries (INTSUM) • Intelligence (Intel) • Situation map update (SITMAP) • Forward line of own troops (FLOT) updates • Coordinated fire line (CFL) updates • Unit movement • Clear fires • Target development • Fire missions • Radar coverage • Rearm • Survivability move

a method that allows measurement to be specifically focused on targeted competencies. Controlling the number of events introduced or sampled allows control over the number of observations obtained, improving the reliability of the resulting scores.

Events can also be targeted to allow deliberate assessment of teamwork skills. In the checklist shown in Figure 1, each of the acceptable responses listed can be coded into one of the teamwork behaviors shown in Table 2. Indeed, this is the approach used for TRACTs. Every behavioral observation shown on the checklists can be linked to a battlefield function (taskwork) and most can also be linked to one of the teamwork functions/team behaviors shown in Table 2. Thus, the same data can be analyzed by teamwork and by taskwork.

Facilitate Observation. Event-based measurement focuses measurement so that not everything has to be observed. This in turn reduces data collector workload and creates a more economical expenditure of time. Importantly, it makes near real-time measurement a realistic pursuit.

Computer-based data collection. Many of the goals identified above such as near real-time measurement, as well as integrating data from

Elapsed Time	FCE EVENT/OBSERVATIONS	Observed Y=yes, N=no, N/A, M=missed	Comments
00:25	**INTEL EVENT: Support team, located in area of interest 3, reports tanks and BMPs in column. Suspected unit of 31st mech div (3 Corps).**		
	Performs tactical fire control (e.g., assesses target size, locations, checks grid; ensures target safe)		
	Reviews computer solution		
	Sends fire mission/fire order to appropriate unit (timely)		
	FCE notifies FSE on status of fire mission		
	Passes End of Mission message when received		
	Enters Mission Fired Report /updates active fire mission log		

Figure 1. Example of TRACTs checklist in paper-based format.

numerous data collectors, will require a computer-based data collection approach. As part of this effort, we took steps to implement TRACTs on a hand-held, off the shelf, pen-based computer. This provided several capabilities.

- The computer-based presentation allowed data collectors to easily access the different checklists. More than one checklist could be opened at a time. In addition, data collectors could make hand-written notes (i.e., "digital ink notes") and voice recordings.
- The handheld devices have a wireless communications capability, allowing data transfer and information to be provided to data collectors. During the Crusader effort, data collectors utilized this capability to communicate among themselves. A future enhancement will be to provide data collectors with information regarding battle ground truth, simulation systems status, and the occurrence of key events.
- Data collectors used the digital ink note capability to record comments and observations. These comments proved useful for providing an additional interpretation of performance data.
- After every battle, data were quickly aggregated to provide performance summaries to the unit. This could not have been accomplished as quickly using a paper-based format.

In the remainder of the paper, we describe the implementation of TRACTs for the Crusader exercise and provide examples of how the resulting data can be used to support T&E.

METHOD

Exercise Description

The purpose of the Crusader Concept Experimentation Program (CEP) was to identify operational concepts for the Crusader weapon system. Crusader is a revolutionary fire support weapon system that will be fielded by the Army in 2005. The exercise was conducted as part of a brigade-level collective training exercise. Approximately 90 soldiers manned the artillery command posts and Crusader workstations. The brigade tactical operations center (TOC) was manned by an additional 30 soldiers. Tactical fire control and battle management functions were observed from the battalion level through individual workstations simulating Crusader self-propelled howitzers and resupply vehicles.

Procedure

Data collection took place on nine days that occurred within a three week period. On each exercise day, participants responded to one of three tactical vignettes—movement to contact, implemented on days 1-3; defense in sector, implemented on days 4-6; and deliberate attack, implemented on days 7-9. On each exercise day, the event ran approximately five hours.

Seven observer controllers (O/Cs) were assigned to one of the seven cells to collect data using TRACTs. All O/Cs were either active duty or retired military personnel and all but one had direct experience with fire support operations. Cell assignments were based on their areas of expertise. O/Cs were tasked solely as data collectors and collateral duties were minimized.

O/Cs received approximately a one-hour training session occurring one to three days prior to the exercise. The training covered the basis for the TRACTs methodology, a review of each of the checklist types pertinent to the cell to which the O/C was assigned, the scoring procedure, and guidelines for sampling events. Training also encompassed use of the automated data collection tool. During each day of the exercise, the O/Cs were stationed in their respective cells during the data collection periods. Finally, the data collected were examined each day to ensure integrity.

RESULTS

The TRACTs methodology provides at least three options for data reduction, two of which are illustrated in Figure 2. Each of these approaches steps down from a look at performance at a summary level to increasingly specific performance scores. Using the taskwork approach, performance scores are first examined for the battlefield events targeted in the exercise. Data for each of the subtasks performed in responding to the battlefield event can then be examined to localize performance strengths and weaknesses. Figures 3 and 4 illustrate this option. Figure 3 presents summary scores for battlefield events examined for the BDE FSE cell. These data can be examined to identify trends. For example, responses to intelligence (intel) events appeared to be lower than the other battlefield events. In addition, performance on fire mission processing, one of the tasks most important to the evaluation during the Crusader CEP, appeared to decrease over the course of the exercise. Figure 4 shows detailed task data that bear specifically on this battlefield event. That is, the specific subtasks that were assessed for fire mission processing (i.e., at the level of individual checklist items) are shown. It can be seen that most of the subtasks increase over the course of the exercise. The exception is plotting the location of fire missions, where performance decreases. The likely explanation for this is that the units were increasingly using digital operations for fire mission processing (which has been documented) and thus the manual plotting of fire missions decreased.

The taskwork approach has high operational relevance and has immediate meaning to soldiers. For the Crusader CEP, fire mission processing for the different cells was examined to address questions pertaining to the interoperability of Crusader with digital systems and the impact of distributed operations.

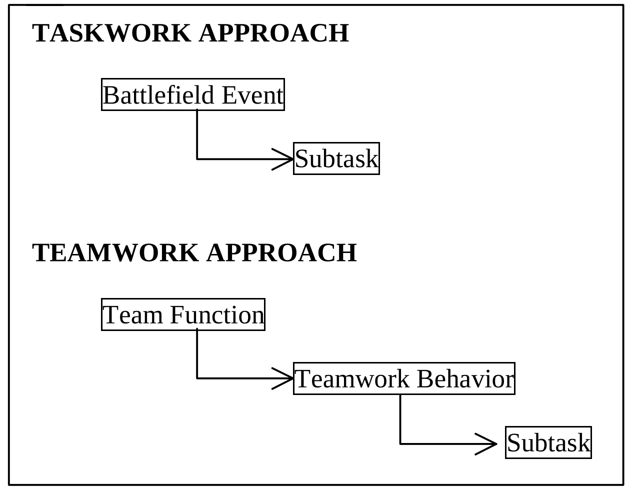

Figure 2. Data reduction approaches.

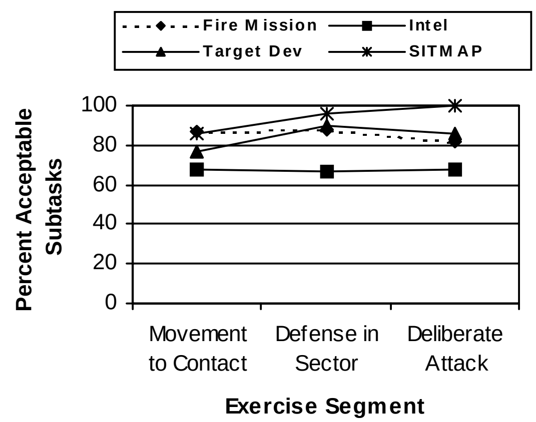

Figure 3. Battlefield events assessed in the BDE FSE cell.

The teamwork approach is illustrated in Figures 5 and 6 which again focus on the BDE FSE cell. Figure 5 shows summary scores for the team functions (note: the communication function was not assessed in the BDE FSE element). Scores for Initiative and Supporting Behavior show improvements over the course of the exercise. In contrast, scores for Information Exchange show no apparent change. Moreover, these scores tend to be lower than the other team functions. To obtain more diagnostic information, each team function can be reduced to its associated behaviors. Figure 6 shows scores for the teamwork

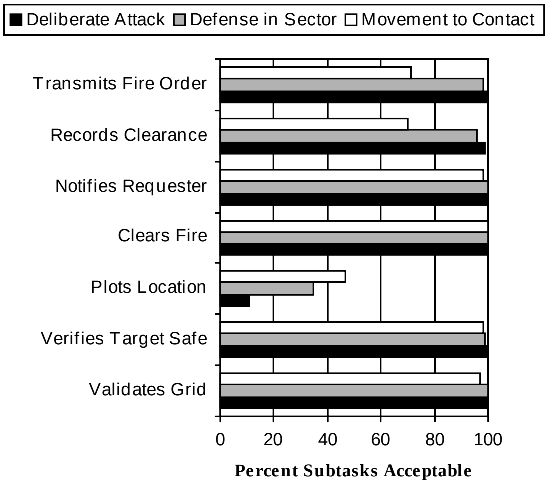

Figure 4. Fire mission processing subtasks shown across exercise phases for the BDE FSE cell.

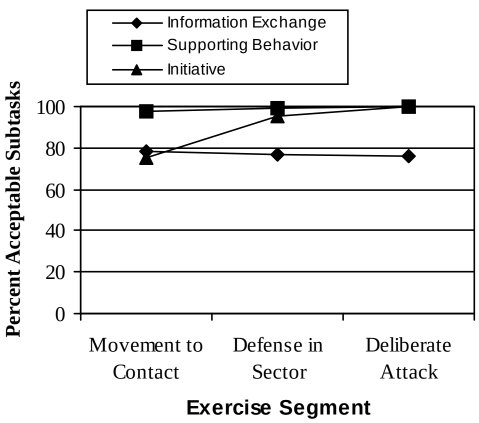

Figure 5. Performance on team functions for the BDE FSE cell.

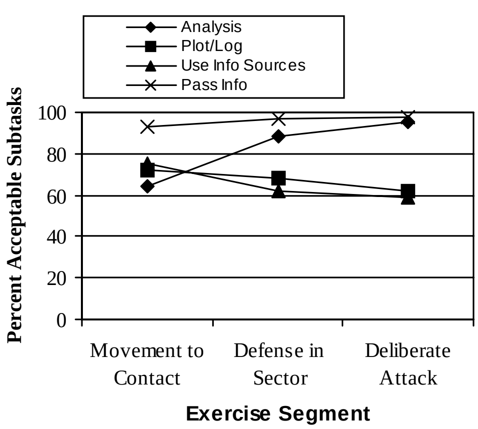

Figure 6. Performance on the Information Exchange teamwork behaviors for the BDE FSE cell.

behaviors associated with the Information Exchange team function. From this view it can be seen that subtasks pertaining to "passing information" are performed well throughout the exercise and those related to "analysis" improve. However, subtasks related to "plot/log" and "use information sources" appear to decrease. To obtain additional information on these, each teamwork behavior can be further reduced to the specific tasks or subtasks that provide operational manifestations of them. For example, the plot/log behavior was manifested for tasks such as plotting fire missions and updating situation maps. When these were examined (not shown), a problem was noted only on the subtask pertaining to fire mission processing. Thus, this was not a general problem and was limited to this one battlefield event. The teamwork approach to data reduction allows analysts (and participants) to see strengths and weaknesses in teamwork as well as the operational impact of good and poor team performance.

The third approach for data reduction is perhaps the most powerful. Using this method, the data analysis approaches are combined to isolate strengths and weaknesses as being due to teamwork, taskwork, or a weak cell. This analysis begins with assessment of team functions at the behavior level. For example, if it is determined that there is a problem with the teamwork behavior "use information sources" (a behavior associated with Information Exchange), we would first determine whether the problem is seen in more than one task (i.e., across battlefield events).

- If the problem is isolated to one task (e.g., to fire mission processing only), we would rule out team functioning as a major cause. We would then determine whether the problem is seen in more than one cell. If the problem is observed across cells, there might be a taskwork problem. We would ask why the task is performed poorly across different cells. Is it because all cells are performing poorly, is it because the task is extremely difficult, or is it because the task has become obsolete with the new system being addressed in the experiment? This situation was observed for the Crusader CEP for the plot/log subtask. A decrement in performance was observed across cells for the fire mission processing battlefield event (but other plot/log functions were maintained) suggesting that this task as a requirement for units to perform should be reexamined. If the problem appears to be limited to one cell, then it might be related to the poor performance of that cell. In the Crusader CEP, it was observed that the POC was one of the weaker elements.

- If the problem in the teamwork behavior is observed across tasks, then we argue that there is likely a teamwork problem at the large team (e.g., brigade) or subteam (e.g., element/cell) level. In the Crusader CEP, we observed problems in the team subcategory "use information sources." Lower performance of this teamwork behavior was observed across cells and across tasks, suggesting a general weakness in the ability to perform this team function. This would suggest a different intervention than a taskwork problem or a problem that was limited to one cell.

DISCUSSION

TRACTs was developed to address challenges inherent to synthetic warfighting environments. With the TRACTs methodology, (1) process performance measurement is utilized to produce diagnostic assessment from both teamwork and taskwork perspectives; (2) event-based measurement is incorporated to allow observation of relevant performance (i.e., control task content), facilitate observation, and enhance the ability to collect real-time observations; and (3) computer-based data collection is used to facilitate real-time data collection as well as data reduction so that data are quickly available to support feedback and decision making.

The capability to capture task-related and team-related performance was demonstrated by the sample of data presented in the Results section. In both of these approaches, it was shown that data could be viewed from global, summary levels—the levels at which general problem areas can be identified—to more detailed breakouts, arriving finally at an analysis of specific task-related behaviors. Collecting data within multiple cells also allowed us to combine teamwork, taskwork, and cell data to better isolate the cause of problems. For example, although poor performance may be noted in a teamwork area, it may be that the poor performance is the result of a particular element and is not a general problem area, or it may be determined that the poor performance is due to a particular task that has become obsolete with the introduction of the weapon system. Separating poor teamwork from poor taskwork or from a poor individual element is information that should be useful to decision makers and analysts involved in the testing of new weapon systems and tactics.

The handheld device was well-received by the data collectors used for this effort. In future projects, we plan to further exploit the technologies associated with these devices. For example, an important use of the wireless links between devices will be to provide information to data collectors about battle status and the occurrence of key events. What they understand about the exercise, battle, and simulation systems will all be a critical determinant of their effectiveness as measurement sources. A question to address in this regard is how should information be provided to data collectors to facilitate their battle awareness (e.g., how often, in what format). In addition, an issue that will be pursued is how near real-time data collection and reduction can be used to support exercise control. For example, if it can be determined that enough data have been collected to address a T&E issue so that the experiment can be concluded, or advanced to the next issue, then the efficiency of data collection can be increased and costs reduced. Similarly, if it can be determined that more data are needed on a particular question, this could improve the usefulness of the T&E exercise and possibly prevent the conclusion of an exercise that has not yet collected the necessary data. Thus, the use of the hand held devices for performance measurement can have the effect of improving the quality of data collected.

The emphasis of this effort was on a T&E application. TRACTs is also well-suited for training applications. For example:

- Process measurement is useful for providing specific feedback on performance strengths and weaknesses.
- The event-based methodology ensures training as well as measurement opportunities. In addition, the up-front work that goes into the identification of battlefield functions can also be used to organize feedback to trainees.
- The handheld data collection strategy offers many benefits to training applications, including rapid data reduction and performance summaries. It is also expected that such a system can support exercise brief and debrief presentation, similar to the way systems have been developed for Navy applications (Pruitt et al., 1997).

In conclusion, most efforts aimed at human performance measurement must rely on multiple methods to provide complete and valid assessments. This is especially true in synthetic warfighting environments in which complex human performances are assessed and in which there are numerous hurdles to effective measurement. We believe that approaches such as TRACTs, which specifically address many of the challenges, can serve as useful additions to measurement systems attempting to characterize combat performance.

REFERENCES

Cannon-Bowers, J. A., & Salas, E. (1997). A framework for developing team performance measures in training. In M.T. Brannick, E. Salas, & C. Prince (Eds.), *Team Performance Assessment and measurement: Theory, methods, and applications* (pp. 45-62). Mahwah, NJ: Erlbaum.

Dwyer, D. J., Fowlkes, J. E., Oser, R. L., Salas, E., & Lane, N. E. (1997). Team performance measurement in distributed envirorments: The TARGETs methodology. In M.T. Brannick, E. Salas & C. Prince (Eds.), *Team Performance Assessment and Measurement: Theory, Methods, and Applications* (pp. 137-153). Mahwah, NJ: Erlbaum.

Fowlkes, J. E., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). Event-based approach to training (EBAT). *The International Journal of Aviation Psychology*, *8*, 209-221.

Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T., & Oser, R. L. (1994). Improving the measurement of team performance: The TARGETs methodology. *Military Psychology*, *6*, 47-61.

Fowlkes, J. E., Lane, N. E., Dwyer, D. J., Willis, R. P., & Oser, R. (1995). Team performance measurement issues in DIS-based training environments. *Proceedings of the 17th Interservice/ Industry Training Systems and Education Conference* (pp. 272-280). Arlington, VA: American Defense Preparedness Association.

Lane, N. E. (1986). *Issues in performance measurement for military aviation with applications to air combat maneuvering* (NTSC-TR-86-008). Orlando, FL: Naval Training Systems Center.

Lane, N. E., & Alluisi, E. A. (1992). *Fidelity and validity in distributed interactive simulation: Questions and answers* (IDA Document-1066). Alexandria, VA: Institute for Defense Analysis.

Lane, N. E., & Kennedy, R. S. (1994). *Design of visual simulation systems in combined arms tactical training (CATT): Volume II: Resolving issues in training effectiveness evaluation* (TCN 94004). Research Triangle Park, NC: US Army Research Institute for the Behavioral and Social Sciences.

McCluskey, M. R., Fowlkes, J. E., Pierce, L. G., Dwyer, D. J. (1998) Measurement of command/control staff performance in tactical training environments [CD-ROM]. *Proceedings of the 20th annual Interservice/Industry Training, Simulation and Education Conference, Orlando, FL,* 826-833.

Pruitt, J. S., Burns, J. J., Wetteland, C. R., & Dumestre, T. L. (1997). ShipMATE: Shipboard Mobile Aid for Training and Evaluation. *Proceedings of the 41st annual meeting of the Human Factors and Ergonomics Society* (pp. 1113-1117). Santa Monica, CA: Human Factors and Ergonomics Society.

Smith-Jentsch, K. A., Johnston, J. H. & Payne, S. C. (1998). Measuring team-related expertise in complex environments. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 61-87). Washington, DC: APA Press.

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training*. (pp. 271-297) Washington, DC: APA Press.

Vreuls, D. & Obermayer, R. W. (1985). Human-systems performance measurement in training simulators. *Human Factors*, *27*, 241-250.