# MEASUREMENT OF COMMAND/CONTROL STAFF PERFORMANCE IN TACTICAL TRAINING ENVIRONMENTS

Michael R. McCluskey
Naval Air Warfare Center Training Systems Division
Orlando, FL
Jennifer E. Fowlkes
Summit Technologies
Orlando, FL
Linda G. Pierce
Army Research Laboratory
Fort Sill, OK
Daniel J. Dwyer
Naval Air Warfare Center Training Systems Division
Orlando, FL

**Abstract**

A performance measurement tool known as TRACE (Tactically Relevant Assessment of Combat Events) was developed and used to assess Command/Control (C2) staff training. TRACE allows the collection of detailed process measures in an efficient, streamlined manner. Alternate versions of TRACE were developed for each of four organizational elements related to the C2 staff. The performance of teams within these elements was assessed during field-based tactical exercises driven by C2 equipment stimulators and a time ordered events list. Observers scored the performance of these teams using TRACE checklists that were either based on the timed sequence of activities in the exercise, or on tasks that were done routinely. Portable, hand-held computers automatically opened the required checklist at the required time for each observer/controller. The computers provided a substantial increase in the efficiency and accuracy of data collection. In general, performance of the C2 staff elements improved over the time period covered by the exercises. This paper will: (1) describe the training environment used for this measurement project, (2) discuss the development of computer-based, hand-held performance measurement tools for use in field-based exercises to assess C2 staff performance, and (3) present selected data summaries of C2 performance as captured by the measurement tools.

**About the Authors**

Mike McCluskey is a Senior Research Psychologist in the Training Technology Development Branch of the Naval Air Warfare Center Training Systems Division. His experience includes developing approaches for measuring unit performance and task-based training, and development of automated collection tools.

Dr. Jennifer Fowlkes is a Senior Research Scientist with Summit Technologies. She has over ten years of experience in areas of human factors and training, training effectiveness evaluations, workload investigations, performance test battery development, and teamwork performance metrics.

Dr. Linda Pierce is Chief of the Army Research Laboratory (ARL), Human Research and Engineering Directorate, Fort Sill Field Element. Her programs include using simulations for training and system acquisition, assessing team performance, and integrating advanced simulations into the classroom.

Dr. Daniel Dwyer is a Senior Research Psychologist at the Naval Air Warfare Center Training Systems Division with 19 years experience in military training research, development, and acquisition support. His experience includes multi-service training, performance measurement, AAR methods, and team training.

# MEASUREMENT OF COMMAND/CONTROL STAFF PERFORMANCE IN TACTICAL TRAINING ENVIRONMENTS

Michael R. McCluskey
Naval Air Warfare Center Training Systems Division
Orlando, FL

Jennifer E. Fowlkes
Summit Technologies
Orlando, FL

Linda G. Pierce
Army Research Laboratory
Fort Sill, OK

Daniel J. Dwyer
Naval Air Warfare Center Training Systems Division
Orlando, FL

## INTRODUCTION

The purpose of this work was to develop a performance measurement system capable of supporting 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 incorporation of advanced technologies into Army systems and the requirement to test systems in realistic battlefield scenarios. STOW environments provide a key function in the test and evaluation (T&E) process, yet they pose huge challenges for effective measurement. Our objective was to develop a measurement system responsive to the challenges inherent in these complex testing environments

While team performance measurement has not typically been a component of the T&E process, it must now play a crucial role as systems and doctrine are tested in operational and simulated settings. Cannon-Bowers and Salas (1997) argue that team performance measurement must occur at multiple levels—that is, at team and individual levels—to provide a complete assessment of performance. There has been a great deal of research activity in recent years aimed at the development of team performance measurement tools and methodologies. Examples of individual performance constructs that relate to teamwork include assertiveness, information exchange, situation assessment and compensatory behavior. These are typically assessed using observational scales, expert ratings, and communication content analysis (Foushee & Manos, 1981). Examples of team constructs that relate to team performance include shared task models and cue/strategy associations.

Another useful distinction typically made (e.g., Cannon-Bowers & Salas, 1997; Dwyer et al, 1997) is between outcome and process measures. Outcome measures generally characterize an end result. As "end results" (e.g., number of fire missions performed) are generally affected by many uncontrolled variables, they have limitations as performance measures as they typically suffer from lack of reliability (Lane, 1986). When this is the case, they have limited utility in deriving answers to complex questions addressed in Advanced Warfighting Experiments (AWEs). Also, while outcome measures can signal the existence of a problem (e.g., a decrease in the number of fire missions performed), they have limited usefulness in diagnosing the cause of the problem. Trainee or soldier opinions and reactions can provide insight into causes but quantitatively-based process performance measures are needed for a truly diagnostic capability.

In contrast to outcome measures, process measures characterize how teams and individuals perform tasks. They can be extremely diagnostic of performance deficiencies if they are driven by a priori expectations. Moreover, the number of observations upon which scores are based can be much greater than outcome measures, giving them greater reliability.

Information age warfare and digitization of the battlefield are changing the nature of battle command and C2 processes. There is no universally accepted method for evaluating teamwork processes, to discriminate between effective and ineffective teams, and to diagnose the effect of teamwork on task performance. The lack of performance measurement in T&E environments limits our ability to evaluate the effectiveness of systems designed to support mission performance.

The objective of this effort is to adapt performance measurement methodologies from training applications to T&E environments. The resulting measurement tools must discriminate between effective and ineffective teams, be usable in T&E settings, and be generalizable across different task situations.

## METHODOLOGY

### Training Audience

Two division level tactical exercises served as the development testbed for the performance measurement tools. The primary training audience for these exercises was the Division Artillery (DIVARTY) Fire Support staff. This battle command staff was composed of several elements or "cells", briefly described below. The number of individuals in each element varied from four to eight.

***Operations Element, DIVARTY Tactical Operations Center (TOC).*** The operations element is responsible for planning and executing the current and future operations. It has the additional function of coordinating field artillery support for the division.

***Targeting Element, DIVARTY TOC.*** This element combines the available intelligence and targeting information. It is also responsible for developing enemy artillery targets for attack.

***Fire Control Element, DIVARTY TOC.*** The fire control element is responsible for analyzing targets for attack by field artillery. This element also issues fire orders to other field artillery units.

***Fire Support Element, Division Command Post.*** The fire support element synchronizes fire support for division operations. It is also responsible for planning fire support for extended operations.

***Fire Support Element, Division TOC.*** This element is responsible for coordinating fire support for the division close operations. It also responds to fire support requests for the maneuver brigades.

### Tasks For Performance Measurement

A training analysis was conducted as a part of another effort and served as input to this project. Training requirements were identified through field surveys designed to identify critical training tasks and unit priorities. Six DIVARTY and three field artillery brigades provided responses to the surveys. The results from these surveys were supplemented with detailed information from various doctrinal sources. The final tasks, presented in Table 1, became candidates for performance measurement.

**Table 1**
Critical Tasks Selected for Training and Performance Measurement

Operations Element
- Prepare the fire support plan
- Establish and monitor the force FA TOC
- Direct and coordinate FA unit operations
- Coordinate FA conventional fires
- Provide counterfire
- Coordinate target acquisition and counterfire

Targeting Element
- Monitor and recommend employment of target acquisition assets
- Predict and produce targets
- Develop targets and potential targets
- Conduct IPB
- Maintain current enemy situation
- Coordinate target acquisition and counterfire
- Prepare intelligence plans and orders

Fire Control Element
- Control and coordinate FA missions
- Prepare schedule of fires
- Establish and maintain the FCE
- Control counter attack

Fire Support Element
- Conduct FS coordination in support of ground operations
- Synchronize FS operations
- Conduct FS operations
- Conduct FS planning
- Coordinate target attack
- Process target attack
- Establish and maintain FS coordination facilities
- Prepare FS portions of plans and orders

**Tactical Training Environments**

Two separate tactical training exercises served as the T&E environment. The first exercise was conducted in October 1997 (Unit A); the second exercise in January 1998 (Unit B). The training audience for both exercises was located in tactical field environments with tactical equipment and vehicles. The tactical operations center and division command posts were located adjacent to simulation center facilities to provide the support of role players and exercise control functions.

**Training Exercise**

The overall training exercise and the operations order that initiated the exercise were developed in collaboration with the division training personnel. The design for the training exercise was intended to incorporate the maximum number of critical tasks identified in the training needs survey as well as satisfy training needs indicated by division personnel. The major events of the defensive tactical scenario were used to generate the exercise details provided in the Mission Events List (MEL). The MEL served as the basis for preparing a time ordered events list to drive tactical equipment and other aspects of the simulation. For both training locations, the exercise was planned to continue for 48 hours. The divisions varied in the approach they chose to represent continuous operations.

**Measurement Instrument Development**

The primary inputs to the development process for the performance measurement tools were the 25 critical tasks identified in the training needs survey, individual and team responsibilities, functions for each of the five organizational elements, and a model of information exchange between the elements. The first step was to match the critical tasks with the appropriate MEL event and each DIVARTY organizational element. This provided a picture of the distribution of tasks across time and elements, and indicated when and where measurement opportunities would occur. The entire MEL was examined again from the standpoint of identifying tasks that occurred more than once over the 48 hour period. These tasks would permit looking at changes over time through repeated measurements on the same behaviors.

The event-based methodology that emerged for assessing within and between cell staff processes was referred to as TRACE (Tactically Relevant

Assessment of Combat Events). TRACE was based on a team performance measurement methodology originally referred to as TARGETs (Targeted Acceptable Responses to Generated Events or Tasks) (Fowlkes, Lane, Salas, Franz, & Oser, 1994). The impetus for its development was a requirement to evaluate the execution of team skills in aircrew coordination training (ACT).This methodology has demonstrated sound psychometric properties and was successful in detecting differences between ACT-trained and untrained crews (Fowlkes et al, 1994).

The implementation of TRACE is based on the notion that, for measurement in simulated environments, it is crucial to control the nature and frequency of observations that are made, similar to the way task content of a multiple choice test can be controlled through selection and number of items. To do this, an event-based methodology (Fowlkes, Dwyer, Oser, & Salas, 1998) is applied whereby exercise events are used to create opportunities for participants to demonstrate key competencies which are traceable to training objectives. For each event, observable, acceptable behaviors are identified a priori through the use of standard operating procedures, technical documentation (e.g., tactics, techniques, and procedures) and expert input. Thus, judgments and performance standards are applied before the measurement opportunity. During the measurement opportunity, the acceptable behaviors expected for each event are scored as either "present" or "absent." The control over the timing and nature of events as well as the observation of behaviors that are determined a priori are all meant to minimize judgments that are required in real time as well as standardize task conditions and observations that are made.

Two broad classes of exercise events were identified after examining exercise materials that included the Operations Order and MEL. The first class of events are those specified in the MEL and, as such, their content, method of delivery, and timing can be specified a priori. For these events, "MEL-driven" checklists were developed and assembled into a TRACE observation tool. There were nine "MEL-driven" checklists that were associated with the following areas and functions: Intelligence reports, intelligence summaries, logistic, call for fire missions, suppression of enemy air defense, high payoff targets, direct support unit movement, direct support target acquisition battery movement, and situation maps.

The other class of events are those we know are important and that we know should occur. However, when they occur is determined largely by the actions of the participants themselves. For example, we know that it is important for situation update briefs to occur. The stimulus events for performing situation updates are varied—it may be the DIVARTY commander entering a cell, a shift change, or some requirement in the unit's SOP. Performance in response to these events was assessed using "on-call" checklists that were included in the TRACE observation tool in such a way that they were easily called up once the observer/controller (O/C) recognizes the requirement to use them. For these, the task for the O/C is more difficult as he or she must remain vigilant to identify the requirement to perform them. Six "on-call" checklists were developed to cover the following areas and functions: Situation updates, general support unit movement, general support target acquisition battery movement, clearing fires, target acquisition, and fire missions.

The development process used for these checklists insured that there were solid linkages between training needs, the doctrinal descriptions of the critical tasks in the Army Mission Training Plans, the MEL items, and the checklist items.

An example of a "MEL-driven" checklist for the operations element is given below. The items listed were to be performed when intelligence information was received in response to the MEL item: "101st reports contact with lead enemy elements along PL Black. Suspected unit of 1st Bde, 20th tank div."

| | | | | |
|---|---|---|---|---|
| Y | N | N/A | N/O | Intel coordinated in TOC |
| Y | N | N/A | N/O | Plotter plots intel & brings to attention of Battle Captain/Ops officer |
| Y | N | N/A | N/O | Ensures intel disseminated |
| Y | N | N/A | N/O | Compares intel to decision support template |
| Y | N | N/A | N/O | TGT & Ops cross talk (intel ramifications) |

Y = yes, performed correctly;
N = no, not performed correctly or omitted;

N/A = not applicable;
N/O = no opportunity

Each checklist item is scored as Y, N, N/A, N/O as described above. The proportion of items performed properly (# yes relative to all checklist items) provides a score.

**Data Collection Procedures**

The procedures implemented for the Unit A and Unit B exercises were similar. For both exercises, four O/Cs were assigned: one to each of the four cells. Their job was to collect data using the TRACE tool. All O/Cs were either active duty or retired military personnel and had direct experience with DIVARTY operations. Cell assignments were based on their areas of expertise. Two of the O/Cs were used in both exercises. To the extent possible, O/Cs were devoted to data collection and collateral duties were minimized.

O/Cs received a one-hour training session occurring one to three days prior to the exercise. The training covered the basis of the TRACE 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 the use of "on-call" and "MEL-driven" checklists. For the Unit B exercise, training also encompassed use of the automated data collection tool. O/Cs received an advance copy of the TRACE checklist as well as the MEL. Prior to data collection, the O/Cs were able to review these materials in detail and address any questions they had with the experimenters.

During data collection, the O/Cs were stationed in their respective cells during the data collection periods. The data collected were examined each day to ensure integrity. For the Unit A exercise, this primarily entailed ensuring that O/Cs completed all of the observations required for the events that had occurred and spotting items that appeared to be ambiguous. For the Unit B exercise, some of the data "QA" checks had been built in to the automated device—for example, an open checklist window could not be closed unless all observations were scored. Other data integrity checks included looking at the data files to ensure they were saved as planned. At both exercises, experimenters routinely queried the O/Cs to help resolve misunderstandings about the scenario and particular checklist items and to ensure that the TRACE tool was being used correctly. In general, there were few problems encountered during data collection.

**RESULTS AND ANALYSIS**

Table 2 shows the number of observations collected for each cell across days for the Unit A and B exercises.

**Table 2**
Total Number of Observations Obtained by Condition

| Scoring Category | Unit A (2360) | | | | Unit B (2316) | | | |
|---|---|---|---|---|---|---|---|---|
| | OPS | FCE | FSE | TGT | OPS | FCE | FSE | TGT |
| No | 164 | 186 | 76 | 250 | 176 | 61 | 90 | 166 |
| Yes | 171 | 208 | 205 | 448 | 214 | 161 | 232 | 442 |
| Missing | 62 | 37 | 3 | 31 | 30 | 251 | 6 | 42 |
| Not Applicable | 89 | 143 | 130 | 157 | 56 | 181 | 81 | 127 |
| Total | 486 | 574 | 414 | 886 | 476 | 654 | 409 | 777 |

Similarities across exercises include a comparable number of observations and more observations collected for the targeting and fire control elements than for the other two cells. This was due to the on-call checklists related to the counterfire missions used for these cells which occurred routinely throughout the exercises. Approximately 31% of the checklist items were scored as "not applicable" or "missing." The bulk of these were due to confusions surrounding unit movements which were not played; events were not counted as occurring and thus the

observations were marked as non applicable. This accounted for 28% of the Unit A and 33% of the Unit B data. Less than .5% of the data collected at the Unit B exercise were lost due to computer error. It appeared that O/Cs were exercising the scoring categories and using them appropriately.

This project collected a considerable amount of data on DIVARTY staff performance. Depending on the particular area of interest, the information may be organized by MTP task and subtask, unit training needs and objectives, items or sections of the MEL, DIVARTY staff element, tasks performed by more than one staff element, or other meaningful "clusters" of specific checklist items. The data presented in the following charts provide an illustration of the nature of the information collected. The first two graphs are more general in nature and the next two show more specific levels of the types of information that are available. These charts represent a very small sample of possible summaries.

Figures 1 and 2 show the overall performance trends for each of the DIVARTY staff elements for each day of the exercises.

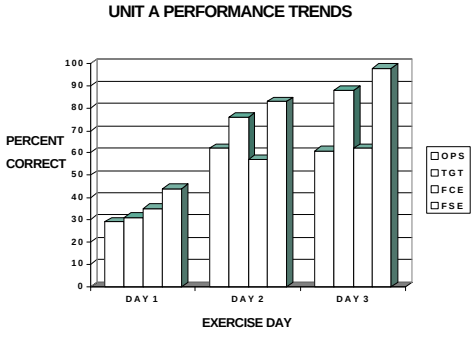


Figure 1. Overall percent correct by staff element over time for unit A

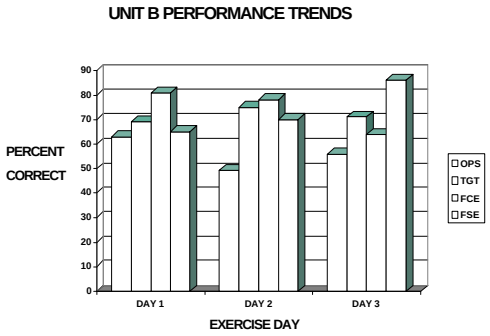


Figure 2. Overall percent correct by staff element over time for unit B

The performance of each element is measured in terms of the percent "present" or "absent" on the total number of checklist items. Overall, Unit A exhibited an upward trend from about the 35% level on day 1 to approximately 75% on day 3. The performance of Unit B was fairly flat at about 65% across the three days. Unit B may have had a repetition of a similar exercise more recently than Unit A, and therefore did not show the lower initial performance levels that resulted for Unit A on day 1.

Figures 3 and 4 illustrate further levels of detail that are possible with all categories of information collected. Performance on the subtask labeled "Process Fire Missions" is graphed in Figure 3.


Figure 3. FCE element performance on process fire missions

Each data point on this figure is based on the performance for completing two fire missions. Although there is considerable variability by day and across the days, there is an overall upward slope to the graph showing improvement in

performance over time. The next level of detail that may be examined is shown in Figure 4.

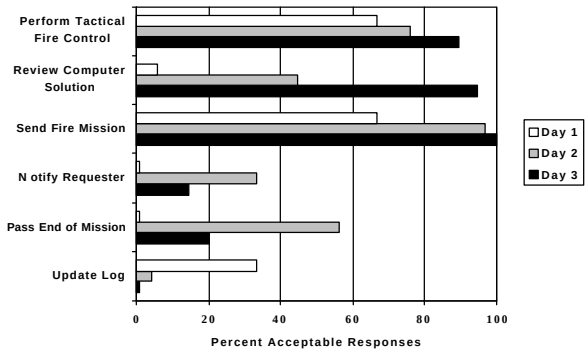


Figure 4. FCE tactical behaviors associated with processing fire missions.

This graph shows performance on six tactical behaviors that are required in order to successfully complete the subtask "Process Fire Missions". The specific behaviors involved in actually sending the fire mission improved over the three days to quite high percent correct levels. The behaviors concerned with reporting and communicating the outcome of the processing activity, however, were at low percent correct levels with considerable variability. Such in patterns performance can provide a diagnostic assessment of where performance strengths and weakness manifested themselves.

## SUMMARY

This project demonstrated a methodology to measure DIVARTY staff performance in a T&E environment. The measurement instrument development process used Army Mission Training Plan tasks as the foundation. For development and analytical purposes, linkages were then made to unit training needs and objectives, organizational elements of the DIVARTY staff, and specific events in the training scenario. These linkages and the development process gave the checklists content validity. This effort was not intended to compare groups or compare their performance against an objective criterion other than the Mission Training Plan tasks. It will be necessary for subsequent studies to establish the validity and reliability of the instrument for measuring DIVARTY staff performance.

## REFERENCES

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

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.

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, & Applications. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Foushee, H. C., & Manos, K. L. (1981). Information transfer within the cockpit: Problems in intra cockpit communications. In C. E. Billings & E. S. Cheaney (Eds.), Information transfer problems in aviation systems. (NASA TP-1875). Moffett Field, CA: NASA-Ames Research Center.

Lane, N. E. (1986). Issues in performance measurement for military aviation with application air combat maneuvering (Technical Report Number 86-008). Orlando, FL: Naval Training Systems Center.

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