MEASURING THE IMPACT OF ADVANCED TECHNOLOGIES AND REORGANIZATION ON HUMAN PERFORMANCE IN A COMBAT INFORMATION CENTER

Jared T. Freeman
Aptima, Inc.
Woburn, Massachusetts

Gwendolyn E. Campbell
Naval Air Warfare Center
Training Systems Division
Orlando, Florida

Greg Hildebrand
Basic Commerce & Industries, Inc
Dahlgren, Virginia

Systematically evaluating the impact of novel technology and organizational structure on team performance is a complex, multidimensional task. We define several of these dimensions that are of particular interest in the development of new command and control teams and technologies for the U.S. Navy. In addition, we describe an approach to stimulating and measuring human behavior on these dimensions, and an experiment in which this approach is applied. Preliminary data are presented.

INTRODUCTION

The goal of the Manning Affordability project is to demonstrate that a Human-Centered Design approach can produce complex Naval systems that can be manned effectively by relatively few warfighters. The project leverages two innovations to achieve its goal: (1) advanced watchstations, which potentially will enable watchstanders to access and control resources with far greater efficiency than in the past, and (2) formal, algorithmic techniques for generating and evaluating watchstanding teams, techniques that leverage characteristics of the new technology to optimize the allocation of responsibilities and resources. This paper, part of a symposium concerning the Manning Affordability project, describes a quasi-experiment evaluation of the combined effects of the advanced Multi-Modal Watchstation (MMWS; Osga, 2000) used by teams that are optimized using TIDE algorithms (Levchuk, et al., 1999a, and 1999b). That experiment began in the fall of 1999 and is currently nearing completion.

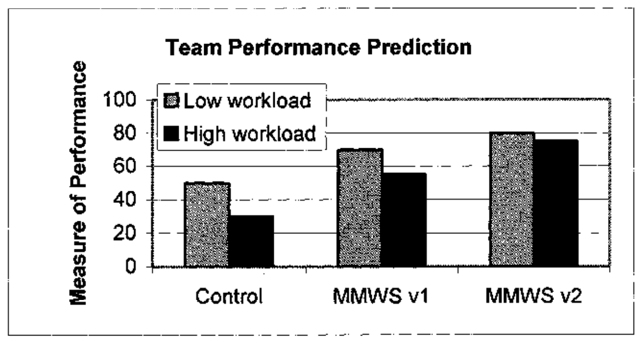
EXPERIMENTAL PLAN

Hypotheses

The study tests two hypotheses. First, we predict that optimized teams using MMWS technology will perform as well or better than standard command and control teams of twice the size using the technology currently found in the Combat Information Center (CIC) of an AEGIS cruiser. Two versions of the MMWS will be evaluated, and we predict that the second generation (MMWSv2) will improve team performance over the first (MMWSv1).

Second, as illustrated in Figure 1, we posit an interaction effect such that performance on various measures (below) will degrade less as mission workload rises for MMWS teams than for conventional teams. MMWSv2 should further minimize performance decay.

We use a mixed factorial experimental design to test these hypotheses. Both team structure and advanced technology vary between participating groups. Workload varies within teams as a function of the number and complexity of tasks demanded by the anti-air warfare scenario developed for this study.



Method

The approach we have taken is to use formal modeling techniques and recent team performance literature to predict the impact of team optimization and MMWS technology on behavior, present a scenario that affords the opportunity for those behaviors to emerge (see below), and empirically measure those behaviors using objective and subjective instruments, as well as the ratings and commentary of domain experts who observe the experimental sessions. Several classes of individual and team behavior are of particular interest in this experiment. Here we discuss these behaviors and the instruments used to measure them.

Detect-to-Engage Behaviors. Among the tasks performed in the CIC, few are more important to mission success and survival of the ship than detecting, identifying, assessing the intent of, and if necessary engaging air platforms operating near the ship and her battlegroup. This set of tasks is called the Detect-to-Engage, or DTE sequence. DTE behaviors are measured here using the Anti-Air Warfare Team Performance Index (ATPI; after Dwyer, 1992). Expert judges monitor participants during the scenario and record the time and accuracy with which they perform standard tasks such as detect an aircraft (or "track"), correlate diverse data concerning it, identify the track, classify it, issue warnings, take defensive and offensive actions, and execute post-engagement activities. Originally a paper form, we have implemented the ATPI on a handheld computer to cue SMEs to take measures regarding key tracks, to automate the collection of time-stamps for each observed event, and to help SMEs associate free text notes with the ratings they make on this highly structured form. Data from the ATPI support assessment of the timing and accuracy of DTE actions relative to expert protocols. A similar, paper instrument is used to assess the performance of the watchstander responsible for managing air intercepts.

Multi-tasking. One objective of the MMWS is to enable watchstanders to handle a greater number of tasks, and potentially a greater variety as well. Thus, a test of the success of the MMWS is its effect on the number of tasks watchstanders successfully handle per unit time. ATPI data will be used to compute relevant measures. One is the average number of DTE actions per track, which we predict will rise. A second measure considers the number and variety of tracks that the team addresses. While all teams may address the highest priority tracks, we predict that MMWS teams will have the spare cognitive resources to complete more tasks associated with the lower priority tracks, as well.

Communications. A central objective of both MMWS technology and team design is to minimize communication and coordination between watchstanders (except where operating procedures require it for legal reasons, ship's safety, and other reasons). Specifically, new technology makes the content of communications accessible across the team and over time (by storing them for playback), and it automates standard outgoing communications. Team optimization attempts (among other things) to ensure that tasks requiring the same body of information are performed by a single watchstander when possible, to avoid errors and delays that occur during communication.

Several methods are used to evaluate communications. Subjective measures are taken using an expanded version of the NASA Task Load Index workload survey (Hart & Staveland, 1988) on which participants rate the communications demands of the scenario at two junctures in the scenario. Trained observers also assess the quantity and content of specific types of communications — such as information passing and situation updates — using a variant of the Team Dimensional Training (TDT) instrument (Smith-Jentsch, et al., 1998). Finally, verbal communications are recorded to support content analysis using a technique that considers the source, destination, and type of content, such as request, information provision, coordination and other factors (Freeman et al., 1997).

It is entirely possible that MMWS teams will talk no less than standard teams, but that they will use the available "airtime" to address different matters. In particular, the MMWS team may be enabled to dedicate a smaller proportion of their communications to passing routine information (e.g., concerning track location) and a larger proportion to discussing high-level issues such as contingency plans and the optimal allocation of defensive assets between threats. Content analysis of taped communications for specific events will provide evidence of these effects. The TDT has also been extended to enable observers to rate and comment on communications that reflect critical thinking. Empirical data concerning communications and coordination will be compared with measures generated by TIDE and by the Integrated Performance Modeling Environment (IPME; Dahn, et al., 1997), which is being used to validate the team design computationally.

Situation assessment. The ability of watchstanders to filter voluminous data, seek key information, plan, and take appropriate actions hinges on the accuracy of their assessments of the tactical situation. In this study, the principle measures of assessment accuracy are participant responses to questions concerning the relative priority of tracks, the potential intent of those tracks, and ownship's intent with respect to them. We predict that improved technology and team designs will improve responses to these questions between teams, relative to experts. Responses should also be more uniform within the MMWS teams, reflecting shared mental models of the tactical situation, the team's priorities, and it's state with respect to those priorities.

It has been proposed that high-performing teams maintain an accurate mental model of the state of the team itself (Cannon-Bowers & Salas, 1990). This capability is particularly important for the team leader(s), whose decisions concerning the assignment of tasks and allocation of resources depend critically on knowing the available capacity of the watch team members. To assess the effects of MMWS technology and team optimization on this capability, we will examine correlations between ratings by the team leader of subordinates' effort with subjective ratings by the watchstanders themselves. Higher correlations indicate better assessments by leaders of team state, and these are expected in the MMWS conditions.

Workload. A difficult aspect of CIC operations is the extreme variability in workload across the team between mission events, and the highly uneven distribution of workload within the team during a given event. The design of teams and technology in this project aims to ameliorate these problems even while cutting team size in half.

Three instruments are used to assess workload. The NASA TLX is administered during planned lulls ("coast" and "cool-down" periods) in the experimental scenario to elicit participants' ratings of the demands of the scenario and its effects on their level of effort, performance, and frustration. SME ratings of workload are taken as well.

Variation in mean TLX and SME ratings over experimental groups within the scenario is a check on the success of our manipulation of workload levels over the span of the experiment. Changes in TLX and SME ratings between groups within the scenario should reveal the effect of team design and technology on workload. The correlation between individuals' TLX ratings should inform analyses of workload balance. Finally, these ratings and responses to an extension of the TLX (concerning the demands imposed by the scenario on communications, monitoring, control, and coordination) will be compared to predictions of the TIDE to refine and validate this tool.

Team process. Work by Smith-Jentsch and her colleagues (1998) defined and validated a set of dimensions along which teams vary by expertise and that predict team performance. The TDT, cited above, is a very simple and efficient instrument with which SMEs can evaluate teams with respect to information exchange, communication, supporting behaviors, and leadership. (It has been extended in this study to capture data about critical thinking). We predict that TDT data will show significant gains on most or all of these dimensions because MMWS teams have more better access to information and lower communication and coordination demands, which should lower workload enough that watchstanders can attend to the important tasks of supporting and leading each other efficiently.

Scenario Design and Description

The scenario used in this experiment is designed to be highly realistic of Anti-Air Warfare missions, to trigger behaviors of the types mentioned above, and to provide opportunities for measurement.

Fidelity. Realism was achieved by building upon an existing, unclassified AEGIS Combat Training System (ACTS) Scenario that is considered to be at an advanced level of difficulty. The scenario includes a hypothetical threat order of battle. It presents a total of 150 tracks, with 20-30 tracks active at any given instant. The participants' ship is tasked to be a critical node in the Air Defense Network, in an Aircraft Carrier Battle Group, under the traditional Composite Warfare Commander (CWC) Concept. The battle group is underway in a littoral environment, in a Condition-III Level of Readiness (typical of deployments today), under Peacetime, Restrictive Rules-of-Engagement (ROE). The test ship is also tasked to control interceptor aircraft, known as Defensive Counter Air (DCA) and serve as the Battle Group coordinator.

The scenario is presented in the realistic work environment. For the AEGIS testing, the scenario stimulates combat systems through an embedded training system (ACTS). For the MMWS testing, the same scenario stimulates the team using the Joint Semi-Automated Forces scenario simulation environment (JSAF). The capabilities and limitations of the weapons and sensor systems are assumed to be equal between AEGIS and the MMWS CIC.

Behavioral triggers. The scenario is seeded with events designed to trigger the behaviors cited above. Those events were developed from successful vignettes in prior Navy research and from the experience of SMEs. A larger scale manipulation is meant to affect workload and, indirectly, the difficulty of situation assessment. It consists of two workload levels within the scenario. During the initial, lower stress period, threat track behaviors are less ambiguous, there is more time for operators to react, and the intelligence reports indicate that attack on U.S. Forces is unlikely. During the subsequent, higher stress period, threat tracks are more ambiguous, they fly directly at friendly forces thus decreasing reaction time, average track load is higher, and the intelligence information incrementally increases to the point that attack on U.S. Forces is probable.

Measurement opportunities. The scenario is one hour and 55 minutes long. It includes a ten min warm-up period, a forty-minute lower stress period, a fifteen-minute coast period, a forty-minute higher stress period, and a ten-minute cool-down period. Objective data are taken during the two stress periods, and subjective SME and operator data are taken during the coast and cool-down periods.

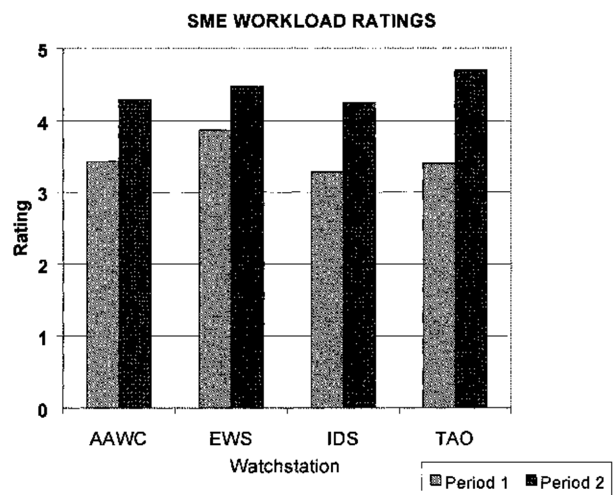
Participants

The officers who execute this scenario are organized as either (1) the anti-air warfare component of a standard AEGIS CIC watchstanding team of nine personnel or (2) a novel Air Defense Warfare (ADW) team of roughly half its size. The sailors playing the standard AEGIS ADW roles are active duty intact CIC watch teams and are tested pier side on their existing AEGIS Systems. Sailors in the experimental condition are members of intact ADW teams from AEGIS Ships. Although these teams have worked together in their old roles, they will require training to conduct the mission in the novel team organizational structure and how to perform ADW DTE Tasks using a MMWS. Testing is being conducted at the Integrated Command Environment (ICE) Lab at the NSWC in Dahlgren, VA.

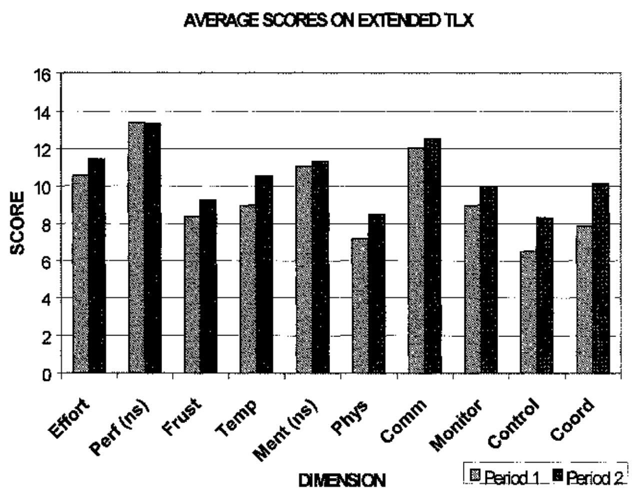
PRELIMINARY RESULTS

Preliminary Results

Currently, only data from the anti-air warfare component of a standard AEGIS CIC team have been collected. While the majority of the interesting comparisons will ultimately be between the AEGIS teams and the MMWS teams, a validation of the workload manipulation was conducted using the data gathered to date. Figure 2 shows the average subject matter expert ratings of workload for several watchstanders during the first and second halves of the scenario.



While these data are suggestive, it could be argued that the raters were not condition-blind, and thus their ratings may be influenced by their knowledge of the scenario design. Figure 3, on the other hand, shows the average NASA TLX ratings produced by the watchstanders themselves, during the coast and cool-down periods. The watchstanders were not aware of the stress manipulation, and thus these data provide strong evidence that the manipulation was a success. In particular, statistical analyses show that ratings on every subscale except mental demand and performance showed a significant increase from the lower stress period of the scenario to the higher stress period of the scenario.



DISCUSSION

The measurement approach described here reflects three core concepts. First, the impact of organization and technology on team performance can be modeled (using TIDE, IPME, and other tools) to predict effects on performance accuracy and tempo, workload, communication, coordination, and other variables. To the extent that these predictions are accurate, we validate the models and can more confidently use them to design new technologies and teams. Errors in predictions offer us guidance in refining these models for future use. Second, this experiment is grounded in a complex, high fidelity scenario. This increases the face validity of the experience for participants. For experimenters, the scenario's value is as a carefully designed stimulus for specific behaviors of theoretical interest. Third, we measure those behaviors using instruments that were refined and validated in prior research, particularly the Navy's program on Tactical Decision Making Under Stress. Model-driven, scenario-based experimentation using well-vetted instruments is a promising method of studying human performance on complex tasks. We look forward to reporting full results of this work in the coming months.

ACKNOWLEDGEMENTS

This work is funded by the Office of Naval Research and supported by the Naval Sea Systems Command. We are indebted to the men and women of the Navy who serve as participants and administrators in the experiments described above. 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

Cannon-Bowers, J.A., and Salas, E. (1990). *Cognitive psychology and team training: Shared mental models in complex systems.* Paper presented at the 5th annual Conference of the Society for Industrial and Organizational Psychology, Miami, FL.

Dahn,D.A.,Laughery, K.R., Belyavin,A.J. (1997), The integrated performance modelling environment, A tool for simulating human-system performance. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting,* Volume 2, pp. 1037-1041. Albuquerque, New Mexico.

Dwyer, D. J. (1992). An index for measuring naval team performance. *Proceedings of the Human Factors Society 36th Annual Meeting, 2,* 1356-1360.

Freeman, Jared T., Cohen, Marvin S., Serfaty, Daniel, Thompson, Bryan & Bresnick, Terry. (1997). *Training in Information Management for Army Brigade and Battalion Staff.* (Technical Report #1073). U.S. Army Research Institute for the Behavioral and Social Sciences Armored Forces Research Unit, Ft. Knox, KY.

Hart, S.G., and Staveland, L.E. (1988). Development of a muli-dimensional workload rating scale: Results of emperical and theoretical research. In: P.A. Hancock and N. Meshkati (Eds.), *Human Mental Workload.* Amsterdam. The Netherlands: Elsevier.

Johnston, J.H., Cannon-Bowers, J.A., and Smith-Jentsch, K.A. (1995). Event-based performance measurement system. *Proceedings of the Symposium on Command and Control Research and Technology,* June 1995 (pp. 268-276)

Levchuk, Y. N., Luo J., Levchuk G. M., Pattipati K. R., and Kleinman, D. L. (1999) A multi-functional software environment for modeling complex missions and devising adaptive organizations. *Proceedings of the 1999 Command & Control Research & Technology Symposium,* War College, Newport, RI, June/July 1999.

Levchuk, Y.N., Pattipati, K.R. and Kleinman, D.L. (1999). Analytic model driven organizational design and experimentation in adaptive command and control. *Systems Engineering,* Vol. 2, No. 2, 1999.

Osga, G. (2000). Task managed watchstanding — Concepts for 21st century naval operations. *Proceedings of the 44th Human Factors and Ergonomics Society Annual Meeting.* San Diego, CA.

Smith-Jentsch, K.A., Johnston, J.H., and Payne, S.C. (1998). Measuring team-related expertise in complex environments. In *Making Decisions Under Stress.* (Janis A. Cannon-Bowers and Eduardo Salas, eds.). Washington, D.C.: APA.