SPAWAR
Systems Center
San Diego

TECHNICAL DOCUMENT 3184
July 2004

Tactical Tomahawk
Weapon Control System v6
Land Attack Combat System
Prototype Human–Computer Interface

Test Report for FY 03 Fleet Operability Test

James Pharmer
NAVAIR Orlando Training Systems Division

Kevin Cropper
Jennifer McKneely
Johns Hopkins University Applied Physics Laboratory

Earl Williams, Ph.D.
SSC San Diego

BEST AVAILABLE COPY
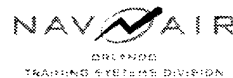
SSC San Diego

20040923 054

TECHNICAL DOCUMENT 3184
July 2004

Tactical Tomahawk Weapon Control System v6 Land Attack Combat System Prototype Human–Computer Interface

Test Report for FY 03 Fleet Operability Test

James Pharmer
NAVAIR Orlando Training Systems Division

NAV AIR
ORLANDO
TRAINING SYSTEMS DIVISION

Kevin Cropper
Jennifer McKneely
Johns Hopkins University Applied Physics Laboratory

Earl Williams, Ph.D.
SSC San Diego

SPAWAR
Systems Center
San Diego

Approved for public release;
distribution is unlimited.

SPAWAR
Systems Center
San Diego

SSC San Diego
San Diego, CA 92152-5001

SSC SAN DIEGO
San Diego, California 92152-5001

T. V. Flynn, CAPT, USN
Commanding Officer

R. F. Smith
Executive Director

ADMINISTRATIVE INFORMATION

ACKNOWLEDGMENTS

LH

EXECUTIVE SUMMARY

Human–computer interfaces (HCIs) have traditionally been structured according to a system's functional steps, which typically force human operators to adjust to an information organization that is foreign to them. In contrast, a task-centered HCI is structured to directly support operators in effectively and efficiently completing their tasks. This report describes an operability test of a prototype HCI developed through a task-centered design process. Using this prototype, participants individually performed a simulated Tactical Tomahawk launch and control scenario. The testing scenario was substantially compressed; using a current fleet Tomahawk Weapon Control System, a comparable tasking scenario ordinarily would be four times as long and would be processed by four operators instead of one. Participants were given a half day of training on the prototype system before their testing. During the testing, data were collected on the participants' task completion timeliness, situational awareness, and subjective workload. Analysis indicates that the participants successfully completed nearly all tasking on time, while maintaining a moderate to high level of situational awareness and reporting perceived workload to be low to moderate. Operators gave generally positive feedback about the prototype system's functionality, ease of use, and trainability.

CONTENTS

FIGURES

TABLES

1. INTRODUCTION

The purpose of this document is to outline the human performance evaluation of the FY03 Fleet Operability Test (FLOT) of the Land Attack Combat System (LACS) Human–Computer Interface (HCI) prototype, developed by the Space and Naval Warfare Systems Center, San Diego (SSC San Diego). The goal of this evaluation was to determine the level of performance that can be expected from the Tactical Tomahawk Weapon Control System (TTWCS) given the latest incorporation of human factors engineering principles and task management techniques developed under an Office of Naval Research (ONR) Future Naval Capability (FNC) research and development effort[1].

It should be stressed that the purposes of this evaluation were to explore the effectiveness of the current TTWCS prototype HCI in supporting performance and to identify any weaknesses in its design that could be addressed during either the ongoing FNC effort or fleet system development. As such, this investigation was not intended to be an experimental comparison. Instead, data were collected under one condition: an operator with prior Tomahawk experience performing a scenario using the prototype HCI. For the scenario, a single operator performed tasking ordinarily executed by multiple operators over a longer time period in less time. Analysis then determined whether the individual operator successfully and comfortably completed tasking or was close to being unable to successfully accomplish tasking, in terms of workload and awareness of the tactical situation. In support of the general experimental process, control principles of experimental design were followed to minimize differences in testing conditions between operators, and standard ethical procedures to protect human participants were followed.

1.1 APPROACH

In field studies of complex military operations, collecting averages of performance across entire scenarios often yields little useful information about how well operators or teams of operators actually performed. A particularly effective measurement scheme that has been used in similar investigations (Pharmer, Campbell, & Hildebrand, 2001; Osga, et al., 2002) is the event-based measurement approach (Johnston, Cannon-Bowers, & Smith-Jentsch, 1995). The first step in this process is to identify the most critical and the most common events required to successfully use the system being evaluated, and to identify how the system may or may not support the operator in responding to these events. The second step is to identify and develop critical events in the scenario, which provide opportunities for operators to perform the tasks associated with these events. Once these critical events are identified and developed, timing and accuracy performance measures of participant actions are tied to them. Careful design of the scenario by subject-matter experts (SMEs) is crucial when using this approach and determining the sequences of actions, potential errors, and acceptable timing for performing the actions associated with the scenario events. To be most effective, the scenario must include events with a range of criticality levels so the operator is not faced solely with high-priority tasks and instead has an opportunity to make tradeoffs.

Timing and accuracy data provide limited information on the outcome of tasks and some clues to the efficiency of operators' perceptual and cognitive processes. A broader measurement approach is required to more fully explore the impacts of the prototype design on these processes. For this investigation, a four-pronged approach was taken to investigate the impacts of the LACS prototype HCI on both task outcomes and processes. First, performance outcomes were measured to determine

---

[1] ONR Code 31 Knowledge Superiority & Assurance (KSA), LACS Decision Support Capabilities; ONR Code 34 Capable Manpower (CM), Task-Centered HSI & Training Capabilities; Principal Investigators: David Kellmeyer & Dr. Glenn Osga, SSC San Diego.

whether operators could meet mission requirements effectively using the prototype. These results are presented in Section 3.1. Second, individual workload was measured using both subjective and objective measures; analysis and discussion of these measures are presented in Section 3.2. Third, the situational awareness (SA) of the operator was assessed through probe questions based on the normal exchanges between the operator and supervisor. Evaluation and assessment of the operator's SA are presented in Section 3.3. Fourth, levels of trust in the automated decision support systems were measured to identify potential issues related to human–automation interaction with task-managed systems. The results, along with other comments from each operator, are addressed in Section 3.4. Finally, human performance was evaluated based on team process measures[2], primarily through subjective ratings of performance across several dimensions shown to be indicative of highly effective teams. Analysis of the team factor is presented in Section 3.5. Before these analysis sections, a detailed description of the test method is provided.

---

[2] This first-year demonstration of the prototype collected data from individual operators manning a single LACS console. However, team process measures were also collected by viewing the individual operator as part of the larger team consisting of the Engagement Control Officer (ECO), Land Attack Coordinator (LAC), and Tomahawk Strike Coordinator (TSC).

2. METHOD

2.1 SYSTEM DESCRIPTION

The prototype assumed fully automated system capabilities for Over-the-Water route planning, mission-to-missile cell allocation (CA), crypto, etc., in which presumably robust and correct algorithms provided decision recommendations and draft reports to the operator, who then could review and either approve or disapprove them. In the scenario, the operator's only recourse if he did not approve was to mention the concerns and alternative suggestions to the Engagement Control Officer (ECO). While useful for prototype testing, such a high level of automation is not indicative of current fleet practice, where automation is generally distrusted, and all levels of operators and supervisors typically desire the ability to drill-down through any system-generated data and edit or override the system suggestion. During training for this evaluation, participants were told to generally trust the system automation in the prototype. Communications, with the exception of limited voice communication with the role-player as described below, were assumed to be digital and were processed by the system.

The prototype HCI used in this test received Electronic Strike Packages (ESPs) containing mission planning information with target and route details. The operator had to review the plans, review and send system-generated reports, authorize the system to execute the engagements, handle failures, and keep the ECO informed. As is standard procedure, the ECO acted as the interface between the operator and all higher ranking officers on the ship for information and approval purposes. All system communications and inputs were simulated by the prototype. Verbal communications were between the ECO and the operator only.

2.2 SCENARIO

A goal of this investigation was to use a scenario with realistic tasking and environmental elements but which also placed substantially higher taskload and time pressure on the operator than would be reasonably expected in real-world operations. An eight-hour, four-operator scenario was used as the basis for constructing the scenario simulated by the prototype. That scenario's events were compressed into a two-hour scenario for execution by a single operator. This compressed schedule resulted in frequent simultaneous taskings for the operator across the various phases of five ESPs, each with a variety of critical events as described in the following section. Such simultaneous operations are comparatively rare in current fleet Tomahawk systems, but are expected to become more and more common as systems become more complex and manning is reduced. SMEs assisted in developing the scenario and establishing performance deadlines associated with scenario subtasks. Using these deadlines, investigators were able to determine whether or not specific operator actions were performed early, on time, late, or not at all.

2.2.1 Scenario Tasks

The five ESPs of the scenario included tasks for pre-launch planning, execution of the launch, and post-launch monitoring (for the BLK IV missiles). Specifically, the scenario required the operator to plan and execute engagements by performing instances of the following tasks:

- Respond to ESP taskings, Call-for-Fire (CFF) taskings, and Mission Data Updates (MDUs)
- Review and send Validation Reports
- Review and send Strike Coordination Overlays (SCOs)
- Respond to pre-launch system failures such as a Hatch Failure, Digital Scene Matching Area Correlator (DSMAC) Failure, etc.

- Conduct Final Review
- Execute launches
- Review and send Post-Launch Reports (PLRs) and Post-Strike Reports (PSRs)
- Monitor, Flex, and Redirect in-flight missiles
- Respond to post-launch system failures such as Missing or Bad Health & Status (H&S) messages
- Respond to questions and requests for information from the ECO

Over the course of the scenario, the operator handled approximately 80 individual task events from the above categories, distributed unevenly over the two-hour timeline. Some of the tasks required immediate attention, while others had larger time windows for successful completion.

Figure 1 provides an overview of the operator's task events and the approximate time-windows available for successful completion of each task. Note that the SA probes from the ECO are excluded from this figure. Even so, the figure provides insight into the number of simultaneous tasks the operator may have been handling during the scenario.

2.2.2 Scenario Observers

During each scenario run, the SME evaluator role-played the position of ECO and provided "filtered" inputs from the Land Attack Coordinator (LAC) and Tomahawk Strike Coordinator (TSC). The use of an SME in this position provided additional realism to the scenario. To investigate the participant's (i.e., operator's) level of SA, the role player asked the participant questions at pre-specified times or events throughout the scenario. The role player was also in a position to observe the operator's actions and to record notes about how tasks were accomplished. Training was covered in a single scenario run-through to familiarize the role player with the scenario and probe questions, and to experience likely participant actions.

In addition to the role player, there were several observers who were not a part of the direct interaction with the participant. The task of these observers was to periodically ask the participant for a workload rating. In addition, they recorded any important or unusual aspects of the operator's actions and comments, and monitored the video recording of the evaluation.

2.3 PARTICIPANTS

Eleven Fire Controlmen participated in this investigation. Six participants were from the Pacific Fleet, and the remainder represented the Atlantic Fleet. All participants on the west coast were Fire Controlmen First or Second Class; the east coast participants were Fire Controlmen Third Class. A prototype failure resulted in dropping one Pacific Fleet participant's data from the analysis. All participants were experienced Tomahawk operators, either with engagement planning and launch control qualifications or recent operational firing experience in Operation Iraqi Freedom (OIF). (The OIF participants where operationally experienced with the currently fielded Advanced Tomahawk Weapon Control System (ATWCS), which does not include the post-launch functionality of TTWCS. The participants with the training qualifications were TTWCS operators.).

Send Validation Message
initiate sel / power / align
accept CA recommendation
review failure
Send Validation Message
Review CA (optional)
Initiate sel / power / align
approve / send SCOs
Send Validation Message
Send Validation Message
accept CA recommendation
Approve Final Review
Authorize execution ESP 0001
Send Validation Message
accept CA recommendation
Authorize execution ESP 0001
Initiate sel / power / align
Send SCOs
Authorize execution ESP 0001
Send Validation Message
Send Validation Message
approve MDU mission
Authorize execution ESP 0001
Send Validation Message
accept CA recommendation
Accept CA (optional)
approve sel / power / align
Approve Final Review
Approve / send SCO
Send PLR
Authorize execution E060
Send PLR
Approve Final Review
Send Validation Message
Authorize execution E068
Accept CA (optional)
Send PLR
Approve Final Review
Authorize execution ESP 0002
Approve CA
Send Validation Message
initiate sel / power / align
approve sel / power / align
Authorize execution ESP 0002
Send SCO
Send SCO
approve Final Review
Operator Approve Final Review
Authorize execution E070
Send PLR
authorize execution E062
Send PLR
Send PSR
Send PLR
Request system rec'dations
Rec'md flex to ECO
instruct system create IMMM
Approves Flex IMMM
Send Validation Msg
Send PSR
Approves / Sends Aimpoint Update IMMM
requests "restrike" options
Rec'md redirect to ECO
instruct LACS to create redirect solution
initiate retarget IMMM process
Approve Retarget IMMM
Send PSR

Time axis: 0:00:00, 0:05:00, 0:10:00, 0:15:00, 0:20:00, 0:25:00, 0:30:00, 0:35:00, 0:40:00, 0:45:00, 0:50:00, 0:55:00, 1:00:00, 1:05:00, 1:10:00, 1:15:00, 1:20:00, 1:25:00, 1:30:00, 1:35:00, 1:40:00, 1:45:00, 1:50:00, 1:55:00, 2:00:00, 2:05:00, 2:10:00, 2:15:00, 2:20:00
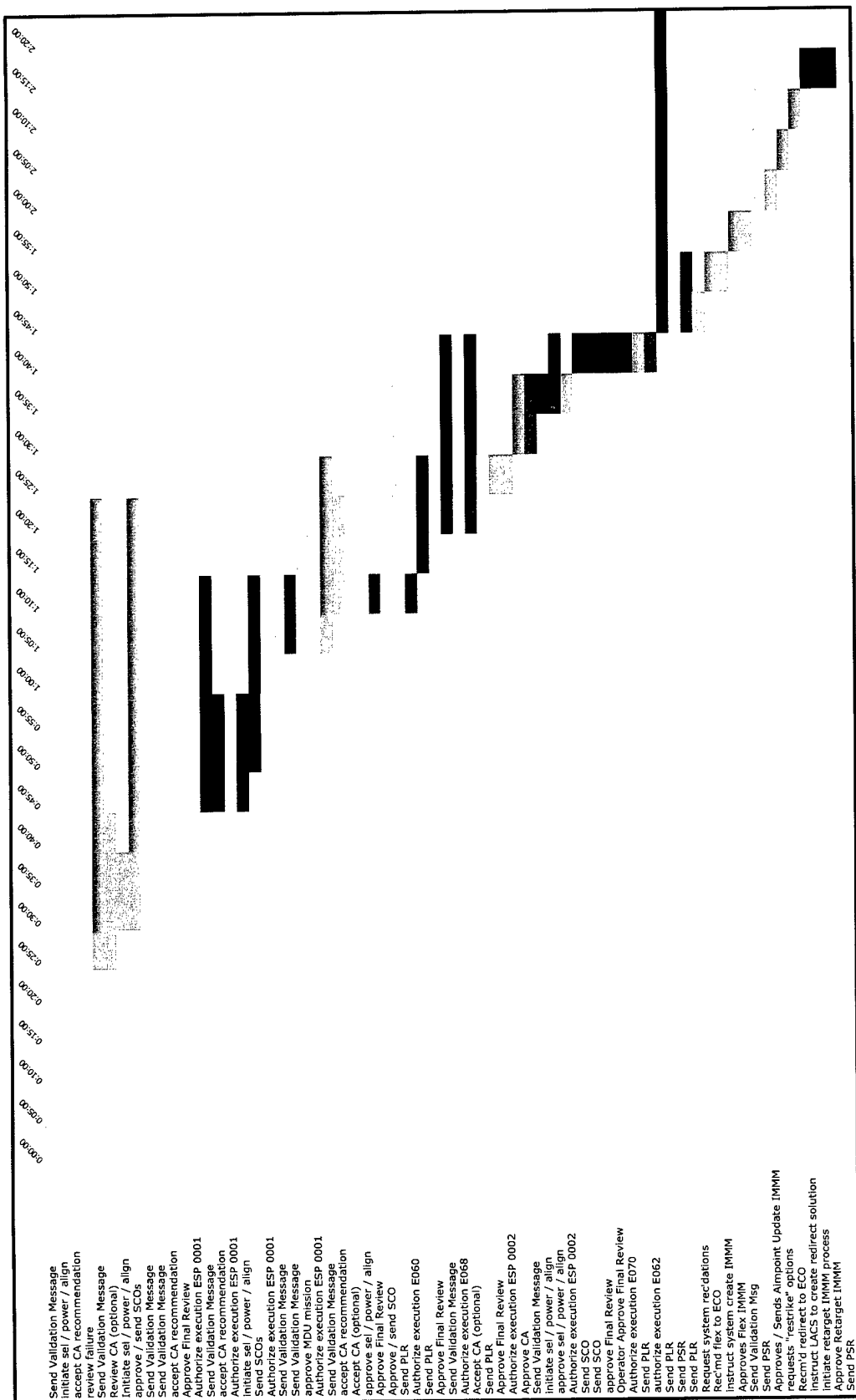
Figure 1. Time windows for successful completion of operator task events, color-coded by ESP (excludes SA probes by ECO). Time is presented in five-minute intervals.

5

Method

2.4 PROCEDURE

Testing on each coast was conducted over four days during a single work week. The first day included administrative setup and a collective training session, and the following three days included the individual testing sessions.

2.4.1 Pre-Test Administrative Work and Training Session

On Monday of the testing week, all operators met to complete paperwork and to be trained in the use of the LACS prototype HCI. The operators were informed of the purpose of the testing and what they would be asked to do during the simulated scenario. The paperwork packet collected the following information: (1) consent forms for both doing the test and being video- and audio-taped, (2) biographical information, and (3) experience information including training, Tomahawk Weapon Control System (TWCS) qualifications, and computer experience. Each participant was then assigned a participant number and testing session time.

To focus the investigation on the effectiveness of the HCI to support operators' performance, training was targeted toward the displays, navigation controls, and general "buttonology" of the LACS prototype. The goal was to minimize the tactical guidance given to the participants, particularly how to prioritize simultaneous taskings within and between ESPs. Some tactical training was necessary, however, because the prototype HCI was intended to support a future weapons system and included some concept-of-operations (CONOPS) changes relative to the current TWCS systems. In total, the participants were provided approximately three hours of instruction and 35 minutes of hands-on training.

The training administrators introduced the participants to the prototype system screens and presented the CONOPS changes. The trainers then demonstrated the use of the prototype by narrating a condensed training scenario. This training scenario contained most of the task and event types discussed earlier, but its duration was kept to approximately 35 minutes through reduction of the tasking level and acceleration of several system performance times. After the training scenario narration, operators gained hands-on experience with the system by executing the same training scenario themselves. During their hands-on session, the operators were encouraged to ask questions and report events appropriately to the ECO. Three single-operator systems were available for training on each coast. On the west coast, two participants executed their hands-on training session immediately before their testing session, rather than on the training day.

2.4.2 Test Administration

The length of the test scenario was just over two hours, followed by the post-test questionnaire and interview. During the test, the operator had to handle ESPs, reports, failures, launches, and monitoring of missiles in addition to keeping the ECO informed. The ECO also asked the operator a series of SA probes, and both the ECO and the trained observer made notes of participant behavior.

2.4.3 Materials

The computer system consisted of a Windows®-based personal computer driving two monitors in a top/bottom layout, which reflects the expected TTWCS operating environment in 2006. Each monitor was set to 1280x1024 resolution. A mouse or trackball was used for operator input. The operating system for the west coast was Windows 2000®; for the east coast, it was Windows XP®. The LACS prototype software, which was also called the Rapid Prototype (RPT), was built in Macromedia® Director® by SSC San Diego. The RPT recorded the time and details of scenario events and operator actions.

Lighting was subdued in an effort to mimic the Combat Information Center (CIC) environment, but was brighter than the typical CIC to support videotaping. On the west coast, a video mixing system was used to record the video feed of the Task Manager screen and an inset image from a video camera of the participant with a time stamp in the background. The video also included audio recordings made with lapel microphones on both the participant and the ECO. On the east coast, a single camera shot of the Task Manager screen was recorded over the participant's shoulder, with audio recorded by the camera microphone. Both systems were used to record participant debrief interviews after the scenario.

2.5 MEASURES

2.5.1 Performance Outcome Measurements

2.5.1.1 Task Latency

Task latencies are defined as the amount of time elapsed from a triggering event to the operator performance of a particular task. Operator tasking and scenario context largely determine the specific tasks to which measurements of this type are appropriate. For example, it may be critical that some tasks be performed as quickly and accurately as possible. For others, it may only be important that the tasks are performed within a certain window of opportunity, which for this test was defined in advance by an SME. Further still, there may be a variety of task sequences that operators can perform and still meet mission tasking.

Throughout the scenario, the prototype recorded times for the start or triggering event, operator response, and operator completion. Triggering events were defined as the time when the HCI indicated to the operator that action was needed. Operator responses were defined as the time when the operator selected the indicator. Operator completions were defined as the time when the operator performed the final HCI interaction for that task. Latency was measured from trigger to completion. The analysis of latency required a thorough understanding of the scenario and fleet-accepted processes and exceptions. This understanding was established with the support of SMEs.

The SME estimates for acceptable time windows were based on the following typical series of steps through the chain of command:

- Message is received by operator
- Operator looks at message
- Operator understands message
- Operator explains to ECO or requests action approval from ECO (or further up the chain of command)
- ECO concurs
- Operator executes action on the system

Table 1 presents descriptions of the tasks and benchmarks for latencies provided by the SMEs as criteria for on-time operator performance, based on the prototype, scenario, and fleet doctrine. Tasks for which completion times include the chain of command shown above are indicated by "(chain)" after the times.

The "Send Validation Report" task completion time assumed the system performed an automated validation and provided the operator with a "No Exception" report. A "No Exception" report lets the tasking authority know the ship can handle the assigned tasking. In contrast, an "Exception" report alerts the tasking authority that the ship cannot handle the tasking and it must be reassigned to another platform. It is important that the tasking authority be informed as soon as possible of the ship's capabilities so reallocation and planning can occur in a timely manner; not receiving a timely

indication is equivalent to receiving an "Exception" report. In this scenario, the system generated only "No Exception" reports for the operator's review and approval.

Table 1. Task latency performance measurement criteria.

Task	Operator Task Completion Action	SME Completion No Later Than (NLT) Time	Doctrine Completion NLT Time[See Table Notes]
Send Validation Report	"Send report"	Trigger event + 30 seconds (chain)	
Send SCO	Approval of Plan to SCO accept	Time of Launch (TOL) – 20 minutes	ASAP and NLT 2 hours prior to first estimated TOL (ref. 1, see table notes)
Select/Power/Align Missiles	"Initiate" s/p/a	Trigger event + 1 minute	
Respond to DSMAC Failure during Powering/Alignment	"Accept" system recommended solution	System recommendation + 30 seconds (chain)	
Do Final Review	"Approve" final review	First TOL - 1 minute	
Execute Missile Launch	"Authorize" execute	Scheduled TOL + 1 minute	NLT X hours (ref. 2, see table notes) (X classified)
Send Post-Launch Report (per ESP)	"Send report"	Last TOL + 20 minutes	NLT 20 minutes after salvo complete (ref. 1, see table notes)
Send Post-Strike Report (per ESP)	Send report	Last Time on Target (TOT) + 20 minutes	NLT 20 minutes (ref. 3, see table notes)
Failure to Transition to Cruise (FTC)	Approves launch of backup missile	FTC time + 1 minute	
Notes: Doctrine Completion References: (1) CTF 60 Standing LAC Intentions Serial 10, 27 Dec 02 (2) NAVY WIDE OP TASK TLAM, 2nd FLEET (3) Exercise Requirement, from interview with a TWCS specialist. (4) "NAVAL SURFACE FIRE SUPPORT REQUIREMENTS FOR OPERATIONAL MANEUVER FROM THE SEA – 1999," Marine Corps NSFS Letter to Navy			

The "Send SCO" task completion time was based on the Time of Launch (TOL) of the first missile in the strike package. It assumed the system provided the operator an acceptable report to review, approve, and send. In this scenario, there were no unacceptable reports, although the results section will address one report provided late by the system.

The "Select/Power/Align Missiles" task completion time assumed the system prompted the operator to begin powering the missiles after validating the strike package. The operator had one minute from the time the button on the Task Manager screen turned white (trigger event) to enter the powering screen by selecting the button (response) and start powering (completion).

The "Respond to DSMAC Failure During Powering/Alignment" task completion time assumed the system would automatically provide the operator a recommended course of action. The operator had

to review the recommendation and inform the ECO before acknowledging the system notification. In this scenario, the missile suffered a DSMAC failure on a GPS-only mission, so acknowledgement of the failure was the only action required; however, the Results section will address some operator issues with this task.

The "Do Final Review" task completion time was based on the TOL of the first missile in the strike package. It assumed the system provided the operator an acceptable report to review, approve, and send. In this scenario, there were no unacceptable reports.

The "Execute Missile Launch" task completion time assumed the system prompted the operator in advance of the scheduled missile launch time and counted the execution as on-time if it occurred up to one minute past the scheduled time. In this scenario, a CONOPS change allowed for multiple execution approvals at once; at the scheduled TOL minus two minutes for the next un-executed missile, the operator was prompted to approve the missile for execution, along with any other missiles scheduled for launch in the next eight minutes following. With this CONOPS change, the operator was prompted to execute the majority of launches well in advance of their latest on-time window.

The "Send Post-Launch Report" task completion time was based on the last TOL for an ESP and assumed the system provided the operator an acceptable report to review, approve, and send. In this scenario, there were no unacceptable reports, although several operators noticed that the Max Follow-on Capability numbers in the prototype were static.

The "Send Post-Strike Report" task completion time was based on the last Time on Target (TOT) for an ESP and assumed the system provided the operator an acceptable report to review, approve, and send. In this scenario, there were no unacceptable reports, although due to long flight times, some did not occur until after time recording had stopped.

The "Failure to Transition to Cruise (FTC)" task completion time assumed the system provided the operator with an acceptable recommendation for a replacement missile to launch. In this scenario, the system was automatically informed of the FTC.

Several tasks an operator would typically handle were not included in the Task Latency measures. "Allocate Missiles" was an automated process in the RPT, and the operator was instructed to trust it so there was no RPT interaction required. "Hatch Failure" was handled automatically by the system because the cell with the failure had an Auto-Ready Spare; the operator made the ECO aware of the failure, and was asked an SA probe about it (discussed below, Section 2.5.3). In this scenario, "Call-For-Fire" launches were treated as defined-time events, thus their results are included in the "Execute Missile Launch" task. "Bad H&S" did not require an operator interaction with the HCI, so no completion action time was recorded. "Flex Inflight Missile" did not have its times recorded by the RPT. "Analyze BDII" (Battle Damage Indication Imagery) is not a typical task for currently deployed Tomahawk systems and prototype-recorded times were unclear in the east coast data, so it was also not analyzed.

2.5.1.2 Task Accuracy

Task accuracy addressed both objective evaluations (was an action performed?) and subjective evaluations (was it performed properly?). Data for the former were provided by the RPT data logging. The latter relied on notes taken by the SME ECO role player and additional observers.

Analysis for the objective data focused on two general categories of errors: omission and commission. First, errors of omission could be demonstrated if operators did not perform tasks that were required in the scenario. Errors of commission could be demonstrated when operators explicitly performed an action that was either inappropriate or unnecessary in the scenario context.

Subjective evaluations were based on observations of operator actions and SME evaluation of the answers provided to SA probe questions (Section 2.5.3) and of the teamwork measures (Section 2.5.5).

2.5.2 Workload Measurements

Workload was assessed primarily by two measures. The first measure was a periodic self-evaluation prompted every five minutes during the scenario. The second measure, administered after the scenario, was based on the National Aeronautics and Space Administration (NASA) Task Load Index (TLX) survey. When the participants were presented with multiple tasks, their task completion prioritization method was also evaluated. These measures will be described more fully in the following sections.

2.5.2.1 Moment-to-Moment Workload Rating

The moment-to-moment workload was collected throughout the scenario and consisted of probing the participants at five-minute intervals for a single rating of overall perceived workload for the past five minutes on a 1 to 7 scale. The prototype provided an audible reminder every five minutes, which was to prompt the data observer to ask the participant to rate his workload. The meaning of "workload" for this investigation and the low and high anchors of this scale were explained during the training session and the training scenario run.

2.5.2.2 NASA TLX Workload Rating

The second measure of workload was the well-validated NASA TLX (Hart & Staveland, 1988), administered post-scenario to each participant (see Appendix 1). This modified version of the index has been used successfully in previous investigations of HCI effectiveness performed by Naval Air Systems Command (NAVAIR) Orlando, SSC San Diego, and the Naval Surface Warfare Center Dahlgren Division under the Office of Naval Research sponsored Surface Combatant 21 (SC-21) Manning Affordability Initiative (Pharmer, Campbell, & Hildebrand, 2001; Osga et al., 2002). The index was modified to include four team workload dimensions in addition to the six individual task workload dimensions of the NASA TLX (see below). While the current study was focused on a single operator performing the scenario, there were important teamwork aspects of the task that required coordination with the role player.

Task Load Index (TLX) Dimensions:

- Effort
- Performance
- Frustration
- Temporal Demand
- Mental Demand
- Physical Demand
- Communications Demand (team)
- Monitoring Demand (team)
- Control Demand (team)
- Coordination Demand (team)

2.5.2.3 Prioritization Methods

Finally, the scenario included both high- and low-criticality tasks. By examining the participant's prioritization decisions at moments of higher workload, this quasi-secondary measurement technique

provided further useful information in post-analysis to understand what aspects of the scenario and the HCI drove the participant's workload.

2.5.3 Situational Awareness and Assessment Measurements

Situational Awareness has gained a considerable amount of attention in the training community. It can be defined as "the perception of the elements in the environment within a volume of time and space" (Endsley, 1988), the comprehension of their meaning, and the projection of their status into the near future. As such, three levels of SA were considered for this investigation:

- *Perception* is the first level of SA and focuses on the individual perceiving (recognizing) the important elements within the environment. These elements are domain specific and have a vital role within the task.
- *Comprehension* is the second level of SA and focuses on understanding the significance of certain cues and events within the task.
- *Projection* is the third level of SA and deals with making predictions of future events based on perceiving and comprehending vital cues and events within the environment.

For this investigation, SA of the operators was determined through two means. First, the role player questioned participants at appropriate times during specific events in the scenario. These probe questions focused on the participant's levels of perception, comprehension, and projection of certain cues within the scenario. Care was taken to ensure the timing of the questions did not alert the participant to events and affect SA. Table 2 lists the full set of probe questions used for the data collection, including the approximate scenario times when they were asked and the associated SA level.

A second approach for investigating SA was through performance-based inferences of operators' actions in response to scenario events. These can be informative about how the operators perceived the event, comprehended its meaning, and were able to project the consequences of the event into the future. Patterns of long delays in recognizing or acting on particular events are often indicative of SA problems.

2.5.4 Human–Automation Interaction Measurements

Understanding how operators perceive their interaction with a HCI is an important factor in designing usable systems. When operators are moved into a supervisory-control role due to increased automation, the potential problem of automation complacency arises. The concern is that operators lose "big picture" SA and respond to tasks without a deeper comprehension of the situation. Coupled with a high level of workload and simultaneous events, operators may begin to "shed" tasks by dispatching them without fully reviewing the automation recommendations. This is commonly seen in the current ATWCS/TTWCS systems, as well as many other military systems, as operators cycle through and dismiss alerts. This phenomenon has been termed the "clean plate syndrome" because operators may endeavor to attain a "clean" display with no operator-actionable items. In such a case, there is an increased risk of a disconnect between the operators' perception of the situation and action; they press buttons because the buttons indicate they should be pressed, not because they understand what pressing the button will do. It should be noted that having no operator-actionable items is a desirable end-goal, but it must be achieved by properly addressing each item.

Table 2. Full SA probe list.

SA Probe Question	SA Level	Scenario Time
What are the estimated times of first and last launches, and their associated Engagement numbers for ESP 001C?	1	0:04:45
What is the estimated time 1st missile mode 7 and its associated plan?	1	0:08:20
Report when 1st missile mode 7/all missiles mode 7.	1	0:10:45
Report current tasking and max follow-on salvo capability	1	0:19:00
Report when Call for Fire (CFF) execute received?	1	0:40:00
What is the estimated earliest time of launch for CFF 1?	3	0:43:00
Report 1st /last launch time for ESP0002A and their associated engagements.	1	0:43:40
Is DSMAC failure mission critical?	2	0:46:00
What is launch time earliest for CFF?	3	0:53:45
Report time of hatch failure.	1	0:56:00
How many engagements left in ESP0001C?	2	0:59:00
Report tasker for CFF engagement.	2	1:03:00
How many missions are in the MDU and can we meet tasking?	3	1:04:00
Report H & S status of CFF1 ESP 0004B.	2	1:13:00
What is TOT for CFF 1?	3	1:16:00
What is TOL for CFF2?	3	1:23:00
What is engagement # and estimated TOL for ready-spare (R/S)?	3	1:25:00
Which ESP is Health & Status (H&S) In-flight Mission Modification Message (IMMM) from E031 associated with?	2	1:35:00
Report max follow-on salvo capability of ESP0002B	3	1:39:00
Why was the flex necessary?	2	1:47:00
How many BLK IVs are in flight?	2	1:51:00
What is TOT for all active ESPs?	2	2:02:00
Report BDI for MSN 0046 ESP 0002C.	2	2:04:00
What is loiter target and estimated time of loiter exit?	3	2:09:00

The task-centered design of the LACS prototype HCI was intended to enable an operator to supervise the system by indicating to the operator if the system detects a problem or produces a recommendation for approval. To date, a relatively small amount of data have been collected to understand how operators perceive their interaction with such supervisory control systems involving automated production of decision recommendations, draft reports, etc. However, the literature on automated systems is replete with examples of operators placing unfounded levels of trust in these systems only to be surprised when things go wrong and SA must quickly be regained. Consequently, it is important to understand operators' reactions to the LACS prototype HCI and to record positive and negative interactions with the system to support consideration of design alternatives.

A post-scenario "Human–Automation Interaction" questionnaire (consisting of a Likert scale and both multiple choice and fill-in-the-blank questions) was administered as an initial attempt to gain an

understanding of operator perceptions of trust in and the reliability of the information provided by the prototype (see Appendix 2). During training, the operators were instructed to trust the automation, and only two or three instances of erroneous automation were (inadvertently) included in the scenario. Therefore, operators' ratings of their level of trust may reflect their previous experiences with other fleet systems, not just their interaction with the prototype.

2.5.5 Team Process Measurements

As stated previously, the current investigation focused on a single operator performing on a single console. However, there were team aspects of the tasks that the operator was required to perform with the ECO. Future demonstration evaluations may focus on multiple operators. Consequently, the following section describes the approach to measuring important team processes in the current investigation.

In the Tactical Decision Making Under Stress (TADMUS) research sponsored by the Office of Naval Research, four major factors emerged as critical to effective team performance (Smith-Jentsch, Johnston, & Payne, 1998): Information Exchange, Communication, Supporting Behavior, and Initiative/Leadership. These factors have been incorporated into a rating scale known as the Air Warfare Team Observation Measure (ATOM; See Appendix 3). A fifth factor, Critical Thinking, was recently added to the ATOM to evaluate the ability of teams to generate, evaluate, and test hypotheses during scenario performance. While the name of this tool implies that it is specific to air warfare, the scales are generic and have been used successfully to assess teamwork across a variety of domains.

Table 3 lists the scales (dimensions) and subscales (behaviors) included in the ATOM. After completion of the scenario, the SME role player rated the participant on a scale of 1 (weakness on the dimension) to 5 (strength in the dimension) for each of the teamwork subscales. The process of rating required very little training and very little time to complete, but it was essential that an individual with expertise in the domain complete the form.

Table 3. ATOM Scales.

Team Dimension	Behaviors
Information Exchange	Seeking Sources Passing Information Situation Updates
Communication	Proper Phraseology Completeness of Reports Brevity Clarity
Supporting Behavior	Error Correction Providing/Requesting Backup and Assistance
Initiative/Leadership	Providing Guidance Stating Priorities
Critical Thinking	Hypotheses Evaluation Testing Hypotheses

3. RESULTS

3.1 PERFORMANCE OUTCOME RESULTS

3.1.1 Task Latency

As mentioned previously, this investigation was not an empirical comparison between a legacy system and the prototype HCI. As such, the criteria for task latency were provided by a SME, who in many cases referred to fleet doctrine and other sources for appropriate time windows. Because this investigation focused on the prototype HCI, several of the SME estimates were conservative "best guesses." Table 4 provides the latency data for the task performance measures in the investigation. Operators performed the vast majority of their tasking in a timely fashion, although some operators did perform some tasks late with respect to SME-provided criteria. These results extended not only to reports (e.g., validation reports) but also to recognition of some faults (e.g., DSMAC fault). The table lists the number of measured events for each task, and the percentage of instances on-time and late based on the SME time windows (see Table 1).

Table 4. Task latency criteria and percentage on time.

Task Event	Overall Number of Events	Percentage On Time (%)	Percentage Late (%)	Average Time, Trigger To Completion	Standard Deviation
Send Validation Report	120	53.7	46.3	0:49	1:00
Send SCO	44	75	25	1:03	0:47
Select / Power / Align Missiles	55	60	40	1:09	1:03
Respond to DSMAC Failure During Powering/Alignment	11	45.5	54.5	6:08	7:43
Do Final Review	63	92	8	0:40	0:48
Execute Missile Launch	312	99	1	0:33	0:27
Send Post-Launch Report	62	100	0	1:04	1:03
Send Post-Strike Report	23	92	8	0:55	1:01
Failure to Transition to Cruise (FTC)	10	50	50	1:04	0:34

The on-time percentage was low for "Send Validation Report," but the degree to which participants were late was relatively small (on the order of seconds). While participants technically sent these reports late 46% of the time, it is possible that the SME-generated criteria may have been too conservative for this measure. Consequently, it is recommended that some attention should be

paid to improving the capability to generate these reports within the prototype HCI, and future investigations should focus on refining the lateness criteria as well. The shorter window was more crucial for sending "Exception" reports than the "No Exception" reports encountered in the scenario. However, one observer noted that the operators typically opened the reports and then sent them, as opposed to evaluating their status before sending. Therefore, it is possible that the delay is not due to operator evaluation of criticality, unless they presumed that an "Exception" report would have been color-coded non-white to indicate its critical nature.

While the "Send SCO" results indicate that the operators were late one-quarter of the time, the scenario only contained four "Send SCO" task events, and on one occasion (ESP 3, the MDU), the system did not provide the operator an opportunity to send the SCOs until approximately 10 minutes before launch, instead of the SME No-Later-Than time of 20 minutes. However, 90% of the operators sent that SCO within two minutes of the system providing the opportunity.

The "Select/Power/Align Missiles" task shows low on-time results. However, like the "Send Validation Report" data, the degree to which operator actions were late was on the order of seconds. In addition, the criterion time window did not address differences in time required for different missile types. Future evaluations should address this issue, as the differences in powering times between BLK III and BLK IV missiles is significant, and it is possible that operators may have taken these differences into consideration in determining when to power.

The results in the DSMAC failure, the only event participants handled late more than half of the time, can most likely be attributed primarily to the participants' lack of knowledge about DSMAC. Typically, the ECO had to explain it to them before they could acknowledge it. Additionally, if they did not immediately respond to the DSMAC failure by pressing the "ACK" button, participants later found it unclear why the button was white. This indicates an improvement is needed in addressing non-mission critical faults.

The traditional measure of successful TWCS operation is on-time execution of missile launches. By this measure, operators had a 99% success rate, with 309 out of 312 missiles executed no later than one minute past their scheduled TOL. This result can be attributed to the task-centered design of the system, which kept the operator informed of when actions needed to be taken, and to the CONOPS changes incorporated in the prototype HCI (e.g., executions were grouped into batches so the operator did not have to handle each one individually). The three late executions, one each by three different participants, were at times in the scenario with multiple simultaneous events requiring processing by the operator, including answering questions from the ECO. Two of the late executions were on an operator's very first execution, and were late by less than half a minute; it may be that the operators had not yet fully explored and understood the prototype HCI, because the rest of their executions were on time. Nevertheless, these late executions do indicate that the prototype HCI still has room for improvement in task prioritization through better attention management. This is similar to the findings of the Multi-Modal Watch-Station 2000 Air Defense Warfare evaluation (Pharmer, Campbell, & Hildebrand, 2001).

The data tend to indicate that the operators noticed tasks soon after they became available and handled them quickly. It is likely that the operators were not merely clearing tasks but were well aware of the actions they were taking, as indicated by the much longer completion time for the DSMAC failure. This was an unfamiliar task for most of them, and their discussion with their ECO before completing the task may imply that their quicker response times elsewhere were on well-understood tasks. It is also noteworthy that the shortest average completion time was for missile launch execution; this indicates that white buttons were not all equal to the operators, who were evidently making evaluations of which to handle first.

3.1.2 Task Accuracy

As indicated in Table 4, the time-critical operator actions had very high rates of successful on-time completion. The RPT did not provide many opportunities for operator error to occur if the operator responded when prompted by the Task Manager and followed the recommendations of the automation. Thus, the analysis shows that the majority of operator actions were done within the SME recommended time window. The late responses reflect a combination of differences in reporting practices between participants and lack of the necessary experience to properly report events using the standard phraseology.

It is important to note that there were no pure errors of omission by the operators. While some events may have been completed late, there were no critical operator actions that were completely missed. This point highlights the capabilities of a task-managed system. In a traditional design, it is up to the operator to remember what steps have been completed for which strike package. In the task-centered design of the RPT, the system kept track of step completion and provided reminders to the operator of steps that still required attention. The operator was always able to see what else needed to be addressed, instead of having to maintain either a hand-written or mental list.

However, there were a few instances of what could be called "operational omission" errors, including not sending reports on time and executing launches late. Such errors occur when the operator completes a task so late that either it would not permit successful system completion of the task, or the time taken is outside the doctrine time window criteria. As shown in the previous section, there were very few of these events. Observations of the operators during the scenario indicate that these instances were generally due to a failure to maintain full awareness of strike timeline progression across simultaneous ESPs. On these occasions, operators appeared to not focus on the next most critical event and would instead become distracted by either trying to answer questions from the ECO or handling other, less critical events. This is another indication that instances of simultaneous tasking and high workload may require additional attention management to achieve even better reliability.

Errors of commission included sending reports before clearing them through the ECO and allocation and alignment of missiles without the ECO permission. These errors could not be identified by the RPT data collection mechanism and were subjectively determined through ECO feedback. The number of these errors appears to have been very low and appeared to be due to the participants' over-enthusiasm. They usually occurred on less-critical events such as sending reports, instead of on crucial events like executing missiles, which indicates that the operators were aware of the need for approval on important events. These are errors that, short of requiring the ECO to perform the action or requiring undue HCI confirmation, would be difficult to eliminate. However, future analysis could look at the relationship between such errors, the "clean plate syndrome" (described in Section 2.5.4), and operator judgment about which tasks need approval given the current workload of the team.

3.2 WORKLOAD RESULTS

3.2.1 Moment-to-Moment Workload

To measure the moment-to-moment workload, the observer would prompt the participant every five minutes to indicate a level of workload over the previous five-minute time window. To assist the observer, the RPT would generate a tone every five minutes. All of the participants, at some point during the scenario, began indicating their workload in response to the tone, obviating the prompt from the observer. However, in some instances the momentary workload was high enough that the observer did have to ask the participant for his workload. This highlights that operators reported

17

aggregate workload ratings for the time windows (as instructed), not the momentarily high workload levels at the five-minute point.

Figure 2 shows the average moment-to-moment participant ratings of workload across the scenario as compared to the actual taskload, which was defined as the number of tasks that operators should perform in the prior five-minute interval. The number of tasks was calculated by adding up the number of SA probes and the number of events requiring action by the operator (e.g., sending an SCO). As the figure shows, average operator ratings followed closely to the actual scenario taskload. In fact, the two factors were highly correlated (r = 0.74). This serves as a good indication that the moment-to-moment ratings appeared to be reliable across the scenario and that the operators had a good level of awareness of their workload during the scenario.
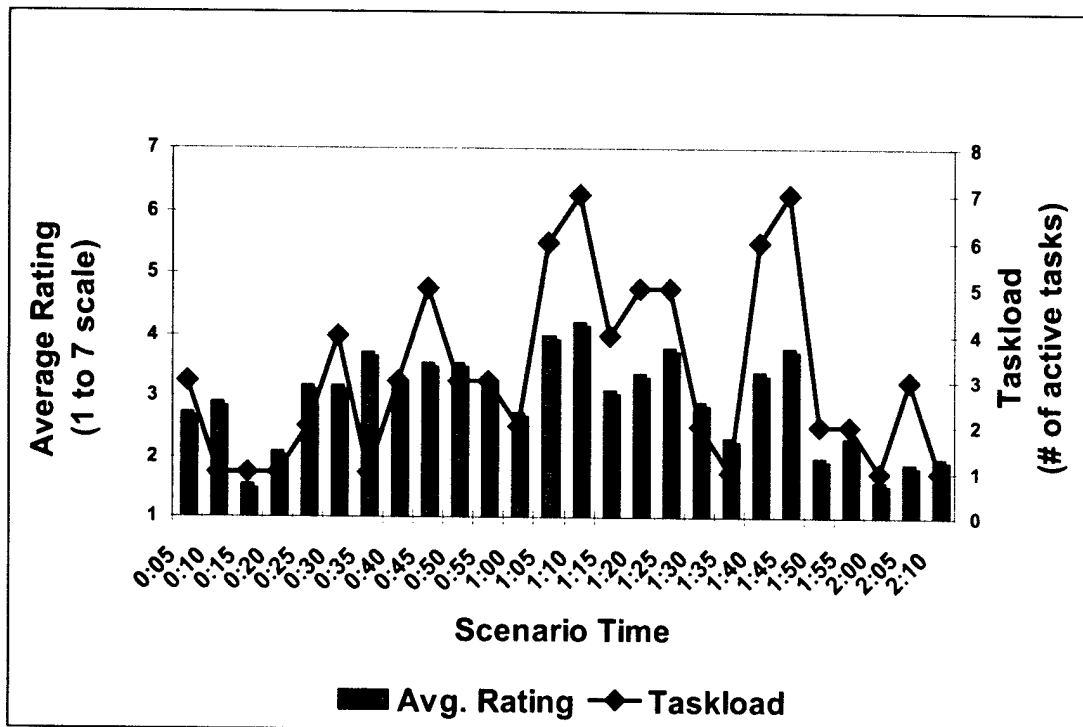
Figure 2. Average workload rating vs. taskload.

After exceeding the taskload measure in over half of the first seven measurements, the reported operator workload rating rarely exceeded it for the rest of the scenario. Observations of the participants indicated that they spent the early, low taskload, portion of the scenario exploring the RPT, which may explain their higher workload reports. Later in the scenario, they were already familiar with the RPT and were able to navigate it quickly, leading to the lower workload reports.

These results are further supported by Figure 3, which shows roughly the same relationship. In this figure, the rating of workload by the participants is compared to a rating of suggested workload, as judged by the SME after numerous testing runs. The correlation between these two factors was also high (r = 0.82).
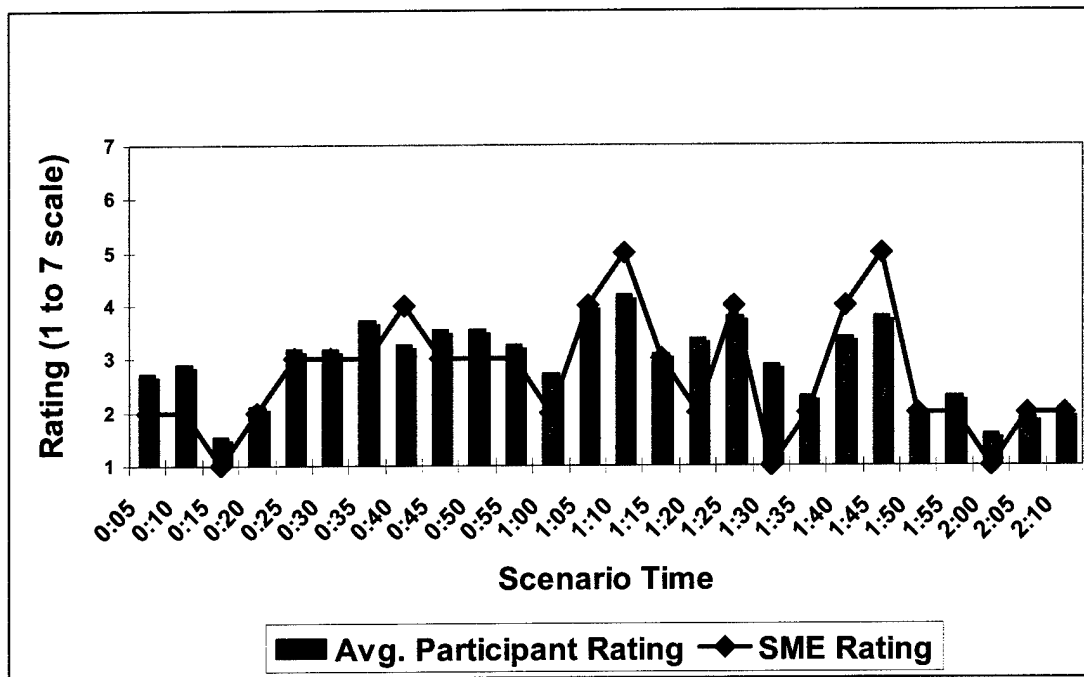
Figure 3. Participant vs. SME workload ratings.

However, these data should be interpreted carefully because the SME was knowledgeable of the scenario events and most likely the capabilities and limitations of the operators as well. As such, the SME was not completely blind to the ratings being provided by the operators. However, when viewed in conjunction with the relationship between taskload and operator workload ratings, these results do support one another and indicate that operators appeared to have good SA of their task and workload demands.

It is also noteworthy that the average operator workload rarely approached even the mid-point on the workload scale, suggesting that the operators did not feel tasking was unmanageable.

3.2.2 NASA TLX Workload Results

At the end of the scenario, operators were given a brief description of the dimensions in the NASA TLX and asked to rate their workload on a 1-to-20 scale for each of these 10 dimensions. Figure 4 shows the average ratings.

As the figure shows, average ratings for all but one of the workload dimensions lies below the midpoint of the scale, which could be interpreted to mean that workload did not appear to be beyond "acceptable levels." It should be noted that the "Performance" dimension is an indication of how the operators felt they performed, not how difficult performance was. In other words, high "performance" ratings on this scale, as are shown here, are considered good. In their self-reporting, the participants believed they performed well, and this was generally backed up by the data and by the opinion of the SME, indicating their self-reporting was accurate.
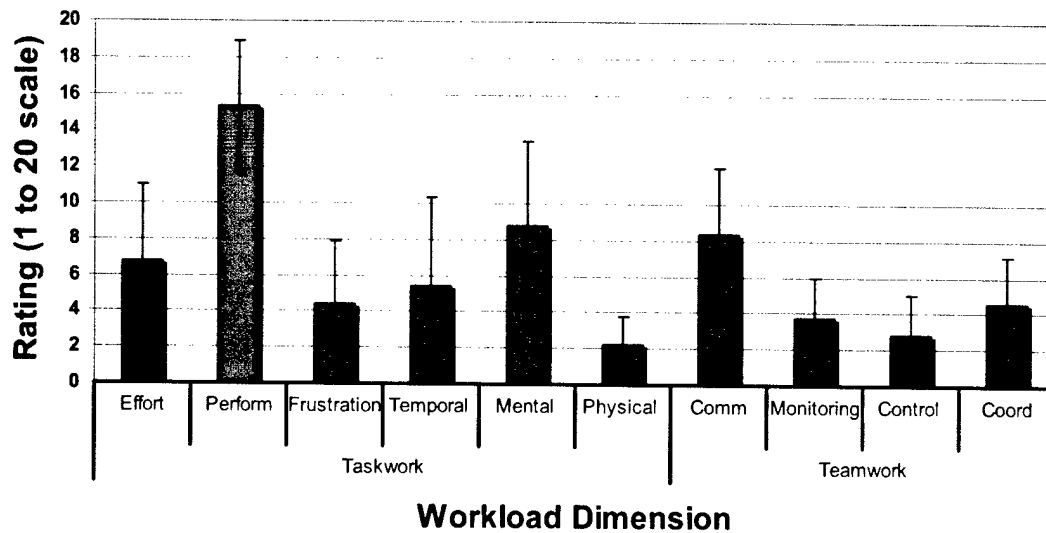
19

Figure 4. Average TLX ratings across scenario.

Because this testing was exploratory in nature, not a comparison between groups, the data presented here must be viewed within that context. The lack of data for comparison allows only for interpretations of relative levels of workload, not the conclusion that the prototype HCI supported better workload than predecessor systems. However, in the experience of the researchers and analysts of these data, the results indicate substantially lower workload than seen in other systems, including a successful application of task-centered design in the air defense warfare domain.

It is evident from these workload results that the operators apparently did not feel overwhelmed by the tasks they were asked to do, and that they felt they performed them well. This indicates that the task-centered design used to develop the prototype HCI supported low levels of operator workload, which is the goal of any task-based system.

3.2.3 Prioritization

Analysis of the operators' prioritization strategies was based on observer notes and comments from the participants in their post-scenario interviews. While most of the participants stated that they evaluated the tasks requiring their attention and addressed the most time-sensitive ones first, this was not the strategy that was always observed. On several occasions, some participants appeared to address their tasks in a top-down or left-to-right order, without consideration of time constraints. (The prototype sorted ESPs by number and engagements by TOL. These were default criteria that operators were unable to change, thus preventing different organization schemes for prioritizing tasking; a fully developed system would presumably provide users with greater sorting flexibility.) Thus tasks that were non-critical or had a long time window for completion, such as sending most of the reports, might be addressed before executing the missile engagements, a task with a strict and short window of opportunity. Future prototypes might want to consider methods for indicating to the operator which tasks have the most urgent deadline, instead of just displaying an engagement's TOL.

3.3 SITUATIONAL AWARENESS RESULTS

The results of the SA probes, provided in Table 5, indicate that the participants appeared to do best on the questions related to Level 3 (prediction) SA (81% correct). Because much of the tasking in TTWCS is schedule-based, the prototype HCI focused heavily on the organization of data into a format that would allow operators to quickly find information about the timing of future events. Consequently, these results may indicate that the design approach did support Level 3 SA. However, there were relatively low scores on the lower level SA questions, dealing with perception and comprehension. This may be attributable to the display design supporting supervisory control of automation, which generally requires a higher level SA than the traditional data-entry and review-function based displays. Alternatively, the low scores in Level 1 and 2 may represent more of a familiarization deficiency with the prototype. As mentioned previously, the training time provided for the RPT was very limited. Additionally, some participants had no prior TTWCS experience. It is possible that Level 1 and 2 scores would rise given additional familiarization and training.

Table 5. Means and standard deviations of the percentage of correct responses to embedded SA probes by level.

SA Questions		Avg. Percentage Correct Responses	Standard Deviation
Level 1	Perception	58.4	27.20
Level 2	Comprehension	64.6	28.55
Level 3	Projection	81.8	23.3

Higher projection (Level 3) scores may also be attributed to the fact that current Tomahawk operator training emphasizes critical thinking and the ability to assess scenario events without the need for step-by-step prompting. It is quite possible that this training, coupled with the increased awareness of task and events afforded by the RPT's cueing of critical events, was the greatest contributor to high Level 3 SA.

As response "correctness" was based on the expert evaluator's judgment, these results should be interpreted with caution. Additionally, a baseline of comparison (i.e., answers to probes of operators using TTWCS or ATWCS systems) would be necessary to determine the comparative value of these results.

The full results of the SA probe questions are provided in Appendix 4.

Looking for consistent delays across operators to make inferences based on their performance in response to the same event indicated several possible points for further effort. The DSMAC problem described previously (Section 2.5.1.1) appeared to be more of a training issue than a problem with the presentation by the HCI. The HCI did not seem to support the arrival of the MDU, as operators seemed unsure about how to handle the notification. One other issue that seemed to cause participants difficulty was the hatch failure and determining the timeline of events (when the first missile tried to launch, when the Auto-Ready Spare launched instead). Participants often wanted to click the "Details" button from the launcher display to learn more, but it was not implemented in the prototype. Overall, few events seemed to cause general consternation. Given the short training and familiarization time the participants had with the RPT, it is possible that more exposure (and a more fully functioning RPT) would lead to even easier handling of events.

3.4 HUMAN–AUTOMATION INTERACTION QUESTIONNAIRE RESULTS

Table 6 provides the mean and standard deviation of participant ratings for the items on the Human–Automation Questionnaire administered at the end of the test session. In general, participants responded very favorably to the prototype HCI, agreeing that the HCI would be an improvement over their current systems. They also agreed that the way the system presented information was easily understood. Moreover, the data indicated that the participants had confidence in the system and felt as though they performed their tasking well. Finally, participants did not believe that the HCI limited their ability to perform their tasks, as indicated by their response to negative characterizations of the system.

Participants also stated that they felt the system was easy to use and navigate. When asked what they liked the least about the system, they cited their unfamiliarity with it. This might indicate that they felt as though the source of some of the problems they had with the system was due to their unfamiliarity and not necessarily an issue with the design.

The observers recorded participant comments and noted several aspects about the design, which may be helpful to future development or prototyping efforts. One participant request was for changing the indication of "executing" and "executed" events because the same green was confusing; perhaps make the launched missiles black like the launcher cell indicator. An observation was that buttons such as "ACK" should indicate either what the operator is acknowledging or how many actions currently require acknowledging. A participant suggested changing the label when an MDU notification comes in to "Ready to Receive," instead of the normal "Send Validation Message," which does not make sense in the context. Several participants complained about the time options, desiring a way to select from Time-on-Target (TOT), Time-of-Launch (TOL), Time-until-Launch (TUL), real time, etc. Due to unfamiliarity with both the RPT and post-launch monitoring for some participants, there were requests for a key or a legend for the symbology of the post-launch monitor, possibly in a help menu. One participant said "I wish I had more information when a new Bravo or Charlie (ESP) comes in," and another suggested a small "ding" sound, similar to the workload prompt, to notify operators of a new arrival. (Obviously, analysis of such an addition must consider all platforms and their requirements; for instance, submarines may require silent running.) On a related note, there were requests for a way to look up the ESP arrival times and keep track of the changes/updates.

Table 6. Average ratings to the human–automation interaction questionnaire items (1 = Strongly Disagree, 5 = Strongly Agree)

Question	Average	Std. Dev.
Higher scores better		
Information on my displays was in a format that was easy to understand	4.57	1.01
This system would be an improvement over the systems that I am currently using.	4.43	0.50
I was always confident that the system was providing me the correct data to perform my task.	4.29	0.90
I feel as though I performed my tasking well.	4.14	0.60
The information provided to me by the system was enough to accomplish the tasks I needed to perform.	4.00	0.94
Information was easy to find when I needed it.	3.57	1.27
Lower scores better		
I felt like I had little control over the way I performed my tasks using this system.	3.57	1.01
At times during the scenario run, I felt as though I was missing pieces of information that I needed to make the appropriate decision.	3.29	1.10
Given the choice, I would have performed my tasks differently than the system prompted me to do.	2.29	0.81
At times, I felt as though I had too much information given to me by the system.	1.86	0.70

3.5 TEAM PROCESS RESULTS

Team process was measured using the previously described ATOM questionnaire. While the investigation focused only on an individual operator performing the tasks, there was interaction between the operators and the SME evaluator playing the role of the ECO. On completion of the scenario, the SME rated the operator's teamwork on a 1- (weakness in the dimension) to 5- (strength in the dimension) scale. The results of the analysis (Figure 5) indicate that operators, on average, were rated highly on Information Exchange, Communication, Supporting Behavior, Initiative, and Critical Thinking as evidenced by above midpoint ratings.

These results suggest that the operators were able to perform tasking at relatively high levels and still manage to effectively contribute to the team. Of note was the high rating of "providing guidance," indicating that the evaluator believed that, in their interactions with the ECO, the operators appeared to be providing useful guidance. However, caution should be used in interpreting these results because the potential does exist for evaluator bias.

Further studies should be conducted to determine how teams of operators using the task-centered prototype HCI would coordinate and communicate within a team setting.
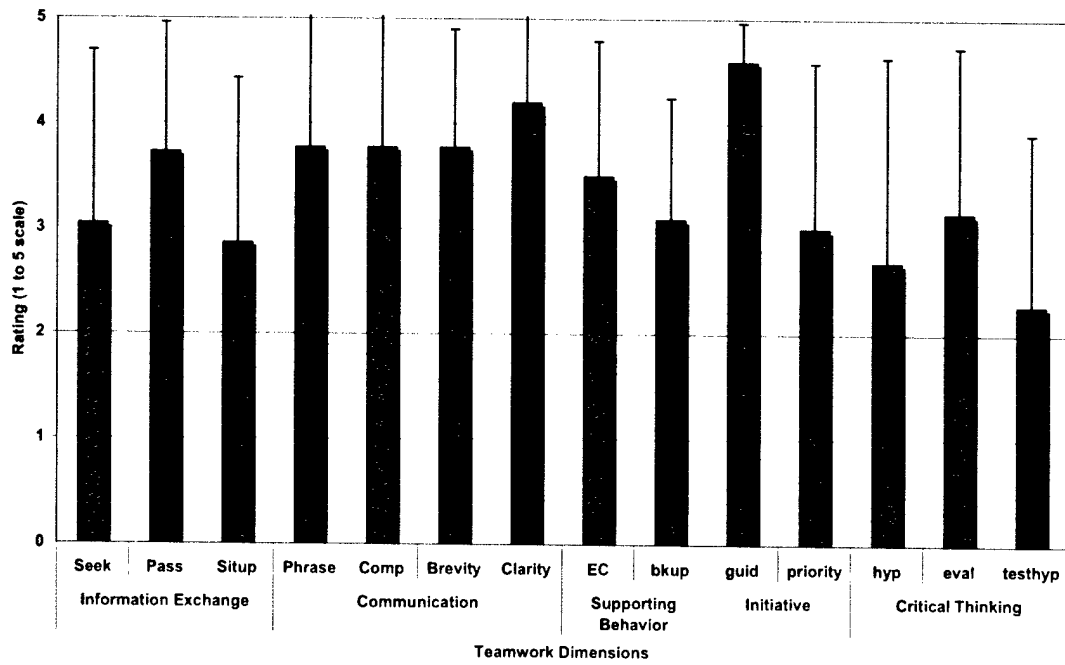
Figure 5. Mean teamwork ratings.

4. DISCUSSION AND CONCLUSION

The FLOT provided an opportunity to examine a task-centered system design in a simulated land attack operational environment. As stated in the Introduction, the goal of this evaluation was to determine the current level of performance that can be expected from TTWCS given the incorporation of human factors principles and task management techniques developed under an ONR FNC research and development effort. The data indicate that the system designed by this process allows a high level of tasking to be successfully handled by a single operator without incurring an unacceptably high level of workload.

The tasks that the operators had to perform were primarily completed within the SME-recommended time window. The tardiness of the other tasks can be attributed to the extremely short training and familiarization time given the participants and changes in CONOPS. The results indicate that the RPT design assisted the operator in completing a high load of tasks with a low level of late or incorrect actions.

The workload reported by the operators and indicated by the SME shows that the system supported the operators and prevented them from feeling overwhelmed even during a highly time- and staffing-compressed scenario. During five-minute time windows, no operator reported a workload of the maximum level, and the highest average barely exceeds the middle rating. Their post-scenario ratings were similarly encouraging. By helping the operators assess task importance and guiding them through the steps for successful completion of the tasks, the task-centered design exhibited the ability to successfully help the operator complete a high load of tasks with an acceptable level of workload.

The SA probes indicate that operator Level 3 SA was generally good, but some attention should be focused on the design to better support lower SA levels (1 and 2). However, it should also be noted that there was a great deal of novelty in the supervisory control design of the prototype HCI. Also, operators received minimal training on the prototype, and it is probable that existing experiences and training may have influenced how they handled information. While it appears that operators were able to adapt their previous training and experience to the new design, it is quite possible that with more system-specific exposure and training, operators would likely be able to improve their SA at all levels.

The automation trust of the operators may also have been impacted by their experiences and training. Their ratings on the automation interaction measurement showed they found the system to be helpful, the information trustworthy, and overall, to be an improvement over their existing systems. However, there were times they would have liked more control over the guidance provided by the system.

Overall, the task-centered design of the system was well liked by the operators and showed itself capable of supporting high tasking levels without associated high-workload reports. It should be tested with teams of operators to further examine teaming issues, and better models of automation should be included. In this evaluation, the task-centered, supervisory-control design of the LACS prototype HCI was shown to successfully support a single operator in performing a highly compressed Tomahawk scenario with multiple simultaneous tasks.

5. REFERENCES

Endsley, M. R. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the 32nd Annual Meeting of the Human Factors and Ergonomics Society*, (pp. 97–101). Santa Monica, CA: Human Factors and Ergonomics Society.

Hart, S. G. and Staveland, L. E. (1988). "Development of the NASA-TLX (Task Load Index): Results of Experimental and Theoretical Research. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload*, (pp. 131–138). Amsterdam: North Holland.

Johnston, J. H., Cannon-Bowers, J. A., and Smith-Jentsch, K. A. (1995). Event-based performance measurement system for shipboard command teams. *Proceedings of the 1st International Symposium on Command and Control Research and Technology*, Washington, D.C. The Center for Advanced Command and Technology, pp. 274–276.

Osga, G., Van Orden, K., Campbell, N., Kellmeyer, D., and Lulue, D. (2002). Design and Evaluation of Warfighter Task Support Methods in a Multi-Modal HCI. (TR 1874). San Diego, CA: Space and Naval Warfare Systems Center, San Diego.

Pharmer, J. A., Campbell. G. E., and Hildebrand, G. A. (2001). Report on the ONR/SC-21 Science and Technology Manning Affordability Initiative Aegis Comparison Study MMWS Build-1 Results. Orlando, FL: Naval Air Warfare Training Center.

Smith-Jentsch, K. A., Johnston, J. H., and Payne, S. C. (1998). Measuring Team-Related Expertise in Complex Environments. In J. A. Cannon-Bowers and E. Salas (Eds.) *Making Decisions Under Stress: Implications for Individual and Team Training*, (pp. 61–87). Washington, DC: American Psychological Association.

6. ACRONYMS

ASAP	As Soon As Possible
ATOM	Air Warfare Team Observation Measure
ATWCS	Advanced Tomahawk Weapon Control System
B/U	Backup
BDI	Battle Damage Indication
BDII	Battle Damage Indication Imagery
CA	Cell Allocation
CFF	Call for Fire
CIC	Combat Information Center
CM	Capable Manpower
CONOPS	Concept of Operations
CPHS	Committee for Protection of Human Subjects
CSCSPLD	Center for Surface Combat Systems Point Loma Detachment
DDG	*Arleigh Burke* Class Destroyer (Guided Missile)
DSMAC	Digital Scene Matching Area Correlator
ECO	Engagement Control Officer
ESP	Electronic Strike Package
FCTCPAC	Fleet Combat Training Center Pacific
FLOT	Fleet Operability Test
FNC	Future Naval Capability
FTC	Failure to Transition to Cruise
GPS	Global Positioning System
H&S	Health and Status
HCI	Human–Computer Interface
HFE	Human Factors Engineering
HSRB	Human Subjects Review Board
IMMM	In-flight Mission Modification Message
JHU/APL	Johns Hopkins University Applied Physics Laboratory
KSA	Knowledge Superiority & Assurance
LAC	Land Attack Coordinator
LACS	Land Attack Combat System
MDU	Mission Data Update
NASA TLX	National Aeronautics and Space Administration Task Load Index
NAVAIR	Naval Air Systems Command

NLT	No Later Than
OIF	Operation Iraqi Freedom
ONR	Office of Naval Research
PLR	Post-Launch Report
PSR	Post-Strike Report
RPT	Rapid Prototype
SA	Situational Awareness
SC-21	Surface Combatant 21
SCO	Strike Coordination Overlay
SME	Subject-Matter Expert
SSC San Diego	Space and Naval Warfare Systems Center San Diego
TADMUS	Tactical Decision Making Under Stress
TBD	To Be Determined
TLAM	Tomahawk Land-Attack Missile
TM	Task Manager
TOL	Time of Launch
TOT	Time on Target
TSC	Tomahawk Strike Coordinator
TTWCS	Tactical Tomahawk Weapon Control System
TWCS	Tomahawk Weapon Control System

APPENDIX 1: MODIFIED TASK LOAD INDEX

MODIFIED TASK LOAD INDEX

Period: _____ Date: _____ Watchstation: _____
Workload Rating Scales

EFFORT — How hard did you have to work (mentally and physically) to accomplish your level of performance? (Check a box on the scale)

Low   High

PERFORMANCE — How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Poor   Good

FRUSTRATION LEVEL — How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Low   High

TEMPORAL DEMAND — How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Low   High

MENTAL DEMAND — How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Low   High

PHYSICAL DEMAND — How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task physically easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low   High

COMMUNICATIONS DEMAND — How much communication was required between you and other team members directly, over nets, or by way of the workstation? Did requesting and transferring information consume a little of your time or a lot? Was it easy or demanding?

Low   High

MONITORING DEMAND — How much monitoring of people did the task require? Was attending to others (directly or through your workstation) easy or demanding, infrequent or continuous?

Low   High

CONTROL DEMAND — How much correction of others did the task require? Was correcting other people or the workstation easy or demanding, infrequent or continuous?

Low   High

COORDINATION DEMAND — How much correction or adjustment of your own actions was required in order to coordinate with others? Was adjusting your actions to improve coordination simple or complex, periodic or constant?

Low   High

APPENDIX 2: HUMAN–AUTOMATION INTERACTION QUESTIONNAIRE

HUMAN–AUTOMATION INTERACTION QUESTIONNAIRE

Instructions. Please circle the rating that best describes how much you agree or disagree with the following statements.

The information provided to me by the system was enough to accomplish the tasks I needed to perform.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Information on my displays was in a format that was easy to understand.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Given the choice, I would have performed my tasks differently than the system prompted me to do.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

I was always confident that the system was providing me the correct data to perform my task.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Information was easy to find when I needed it.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

I feel as though I performed my tasking well.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

At times during the scenario run, I felt as though I was missing pieces of information that I needed to make the appropriate decision.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

At times, I felt as though I had too much information given to me by the system.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

This system would be an improvement over the systems that I am currently using.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

I felt like I had little control over the way I performed my tasks using this system.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Instructions: Please provide any comments you may have about this system.

The thing I liked most about this system was:

_____
_____
_____
_____

The thing I liked least about this system was:

_____
_____
_____
_____

The thing I would change about this system is:

_____
_____
_____
_____

Appendix 2: Human–Automation Interaction Questionnaire

If I were designing this system I would add:

_____
_____
_____
_____
_____

Additional Comments:

_____
_____
_____
_____
_____

APPENDIX 3: ATOM SCALE

TDT RATINGS FORM

Team #: _____

Scenario #:_____

Rater: _____

Information Exchange

<u>Seeking sources</u> - Proactively asking for information from multiple sources in order to establish an accurate assessment of the situation. These sources may be internal or external to the team and may include written documentation.

```
1        2              3              4              5
|--------|--------------|--------------|--------------|
```

Seeking information
is a real weakness for
this team.

Seeking information
is a real strength for
this team.
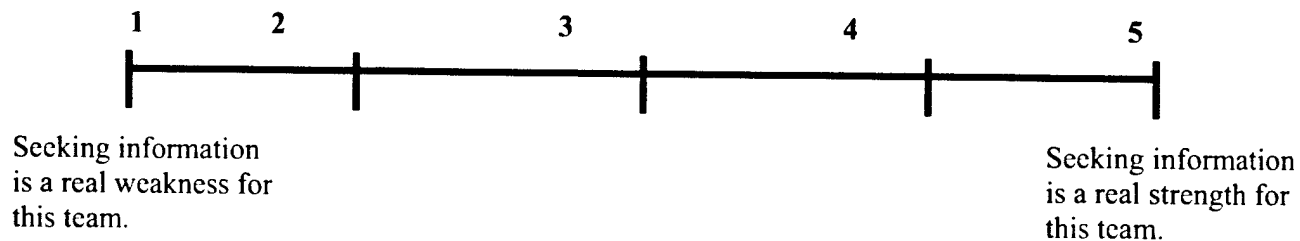
Please list one concrete example (this may be a positive or negative example):

_____

<u>Passing information</u> - Anticipating another team member's need for information and passing it to him/her without having to be asked. This could be a single piece of information passed to an individual or group of individuals.

```
    1       2              3              4              5
    |-------|--------------|--------------|--------------|
```

Passing information
is a real weakness for
this team.

Passing information
is a real strength for
this team.

Please list one concrete example (this may be a positive or negative example):

_____

==

<u>Situation update</u> - An update given by a team member either to the entire team or a subset of the team, which provides an <u>overall summary</u> of the <u>big picture</u> as they see it. This can include updates reported internally within the team and updates that go out from the team to others.

```
    1       2              3              4              5
    |-------|--------------|--------------|--------------|
```

Providing situation
updates is a real
weakness for this
team.

Providing situation
updates is a real
strength for this
team.

Please list one concrete example (this may be a positive or negative example):

_____

Communication

Proper phraseology - Use of standard terms or vocabulary when sending a report.

```
    1         2              3              4         5
    |---------|--------------|--------------|---------|
```

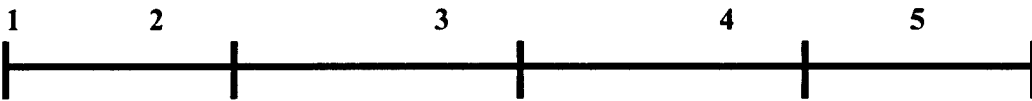Phraseology is a real
weakness for this
team.

Phraseology is a real
strength for this
team.

Please list one concrete example (this may be a positive or negative example):

_____

Completeness of reports - Following standard procedures that indicate which pieces of information
are to be included in a particular type of report and in what order.

```
    1         2              3              4         5
    |---------|--------------|--------------|---------|
```

Incomplete reports
are a real weakness
for this team.

Providing complete
reports is a real
strength for this
team.

Please list one concrete example:

_____

Brevity - The degree to which team members avoid excess chatter, stammering and long winded reports, which tie up communication lines.

```
  1        2              3                  4     5
  |--------|--------------|------------------|-----|
```
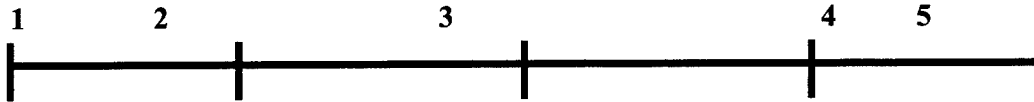
Brevity is a real
weakness for this
team.

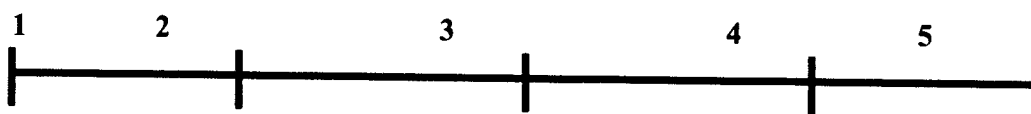Brevity is a real
strength for this
team.

Please list one concrete example (this may be a positive or negative example):

_____

Clarity - The degree to which a message sent by a team member is audible (e.g.., loud enough, not garbled, not too fast).

```
  1        2              3             4          5
  |--------|--------------|-------------|----------|
```

Communication/clarit
y is a real weakness
for this team.

Communication/clarit
y is a real strength for
this team.

Please list one concrete example (this may be a positive or negative example):

_____

Supporting Behavior

<u>Error correction</u> - Instances where a team member <u>points out</u> that an error has been made and either corrects it him/herself or see that it is <u>corrected</u> by another team member.

```
    1       2              3              4              5
    |-------|--------------|--------------|--------------|
```
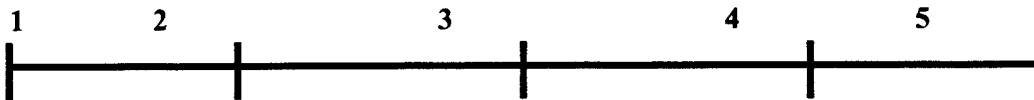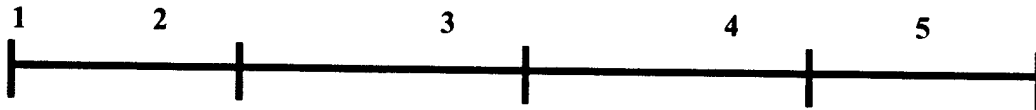
Error correction is a
real weakness for
this team.

Error correction is a
real strength for
this team.

Please list one concrete example (this may be a positive or negative example):

_____

<u>Providing and requesting backup/assistance</u> - Instances where a team member either requests assistance or notices that another team member is overloaded or having difficulty performing a task and provides assistance to them by actually <u>taking on</u> some of their <u>workload</u>.

```
    1       2              3              4              5
    |-------|--------------|--------------|--------------|
```

Offering and requesting
Backup is a real
weakness for this team.

Offering and requesting
backup is a real
strength for this team.

Please list one concrete example (this may be a positive or negative example):

_____

Initiative/Leadership

Providing guidance - Instances where a team member directs or suggests that another team member take some action or instructs them on how to perform a task.

| 1 | 2 | 3 | 4 | 5 |

Providing guidance or
suggestions is a real
weakness for this
team.

Providing guidance or
suggestions is a real
strength for this team.

Please list one concrete example (this may be a positive or negative example):

Stating priorities - Instances where a team member specifies, either to the team as a whole or to an individual team member, the priority ordering of multiple tasks.

| 1 | 2 | 3 | 4 | 5 |

Stating priorities is a
real weakness for
this team.

Stating priorities is a
real strength for this
team.

Please list one concrete example (this may be a positive or negative example):

APPENDIX 4: SITUATIONAL AWARENESS
PROBE QUESTIONS AND RESULTS

SITUATIONAL AWARENESS PROBES ASKED BY SME ECO ROLE-PLAYER

Approximate Scenario Time	Probe Question	SA Level	Correct Answers (out of 10)	Percent Correct
0:04:45	What are the estimated times of first and last launches, and their associated Engagement numbers for ESP0001B?	1	10	100.0
0:08:20	What is the estimated time 1st missile mode 7 and its associated plan?	1	7	70.0
0:10:45	Report when 1st missile mode 7/all missiles mode 7	1	4	40.0
0:19:00	Report current primary tasking and max follow-on salvo capability for ESP0001B	1	5	50.0
0:40:00	Report when Call for Fire (CFF) execute received	1	4	40.0
0:43:00	What is the estimated earliest time of launch for CFF 1?	3	8	80.0
0:43:40	Report 1st/last launch time for ESP0002A and their associated engagements.	1	9	90.0
0:46:00	Is DSMAC failure mission critical?	2	5	50.0
0:53:45	What is launch time earliest for CFF?	3	10	100.0
0:56:00	Report time of hatch failure.	1	4	40.0
0:59:00	How many engagements left in ESP0001C?	2	9	90.0
1:03:00	Report tasker for CFF engagement	2	10	100.0
1:04:00	How many missions are in the MDU and can we meet tasking?	3	8	80.0
1:13:00	Report H & S status of CFF1 ESP 0004B	2	10	100.0
1:16:00	What is TOT for CFF 1?	3	10	100.0
1:23:00	What is TOL for CFF2?	3	10	100.0
1:25:00	What is engagement # and est. TOL for R/S?	3	9	90.0
1:35:00	Which ESP is Health & Status (H&S) IMMM from E031 associated with?	2	7	70.0
1:39:00	Report max follow-on salvo capability of ESP0002B	3	10	100.0
1:47:00	Why was the flex necessary?	2	6	60.0
1:51:00	How many BLK IVs are in flight?	2	6	60.0
2:02:00	What is TOT for all active ESPs ?	2	8	80.0
2:04:00	Report BDI for MSN 0046 ESP 0002C.	2	1	11.1
2:09:00	What is loiter target and estimated time of loiter exit?	3	3	30.0

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-01-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden to Department of Defense, Washington Headquarters Services Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)* 07–2004	2. REPORT TYPE Final	3. DATES COVERED *(From - To)*	
4. TITLE AND SUBTITLE TACTICAL TOMAHAWK WEAPON CONTROL SYSTEM V6 LAND ATTACK COMBAT SYSTEM PROTOTYPE HUMAN–COMPUTER INTERFACE Test Report for FY 03 Fleet Operability Test		**5a. CONTRACT NUMBER**	
		5b. GRANT NUMBER	
		5c. PROGRAM ELEMENT NUMBER	
6. AUTHORS James Pharmer (NAVAIR Orlando Training Systems Division) Kevin Cropper, Jennifer McKneely (JHU APL) Earl Williams, Ph.D. (SSC San Diego)		**5d. PROJECT NUMBER**	
		5e. TASK NUMBER	
		5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NAVAIR Orlando Training Systems Division SSC San Diego 12350 Research Parkway, Orlando, FL 32826 San Diego, CA 92152–5001 Johns Hopkins University Applied Physics Laboratory (JHU APL) 11100 Johns Hopkins Road, Laurel, MD 20723–6009		**8. PERFORMING ORGANIZATION REPORT NUMBER** TD 3184	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research 800 North Quincy Street Arlington, VA 22217–5660		**10. SPONSOR/MONITOR'S ACRONYM(S)** ONR	
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)	

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
This is the work of the United States Government and therefore is not copyrighted. This work may be copied and disseminated without restriction. Many SSC San Diego public release documents are available in electronic format at http://www.spawar.navy.mil/sti/publications/pubs/index.html

14. ABSTRACT
This report describes an operability test of a prototype HCI developed through a task-centered design process. Using this prototype, participants individually performed a simulated Tactical Tomahawk launch and control scenario. During the testing, data were collected on the participants' task completion timeliness, situational awareness, and subjective workload. Analysis indicates that the participants successfully completed nearly all tasking on time, while maintaining a moderate to high level of situational awareness and reporting perceived workload to be low to moderate. Operators gave generally positive feedback about the prototype system's functionality, ease of use, and trainability.

This report describes an operability test of a prototype HCI developed through a task-centered design process. Using this prototype, participants individually performed a simulated Tactical Tomahawk launch and control scenario. During the testing, data were collected on the participants' task completion timeliness, situational awareness, and subjective workload. Analysis indicates that the participants successfully completed nearly all tasking on time, while maintaining a moderate to high level of situational awareness and reporting perceived workload to be low to moderate. Operators gave generally positive feedback about the prototype system's functionality, ease of use, and trainability.

15. SUBJECT TERMS
Mission Area: Command and Control Tactical Tomahawk Weapon Control System V6, Land Attack Combat System Human–Computer Interface

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Earl Williams, Ph.D.
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			
U	U	U	UU	62	**19B. TELEPHONE NUMBER** *(Include area code)* (619) 553–9561

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

INITIAL DISTRIBUTION

SSC San Diego
San Diego, CA 92152–5001