ARMY RESEARCH LABORATORY

Janus Digitization Test Bed: Assessing Performance in a Simulated Battlefield Environment

John K. Hawley
James P. Flanagan
Tary D. Wilkinson
Linda G. Pierce

ARL-TR-1437

DECEMBER 1997

ERRATA SHEET

re: ARL-TR-1437, "Janus Digitization Test Bed: Assessing Performance in a Simulated Battlefield Environment,"
by John K. Hawley, James P. Flanagan, Tary D. Wilkinson
of Hughes Training, Inc., and
Linda G. Pierce of
Human Research & Engineering Directorate, ARL

The following pen-and-ink change should be made on page 45 in the sixth line of the first paragraph under the heading "Standard Human Performance Measures and Moderators":

> performance measures contained in Lowry (1955)

should be changed to read

> performance measures contained in Lowry (1995)

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5425

ARL-TR-1437 December 1997

Janus Digitization Test Bed: Assessing Performance in a Simulated Battlefield Environment

John K. Hawley
James P. Flanagan
Tary D. Wilkinson
Hughes Training, Inc.

Linda G. Pierce
Human Research & Engineering Directorate, ARL

Abstract

An evaluation of the combat information processor (CIP) was conducted by the Fort Sill Field Element, Human Research and Engineering Directorate (HRED) of the U.S. Army Research Laboratory (ARL), and Hughes Training, Inc., (HTI) in support of the Depth and Simultaneous Attack Battle Lab at Fort Sill, Oklahoma. The Training and Doctrine Command, through the Concept Evaluation Program, funded the research. Our objectives were to (a) evaluate the CIP in a simulation test bed to provide user feedback to the system developers and (b) develop guidance for using simulation test beds such as the Janus Battle Simulation Center (JBSC) in operational test and evaluation, particularly early user tests. More specifically, the CIP user test was conducted to (a) evaluate the battle command utility of the CIP software, (b) provide formative feedback to the CIP system developer, and (c) recommend technological and procedural enhancements to improve information management during the conduct of battle exercises within the JBSC. It was hypothesized that by establishing a test bed in conjunction with a multi-user simulation center, developmental or prototype hardware and software products could be evaluated in relationship to fielded systems, allowing system developers to receive early user feedback regarding the suitability of the product for user application. This process would result in the enhancement of the acquisition cycle by providing a high fidelity, low cost environment to support early and frequent user testing. The first phase of the test, a functional review of the CIP software, revealed a number of deficiencies that would limit the usability of the CIP by a battle command staff, whether in a simulation test bed or in the field. The second phase of the test involved a limited user evaluation during two Janus battle simulations. A number of deficiencies were identified in the use of the CIP in an operational environment, especially in the use of software control measures. Deficiencies and our observations are included in the report, along with recommended solutions to aid in the design of the next generation software. The current research program was initiated to address the use of simulation test beds to support the acquisition of battle command systems. Although the current simulation test bed was adequate for conducting a limited user evaluation, it was suggested that future simulations-based testing be developed using distributed interactive simulation (DIS) technology. The use of a DIS environment will allow for immersion of the test systems and operator into the synthetic environment to increase the realism of the training and ensure the validity of the user assessment.

CONTENTS

APPENDIX

TABLES

EXECUTIVE SUMMARY

An evaluation of the combat information processor (CIP) is described in the current report. The research was conducted by the Fort Sill Field Element, Human Research and Engineering Directorate (HRED) of the U.S. Army Research Laboratory (ARL), and Hughes Training, Inc., (HTI) in support of the Depth and Simultaneous Attack Battle Lab at Fort Sill, Oklahoma. The research was funded by the Training and Doctrine Command (TRADOC) through the Concept Evaluation Program (CEP).

The objectives of the research program were to (a) evaluate the CIP in a simulation test bed to provide user feedback to the system developers and (b) develop guidance for using simulation test beds such as the Janus Battle Simulation Center (JBSC) in operational test and evaluation (OT&E), particularly early user tests. More specifically, the CIP user test was conducted to (a) evaluate the battle command utility of the CIP software, (b) provide formative feedback to the CIP system developer, and (c) recommend technological and procedural enhancements to improve information management during battle exercises within the JBSC.

It was hypothesized that by establishing a test bed in conjunction with a multi-user simulation center, developmental or prototype hardware and software products could be evaluated in relationship to fielded systems, allowing system developers to receive early user feedback regarding the suitability of the product for user application. This process would result in the enhancement of the acquisition cycle by providing a high fidelity, low cost environment to support early and frequent user testing.

The CIP, in development by ARL, is a mobile test bed system designed to demonstrate real-time situation development, multi-sensor fusion, and horizontal integration of the battlefield. The CIP's functionality during the test included a set of map and graphics utilities and accompanying support tools. Other aids designed to reside within the CIP included distance estimation, field-of-view (FOV) estimation, line-of-sight (LOS) estimation, corridor-maneuvering calculation, mobility prediction, peak search, perspective view generation, and route planning. The test consisted of a functional review of the CIP software and an evaluation of the CIP in an operational setting.

The first phase of the test, a functional review of the CIP software, revealed a number of deficiencies that would limit the usability of the CIP by a battle command staff, whether in a simulation test bed or in the field. While these deficiencies are documented in the report, many of

them were identified previously and will have been resolved before the next software release. Our observations are included to aid in the design of the next generation software.

The second phase of the test involved a limited user evaluation during two Janus battle simulations. The CIP was used by a researcher to follow the Janus exercises, while the students participating in the Janus exercises executed the battle manually using paper maps and overlays. A number of deficiencies were identified in the use of the CIP in an operational environment, especially in the use of software control measures. These deficiencies are detailed in the report, along with recommended solutions.

The current research program was initiated to address the use of simulation test beds to support the acquisition of battle command systems. A user test of a battle command system was successfully conducted in the JBSC, and although the current simulation test bed was adequate for conducting the limited test, it was suggested that future simulations-based testing be developed using distributed interactive simulation (DIS) technology. The use of a DIS environment will allow for immersion of the test systems and operator into the synthetic environment to increase the realism of the training and ensure the validity of the user assessment.

JANUS DIGITIZATION TEST BED: ASSESSING PERFORMANCE IN A SIMULATED BATTLEFIELD ENVIRONMENT

INTRODUCTION

The Problem

Information technology is fostering a revolution in military operations. The future warfighting environment will be geographically and temporally dispersed and populated by many small, mobile, and semi-autonomous units possessing weapons of considerable range, accuracy, and lethality. In this future environment, mission planning, decision making, and communications capabilities will be as important as weapons platforms. A term coined to characterize this emerging warfighting environment is "Information Age Warfare."

More than ever, the information age battlefield will include complex socio-technical systems that depend upon the human component for success. Advances in information technology have forced a rapid evolution in combat operations hardware and software. These technical advances have, in turn, demanded entirely new doctrinal and tactical concepts. Interacting and compounding technical, doctrinal, and tactical changes have had a significant impact on the personnel who must use, maintain, and support this emerging class of systems. It has been said that information technology is creating a situation in which we are moving from soldiers organized around systems to one of soldiers organized around information. The human performance implications of this transformation are profound.

Because of these trends, the military services are faced with a dilemma. Information age systems have high performance potential, but debugging and proving them is often difficult or impossible using conventional operational test and evaluation (OT&E) methods and assets. Realistic testing is often hampered or restricted for reasons of cost, range limitations, emission restrictions, safety, and system capabilities and complexity. It has been observed, for example, that using a modern, high-performance aircraft and its weapons suite within the confines of most training ranges is like trying to "fight a war in a phone booth." The same command also applies to the emerging generation of fire support, air defense, armor, and command and control systems.

Another complicating factor is that the systems that support critical information age combat functions frequently are not systems in the traditional military use of the term. Often, "the system" is represented by changes in doctrine or tactics partly based in software and partly based in user training and supported by an item of commercial or Government off-the-shelf equipment. Ambiguity concerning what comprises the system makes traditional field testing difficult.

Project Objectives

The objectives of the current effort are twofold:

1. Conduct a user evaluation of the combat information processor (CIP) using a simulation test bed to

- Evaluate the battle command utility of the CIP software.
- Provide formative feedback to the system developer.
- Recommend technological and procedural enhancements to improve information management during Janus Battle Simulation Center (JBSC) exercises.

2. Develop guidance for using simulation test beds such as the JBSC in OT&E, particularly early user tests.

The first objective is addressed in the body of the current report, while the second objective is presented in an appendix to the report. Many of the issues addressed in the appendix pertain to OT&E in general, but the primary thrust of the discussion is centered on obtaining quality soldier-in-the-loop performance data during manpower and personnel integration (MANPRINT) investigations.

USER TEST OF THE COMBAT INFORMATION PROCESSOR (CIP)

Background

The Combat Information Processor

The combat information processor, or CIP, is being developed by the Information Processing Branch of the U.S. Army Research Laboratory (ARL) as part of the very intelligent surveillance and target acquisition (VISTA) program. The system was designed as a mobile test bed to demonstrate (a) real-time situation development, (b) multi-sensor fusion, and (c) horizontal integration of the battlefield. It is hosted on a UNIX™ platform running X windows and the Open Software Foundation (OSF) Motif user interface.

The CIP's functionality (at the time this test was conducted) is best characterized in terms of a primary set of map and graphics utilities with an accompanying set of support tools. The CIP map and graphics utilities are intended to assist users in performing a variety of generic command and control functions. These functions include the following:

- Process message traffic.
- Display the tactical situation.
- Create overlays and display control measures.
- Access and manipulate terrain databases.

Inter-system communications are enabled using Unix's interprocess communications (IPC) standard. Possible connection media include (a) radio frequency (RF) links, (b) ethernet, or (c) RS232. Entities used in the tactical situation display and control measure symbols are taken from Field Manual (FM) 101-5-1 (Headquarters, Department of the Army, 1985). The CIP's terrain database is generated from standard Defense Mapping Agency (DMA) products. With the CIP's terrain database, users can develop tailored map images by selecting supported map features. Examples of supported features include roads, rivers, bridges, and railroads.

Beyond the primary utilities just listed, the CIP gives users a set of tools intended as aids in processing real-time battlefield information. CIP support tools include the following:

- Distance Estimation. The distance tool is used to estimate the distance between two points.

- Field of View (FOV) Estimation. The FOV tool is used to estimate the coverage of a line of sight (LOS) sensor for a given range.

- Line of Sight Estimation. The LOS tool queries the terrain database and returns a terrain elevation profile between an observer's location and a target location.

- Corridor-Maneuvering Calculation. The corridor-maneuvering tool is used to determine the trafficability of terrain for various classes of vehicles.

- Mobility Prediction. The mobility prediction tool is used to estimate where a selected entity could have traveled within a given time period.

- Peak Search. The peak search tool queries the terrain database for the highest elevation within a specified geographic area.

- Perspective View Generation. The perspective view generation tool gives the user a wire frame mesh portrayal of the elevation of selected terrain.

- Route Planning. The route planning tool is used to determine the optimal route between two points on the map.

The Janus Battle Simulation Center

The Janus Facility

Data collection for the current effort was conducted in the JBSC at Fort Sill, Oklahoma. The Janus facility consists of a suite of player workstations. Each workstation represents an operation or fighting unit that controls a designated portion of the friendly or enemy force. The workstations are not close to one another. Communications between workstations are conducted via electronic means, and the nature of the task requires coordination among the various team activities to complete the mission successfully. Combat actions are driven by user input, movement, and other instructions. Target acquisition, delivery of direct fire, and the results of individual fire events are automatically determined by the simulation according to user-established priorities and probabilities.

The host workstation with a graphics display terminal, external tape and disk drive units, and CD ROM drive is connected to the student workstations by an ethernet cable. A single workstation consists of a Hewlett-Packard processor, graphic display panel, and Summagraphics Sketch III digitizing tablet. The system can be configured with either eight or sixteen student workstations.

Battles conducted within the simulation facility typically are configured with eight workstations comprised of several task forces that represent operational or fighting units and six cells that serve as command and control centers. Six workstations support the Blue forces, which are played by students. Three of these workstations represent the three maneuver battalions. The other three workstations represent (a) close air support and reinforcing artillery, (b) general support and general support reinforcing artillery, and (c) a direct support artillery battalion.

Besides the eight workstations, six player (student) cells represent the tactical operating centers (TOCs) for the Blue forces. Two workstations support the Red force, which is operated by Janus instructors. One of the Red force workstations controls the Red maneuver forces and the other controls Red artillery and close air support.

A typical Janus exercise is organized into three distinct phases, each providing the opportunity for collecting one or more types of data. These phases include (a) planning and input, (b) execution of the plan, and (c) the after-action review (AAR). The entire operation can be recorded, and the measures can be used in an AAR. These measures can also be used for battle outcome calculations. In addition to the tabular reports available from the Janus system, the Janus analyst workstation (JAAWS) provides a graphic representation of these variables. Using JAAWS, a user can replay any portion of an exercise and view those aspects of the operation that are of interest.

The Janus Model

Janus is a two-sided, interactive, closed, stochastic ground combat simulation. The model can portray virtually any tactical situation and the effects of most weapon systems. Players must consider all aspects of employing their forces just as they would in combat. Janus accurately models both friendly and enemy weapons systems with resolution down to the individual platform. These systems have distinct properties such as dimension, weight, carrying capacity, weapons, and weapons capabilities--all of which can be affected by terrain and weather. Recent enhancements include the ability to conduct military operations in urban terrain and improved dismounted infantry functionality.

Janus uses digitized high resolution terrain, displaying it in a format similar to a standard military map representation; contour lines, roads, rivers, vegetation, and urban areas are all represented. System capabilities allow a play box as large as 60 km x 60 km during the simulation. At the battalion and brigade level, Janus serves as an excellent training simulation, requiring detailed interaction between staff members as they develop and execute the ground tactical plan. Users must apply sound warfighting principles and achieve full synchronization of all battlefield operating systems to fight a successful Janus battle.

The Janus simulation system is improved and maintained by the TRADOC Analysis Command (TRAC). It is written in Formula Translator (FORTRAN) and runs on Hewlett-Packard™ or Digital Equipment Corporation (DEC) MicroVAX computers with Hewlett-Packard™ or Tektronix color graphics workstations. Janus uses an accessible database to establish the characteristics of all weapons and systems played in the simulation. New systems can be prepared quickly, and terrain data can be tailored to achieve desired simulations in combat. Janus can be operated by a few specially trained personnel.

METHOD

As noted in the previous section, the purpose of the user test was to evaluate the CIP software in the JBSC. During this assessment, we were to evaluate the utility of the CIP's current[1] capabilities in fire support planning and direction by comparing the performance of battle commanders and their staffs in a manual setting (the current situation in the Janus facility) with performance in an automated environment (CIP-aided operations). The assessment was performed in two phases: (a) CIP software functional review and (b) user evaluation in the JBSC. Both of these phases are discussed in the following subsections.

CIP Software Functional Review

The CIP software functional review was conducted in two steps: (a) a preliminary review of CIP capabilities based on available documentation, and (b) a hands-on assessment of CIP features by a senior member of the project staff. Documents examined during the preliminary review of CIP capabilities included a draft information paper about the CIP (ARL, 1992a) and a CIP software user's manual prepared for the U.S. Marine Corps (ARL, 1992b). These materials were used to orient the reviewer to the CIP system and prepare for the hands-on sessions to follow.

The hands-on assessment was supported by several members of the ARL software development staff during two separate review periods of 3 days each. During these sessions, the software reviewer was trained to use the system and was guided through each of the CIP's features and supporting tools. The software reviewer was experienced in using mission planning tools such as the CIP. Therefore, the two review periods permitted him to become proficient in CIP use and assess the CIP's strengths and weaknesses vis-à-vis competing systems. The reviewer became the project staff's CIP subject matter expert (SME) and trainer for the subsequent user test.

User Evaluation

The original plan for the user evaluation was to compare the performance of two groups of Janus users, one using manual planning and direction methods and another employing the CIP. Delays in getting the CIP software operational in the JBSC made it impossible to follow the original plan. Accordingly, a fall-back option for the user test was employed. The fall-back

---

[1] Presently, there are two versions of the CIP software, denoted as "old" and "new." We evaluated the old version because the new version was not ready for use.

option called for a senior member of the project staff having experience in fire support mission planning and direction to be trained in CIP use. This staff member would then serve as a surrogate for personnel who might eventually use the CIP in the JBSC.

The user test was scheduled to take place during two Janus exercises: one involving a U.S. Marine Corps Reserve battle staff and a second conducted by a Field Artillery Officers' Basic Course (FAOBC) class. During these sessions, the project staff member serving as a surrogate user was to set up the exercise and attempt to use the CIP during battle planning and direction. The primary outcome measures for these trials were (a) the surrogate user's success in keeping pace with the parallel exercise (yes or no), and (b) his opinions about the CIP's usability and utility in fire support mission planning and direction.

Because of a series of unfortunate circumstances, the U.S. Marine Corps unit did not participate in their scheduled Janus exercise. The FAOBC exercise was executed as planned. The user test results reported in the next section are from the FAOBC exercise only. Also, because of restricted opportunities for data collection resulting from delays in getting the CIP operational at Fort Sill, we were not able to conduct a test that would fully demonstrate the testing concepts and principles addressed in Appendix A. We were left in the position of only being able to conduct a cursory CIP user evaluation within the context of two Janus exercises. However, the concepts and principles presented in the appendix, coupled with the lessons learned from a "real-world" simulations-based test of an automated system, should provide valuable insight for the testing community about the potential advantages and disadvantages of using a simulation test bed to evaluate future systems.

RESULTS

CIP Software Functional Review

During the first phase of the user test, CIP software was evaluated on a functional basis, that is, what it would and would not do. Documentation, software code, and models were not analyzed because of problems with system availability and lack of documentation. System developers were asked about obvious problems and glitches, but a systematic analysis was not conducted. One major reason for many problems encountered was the state of the version used. For lack of a more precise term, we will call this version the "old" version. A newer version is being developed. This analysis will support development of the next version of CIP software.

The old version has not been maintained for an extended period of time and, according to system developers, has "lost some functionality." During our evaluation, much of the old version was nonfunctional. We did not attempt to determine the utility of the nonfunctional software because of the lack of documentation describing its intended use. Some features or applications were probably never functional, in that menu selections were in fact place holders. Another constraining factor in our evaluation was the accuracy of the products generated by the CIP. No documentation was provided which discussed the logic behind the products.

The results reported next are based on two visits to the Adelphi Laboratory Center and the Fort Sill training exercises.

Maps

The CIP uses a computer-generated map based on DMA databases. Unlike most systems in existence today, the displayed map is not an image of a paper product. This allows greater flexibility in displaying an image on the screen. The user can query the database and selectively toggle any feature on or off. This "declutter" feature is very helpful. The disadvantage of this capability is that helpful information on paper maps is not available. Grid lines, legends, scales, and labels (names of towns, cities and villages, rivers and selected features) are not part of the database. A software capability allows the user to turn on grid lines as a separate feature. Unfortunately, the declutter function (turning features on and off) did not work and actually crashed the system. Also, in the old version, the user could not adjust or move the map around. The new version is reported to have this capability with a display of approximately 1:250,000.

There is a "snapshot" capability that results in about a 1:50,000 display, but the information is the same, which is a major difference from paper maps. How useful this will be is difficult to predict. At lower echelons (battalion and below), most users rely on 1:50,000 maps for tactical application. Higher echelons (divisions and above) routinely use 1:250,000. Perhaps, with the ability to query the terrain database, this will not be an issue. Automated map displays will allow much greater flexibility than paper maps in terms of scale. There is nothing magical about a specific scale, and automation will allow a user to determine the best scale for a given task. To interface with people using paper maps, the system should be able to go to the paper map scale. At the very least, the user should always know what scale of map is being used.

Understanding the map display is another potential problem. A map of Camp Pendleton, California, and the adjacent coastline was being used in the current exercises. With the feature menu not working, understanding the meaning of all the displays was difficult. The Fort Sill and National Training Center (NTC) maps were less cluttered and easier to use. The increased utility of these reports may be attributable to greater familiarity with the Fort Sill and NTC terrain by the system reviewers. The computer-generated map seems useful for conducting general or relatively simple map tasks. The completion of more specialized technical tasks in terrain analysis, such as those performed by intelligence personnel, will be hampered by the display.

The size of the screen display may present a problem for fire support. Fire support planners must be able to position friendly and opposing force (OPFOR) artillery and deep targets on the map. The location of the forces on the map depends on the level of command, scenario, and terrain. A large map display will be needed to support exercises in which the field artillery is positioned on the battlefield with units some distance behind the forward line of own troops (FLOT) and who are interested in deep attack on units far beyond the FLOT. Automated systems must support map manipulation to satisfy this requirement.

Finally, not all tasks require a map to be displayed at all times. The map may become distracting or even obscure other applications. For example, a map may be needed initially to construct an overlay with appropriate battlefield geometry but later work may only involve use of the overlay. Being able to turn off the map while leaving the overlay up with grid lines present would be helpful. The CIP allows users to decrease the intensity of the map display to 40%. This should be revised and lowered to zero. Users should be able to view grid lines. A good solution would be to turn off the map and have a light gray background upon which boundaries, units, and grid lines could be easily seen.

Overlay Data

Creating battlefield geometry (graphics) is a simple task using the CIP. The system is menu driven and the user is prompted by a system of user aids. However, positioning graphics is somewhat difficult, especially when the information is acquired manually from a paper overlay or an operations order (OPORD). The presence of grid lines supports the execution of this task by providing visual cues that allow users to place objects correctly. The system also has a tracker that provides cursor positions in Universal Transverse Mercator (UTM) coordinates. This is helpful, although overlays received from higher or adjacent units must be imposed on a proper scale paper map and then manually created in an automated system. A solution to the problem created when transferring information from manual data such as maps and overlays to an automated system is to require that all map overlays be created and transmitted only on the automated system.

As noted earlier, preservation of the integrity of the CIP's database is important. However, if all workstation operators have the ability to create and manipulate graphics, maintaining quality control of displays will be virtually impossible. There is a need to ensure that safeguards are in place to maintain the common database to support maintenance of situation awareness.

Multiple maps and overlays are standard features in most operations centers. The CIP was only capable of maintaining one overlay and it generally represented the current situation, although the overlay could represent another point in time. This is a limitation because planners require multiple overlays to support the planning process. Planners view the current situation as a starting point from which they can reposition forces, create new control measures, and introduce new forces. Thus, the ability of the CIP to support war gaming and planning activities is limited.

Another function inherent in overlay creation is the depiction of unit symbology. As part of the database initialization, units are created with resulting symbology. Records are established that contain unit strength and supply data. Information about unit location, velocity, and direction of movement is also stored in this database. Users can create units and place them on an overlay using the entity database. This operation is limited to creation of a symbol and geographic placement of the symbol. Task organization and order of battle information are nonfunctional in the old version.

As noted, the creation and use of overlays within the CIP was difficult. The overlay capability was limited and the functionality that did exist rarely functioned correctly. This is a critical shortcoming and repeatedly hindered our ability to support most tasks. Users must be able to create many overlays and turn them on and off at will.

Distance Estimation Tool

This is a simple tool used to measure the distance between two points. If a third point is designated, a table is used to present the distance from each point in order and to provide a cumulative distance. It is a useful low order tool that has some utility and will probably elicit a favorable response from ultimate users. By low order, we mean that it will function as a yardstick and not as a cognitive aid. There are many possible applications for this tool in estimating and determining distances and ranges. The tool will support a range of operational and logistical tasks.

Field of View Estimation Tool

The FOV estimation tool was developed to estimate the coverage of an LOS sensor for a given range. Users specify the height of the observer, the height of the target, the range of the sensor, and the angular spacing for the radii of the circle to be searched. The default spacing is 2°. In a circle with this default, 180 rays are then calculated and displayed. The display is a circle with rays showing dead spots where LOS is not possible. This tool is useful in positioning observer teams, sensors, or radio relay stations.

The terminology presented with this tool is generally geared to sensors and should be changed or expanded for other applications. A brief explanation regarding the output of the tool would eliminate the initial confusion evidenced by most users who were unsure whether the display represented visual or nonvisual areas.

Line of Sight Estimation Tool

The LOS estimation tool is similar to the FOV estimation tool, but the focus is on a single ray. Two points are indicated and the system produces an LOS profile. This tool, while somewhat more limited in scope, has the same potential application as the previous tool. A potential problem exists in designating precise points using a cursor, which may limit the usefulness of this tool.

Corridor-Maneuvering Calculation Tool

This tool is designed to determine the "trafficability" of terrain. Two displays are produced. Initially, a go, slow-go, no-go overlay is produced, and then a preferred route is indicated on the terrain background. Input items include time of day, visibility, ground condition, type of unit, direction of travel, grid size, and sensor coverage. Time of day, visibility, grid size, direction of travel, and sensor coverage have little apparent relationship to calculating a maneuver corridor. Users will be curious about these input requirements and are likely to question the output accordingly. Further, direction is seldom defined as one specific value since units maneuver left and right as needed. A major drawback to this tool is the color display. The overlays are presented using red, yellow, and green, but the differentiation between yellow and green areas was sometimes minimal.

Mobility Prediction Tool

This tool is designed to predict how far an entity could travel from a known point in a specified time. Using the terrain, features, and trafficability databases, the user selects a point of origin on the map and then inputs computational criteria. The criteria include time of day, visibility, grid size, type of unit, and prediction time in seconds. Based on these factors, the system generates a red, yellow, and green display around the point of origin depicting possible but unlikely, probable, and highly probable locations of travel. The prediction does not consider roads that a vehicle would probably follow to make better speed. The tool does have some limited application for verifying targets that are moving about the battlefield. It might also be useful in tracking and counting enemy units. One would need a rich scenario to exploit this capability.

Peak Search

The peak search tool scans a designated area to determine the highest peak in an area along with the associated grid coordinates. There are many potential applications for the peak search tool. For example, it provides a means to search for dominant terrain features in a specified area. Again, this is a low order tool, but it could be useful in determining observation points, radio relay sites, and radar sites. One problem with the tool is that the display is temporary and does not remain on the screen for a long enough period. Also, it would be useful to permanently mark high elevation points.

Perspective View Generation Tool

This tool gives the user a wire frame mesh view of the elevation of specified terrain. It provides a “snapshot” profile of the terrain. Most tools we have seen with this capability also have a real-time fly-through capability that greatly enhances their utility. The utility of this tool is limited.

Route Planning

The route planning tool is used to decide the optimal route between two points on the map. An optimal route is defined in terms of the shortest time needed to travel from one point to another. The user indicates a start point and an end point, and the system quickly generates a preferred route. One drawback to this tool is that the route selected is often contrary to what is expected and no explanation for the selected route is provided. Routes were frequently depicted with hard angular turns rather than, as would be expected, a smooth, flowing course. This effect was attributable to cell-to-cell calculation of the route with route drawing done on a connect-the-dots basis. For usability, the preferred route should be redrawn either by the system or manually by the user using the CIP's control measures capability.

Our analysis revealed that the products generated by this tool appeared irregular or nonstandard and were affected by irrelevant data such as grid size. Also, the route is not selected based on unit size or frontage, and the presence of roads did not seem to affect the preferred route. Another area of concern is the lack of documentation concerning the rules used to produce products. User confidence is enhanced when reasonable products are generated from automated systems or when explanations are available to document unexpected results. At a minimum, a user should be able to view the output from a tool and, if required, obtain information regarding the development of the output. An explanation for nonintuitive results is that the rules may be incorrect or a tool may not support a specific application. Another potential explanation is that the rules may be correct, but the data that support the rules may be inadequate.

Additional Observations

The CIP user's guide lists many mission planning applications for the system. However, we could only use two applications (the control measures and the entity database) and those two were only partially functional. The control measures were time consuming and had a poor, nonintuitive user interface. Creating control measures and displaying them in the correct color was cumbersome. Editing control measures was limited to changing the color of the measure. The moving of labels was nonfunctional. An additional edit capability that would have been desirable was a "move" function. Creating units was easy and positioning them on the map was also easy. However, units could not be posted to an overlay--a major limitation precluding planning and war gaming. Also, as a matter of utility, a capability to "hide" a unit would have been useful. This would allow for the creation of units to be used later in the exercise. The units could then be "uncovered" as required.

Control measures were individually reviewed on both the large map display and the snapshot map. The measures that did not correspond to FM 101-5-1 or measures that presented specific problems are listed in Table 1, and general problems associated with the use of the control measures are reviewed.

Table 1

Specific Control Measure Deficiencies

Control measure	Deficiency
Air space coordination area	To draw a closed box, width must be left blank.
Attack point	This measure is acceptable, but a better measure would be an attack position.
Axis of main advance	Does not produce double headed arrow for main attack.
Check point	Label is outside symbol; should be inside.
Contact point	Should be square instead of round.
Coordination point	Extremely large for snapshot map.
Direction of attack	No arrow on stem.
Drop zone	Label is outside; should be inside symbol.
Feint	No arrowhead on line.
Follow and support force	No arrowhead on line.
Infiltration route	Should be open line instead of closed lines.
Objectives	Objective name appears to the right side of the objective instead of inside the objective.
Phase line	In order to display smaller coordination points, the forward edge of the battlefield (FEBA) feature must be used and named as a phase line.
Target	For artillery use, the target reference point symbol is preferable.

1. Some symbols such as the "no fire areas" left ghosts on the screen when deleted. To delete the ghost, the screen had to be switched between the normal map display and the snapshot display. Sometimes, several attempts were required. This problem also existed with some labels when they were deleted from the snapshot map. Switching back and forth from the normal map display was required to delete the label from the display.

2. Control measures were created in black. Display of the control measure in color to support identification of friendly or enemy measures required 11 additional mouse clicks.

3. The only operable control measure edit feature was the color selector. When control measure feature labels changed, such as the renaming of a boundary, the feature had to be deleted and reconstructed with the new name.

4. The control measure edit feature did not allow for the feature to be selected by name. The feature could only be edited by selecting it with the cursor. When a person was working near the edge of the map, the measure could not be edited if the label was not on the map. The measure had to be deleted and reconstructed before it could be edited. This deficiency made creating control features in color very difficult, especially near the edge of the map.

5. The creation of control features was time consuming. Overlays to support the offensive phase of the exercise required 2 minutes to be transferred from the OPORD to the paper map. This same process required more than 1 hour using the CIP.

6. Positioning the control measures was difficult and time consuming. The grid lines were far apart (10,000 meters) and were not present in the snapshot map if the selected area fell between the grid lines. Control measures that were map coordinate dependent had to be placed by use of the tracker readout. The longer the measure, the more points that must be used, and the longer the process became to execute.

7. During the user test, the overlay feature did not work. An operator could not build overlays for contingency plans. Thus, when a new OPORD was initiated or the current order changed, the construction of the measures from the new paper overlay resulted in a significant delay.

8. In addition to problems associated with creating overlays to support the current situation, the inability to plan on the CIP was a major problem. The CIP has potential as a planning tool, but the absence of a "save" capability for overlays prevents its use for planning.

9. The creation of unit symbols was simple and relatively fast when compared to the creation of control measures. A knowledge of order of battle and labeling of map symbols was required to properly designate the individual units. Menus do not present all units, and therefore, the operator must be able to relate units of relative size for those not presented. For example, "company" size designation must be used for a "team" and "battalion" for "task force." No discrepancies in unit symbols were discovered.

10. Relocation of units was simple and fast, although a quick method of showing proposed unit locations, such as a dashed unit symbol, was not available.

11. During construction of the control measures to support the exercise, the system locked repeatedly. These lock-ups ranged from a complete system lock-up to loss of a single application. In several cases, this was not evident until use of the application was attempted. A review of the error windows usually indicated that two or more applications were running simultaneously.

12. During the test, the map database was changed from Fort Sill to the National Training Center (NTC) at Fort Irwin. A simple method for changing databases was not available and required a UNIX™-knowledgeable operator to load the other database. Written step-by-step instructions (which were not available) might also have solved the problem.

13. One brigade boundary entered on the Fort Sill map could not be edited or deleted, either by name or by using the cursor. When the database was changed to the NTC at Fort Irwin, the boundary appeared by name on both the delete and edit menus. Access was not possible from either menu. When the database was changed back to Fort Sill, the boundary was still present and could not be deleted or edited. A problem of this kind is very disconcerting to a user and requires a UNIX™ programmer to correct.

14. When an application was selected, a window showing prompts opened on the display. Prompts were added to the bottom of the window scroll and prior prompts usually appeared at the top of the window. This caused some confusion for the new operator and required caution to assure use of the proper prompts.

15. The Fort Sill database did not include the easternmost part of the Fort Sill reservation, which included part of the area played during the user test. As a result, only part of the maneuver area could be displayed on the CIP. This caused several problems. When a snapshot area was selected near the edge of the display and the snapshot selection area was allowed to extend past the edge of the map, the system locked. The second problem occurred in establishing control measures. The labels for many of the measures, such as objective, are automatically placed on the right side of the map. The label would not appear on the map. Thus, the measure could not be edited from the snapshot. Switching to the main map made editing possible only if the label was visible.

16. Selection of applications for "deletion" or "editing" was not consistent. Units had to be "deleted" by placing the cursor in the middle of the unit symbol. Selection of control measures required placing the cursor on the label. Consistency would support usability.

17. The system sub-menus on the bottom right of the screen were too faint to read. Variation of the color and intensity controls did not solve the problem. Because the menu selections were difficult to read, the wrong items were selected. Selection of the wrong command resulted in loss of the working window and, in extreme cases, system or application lock-up.

18. Networking posed an additional problem. The CIP has a very limited capability to send text messages from station to station, although a software package is reportedly available that can support this function. Also, the CIP automatically broadcasts changes in the database as they occur. This means that as users create control measures, they are instantly reproduced at other workstations. Users cannot review their work before it is transmitted to other workstations where it appears on the networked workstations without warning. This limitation will lead to confusion and, rather than facilitate battle command, will complicate staff operations.

Training

Training for the CIP operator was accomplished immediately before the user test, as would probably be the case when the CIP is used to support a Janus exercise. Selected observations about user training are provided next.

The operator trained for the CIP user test was a former artilleryman, a graduate of the Field Artillery Advanced Course, a trained tactical fire direction system (TACFIRE) operator, and computer literate. However, the operator did not have knowledge of UNIX™ or the Janus system. The qualifications of the operator are consistent with what would be expected among the user population--an advanced course student.

The operator received training for approximately 6 hours, during which time, he entered all the control measures and all the friendly and known enemy units in their initial positions to start the defensive portion of the exercise (OPORD TOMAHAWK A). During this time, numerous lock-ups and degraded conditions were encountered with the CIP. It is likely that initial training could have been accomplished in 3 hours if the system had functioned reliably. This estimate is contingent upon the qualifications of the operator. The operator must have knowledge of the order of battle and military tactics. Thus, the 3-hour training estimate does not include troubleshooting and UNIX™-related operations.

Selected Observations and Impressions

1. During the user test, a single CIP operator was able to keep the battle current as to the location of friendly and enemy forces, but the operator was not able to update control measures with the speed of the paper map users.

2. Replacing paper maps with the CIP in an exercise headquarters is a viable option, although the present version of the CIP software is too slow and cannot be used to quickly and reliably update the map display when new control measures (overlays) are required.

3. The CIP must support the operator in creating and saving multiple overlays. Currently, only one overlay can be maintained. This capability would allow exercise participants to conduct fire support planning, which is an essential part of an artillery battle command exercise. In the three artillery exercises conducted during CIP operator training and user testing, fire support planning occupied 35% to 50% of the students' time.

4. A major deficiency of the CIP was the repeated system lock-ups which caused complete system failure or a degradation in functionality. During input of the control measures for the offensive phase of the command post exercise (CPX), the system experienced three complete failures, which required rebooting the system, and four partial failures, which resulted in two rebootings of the system. In two instances, the system cleared itself when the map was reconstructed from the map display menu.

5. In its present configuration, the CIP requires a trained UNIX™ operator to be available to ensure continued operation. The system also requires a UNIX™ operator to perform functions such as changing the database, significantly limiting the usability of the CIP in an operational environment.

DISCUSSION

Based on the detailed review of the capabilities of the CIP software presented in the results section of this report, an evaluation of selected CIP capabilities, as observed in the JBSC is presented in Table 2. Map presentation, terrain analysis, and currency of operations were evaluated as positive components of the CIP. However, problems with setup time, planning, and reliability significantly degraded the usefulness of the CIP for supporting battle command activities. Each of these attributes is presented in Table 2, along with an evaluation of the attribute.

Table 2

CIP Capabilities in the JBSC

Attribute	Evaluation
Map presentation	The map presentation capability of the CIP is excellent and complements the Janus operation.
Terrain analysis	The CIP is a good tool for terrain analysis, especially in the snapshot mode. Greater effectiveness could be achieved if several applications could be run at the same time, such as corridor maneuvering and route planning.
Currency of operations	The CIP was used as a replacement for paper maps to show current position of both friendly and enemy units. Movement of units was simple and fast.
Setup time	Creation of control measures required too much time compared with the use of paper map overlays.
Planning	The lack of an "overlay save" capability prevented the use of the CIP as a planning tool.
Reliability	The CIP repeatedly experienced system failures resulting in system lock-ups or degraded operations.

Recommended CIP Enhancements

The following recommendations are given to enhance the CIP:

1. Debug and simplify the system. To be an effective tool in a training exercise or within the Janus environment, the system must be debugged to eliminate the frequent lock-ups. The system must be simplified to reduce the need for a UNIX™ operator during normal operations.

2. Provide documentation. Documentation must be provided to enable the operator to solve common problems that occur during the running of the program and prevent the "trial and error" method employed in the user test.

3. Provide a readable navigation feature. A readable navigation feature must be provided to enable the operator to determine where he or she is in the system and which applications are running.

4. Allow control measures to be created in the required map color. Control measures should be created in the required map color, rather than requiring editing from the default color of black.

5. Provide a "save" selection on the control measure menu to prevent loss of control measures when lock-up occurs while a person is working in the control measure database. Presently, control measures are saved only when the user exits the control measure menu.

6. Provide an "overlay save" capability. This feature must be usable for planning and time conservation.

7. Reconfigure the prompt menu. The prompt menu needs to be reconfigured to display only current prompts.

8. Enable all applications or highlight those available on the current menus. It is very confusing to the operator to view many selections, of which only a few can be chosen. However, if all the listed selections were available, the system would be much more acceptable to the exercise participants.

9. Provide a capability to "change the terrain database" from the map menu.

10. Provide a capability to "change the grid line interval" so it is useful on the snapshot map.

11. Provide an additional capability to "add control features" such as boundaries by entering a series of map coordinates. This would accelerate the entering of control features.

12. Allow for the simultaneous running of multiple applications and tools. This would provide the operator with the capability to use several tools at the same time in order to use multiple capabilities.

13. Allow the snapshot map to extend past the map boundary to permit editing of control measures.

14. Design a software interface for the CIP to allow it to transmit and receive distributed interactive simulation (DIS)-compatible protocols that allow the CIP to interact with other DIS-compatible simulations, simulators, or live tactical equipment.

Potential Uses of the CIP in the Janus Facility

Following is a list of potential uses of the CIP in the JBSC. These applications are only valid to the extent that the enhancements listed previously are implemented.

1. Replacement for paper maps. The CIP could replace paper maps by providing an up-to-date display of the situation.

2. Mission planning. The CIP could provide a mission planning capability both for planning unit position and routes as well as battlefield geometry.

3. Terrain analysis. The CIP could provide terrain analysis and map support. The present capability of varying the map scale and extensive use of existing tools would be required to provide distance, route, and corridor-maneuvering information for terrain analysis.

4. Developing and distributing battlefield geometry. The CIP could be used to distribute battlefield geometry to other workstations. Thus, created or saved geometry, including fire support measures, could be rapidly distributed throughout the facility through network updating, negating duplication of work across cells.

5. Planning future operations. The CIP, with an "overlay save" capability and the terrain map combined with the planning tools, could be used to support planning operations.

6. Fire support planning and distribution. The CIP could provide the Fire Support Officer with a tool to develop and distribute the fire support plan.

7. CPX Communications. The CIP could be used to replace field phones and paper maps by use of a message feature within the CIP network.

8. Placement of observers. The CIP terrain database and terrain planning tools afford the capability to position observers by determining visibility conditions.

Summary

CIP users, based on the preliminary functional assessment and the user test, identified a number of deficiencies in the CIP software. Their joint summary evaluation of the system and its potential use in the Janus facility is as follows. First, the old version of the system is obsolete and the software is not being supported (maintained or upgraded). Any further work with the CIP should be restricted to a new version of the software, although many existing commercial and Government systems are currently available that may be more advanced in terms of usability and functionality than the CIP. Before we commit to an expansion of the CIP's capabilities or its use in the JBSC, the use of one of the more advanced alternatives or upgrades to the CIP should be evaluated. Finally, the CIP, or any other digital system being tested in the simulation center must be integrated into the Janus system. Running a command and control system manually in parallel with an ongoing Janus exercise is a difficult and error-prone process. It is recommended that the CIP or the digital system to be tested be integrated into the Janus system, possibly through protocols developed for DIS applications. At a minimum, this would require the Janus model being used in the JBSC to be replaced by a DIS-compatible simulation (either Janus or some other comparable simulation driver) and would require all systems to be tested in the facility to meet the specifications for DIS compatibility.

CONCLUDING REMARKS

The realization that traditional testing concepts, methods, and assets may not be suitable for the next generation of battlefield systems has provided the stimulus for research into new approaches to testing. By new approaches, we are referring to technical developments such as DIS and virtual reality. The ability of these technologies to create artificial performance environments will stand the traditional notion of field testing on its head. Already, for example, some user and operational tests have been performed using the mounted warfare test bed at Fort Knox, Kentucky. Recent fire support and air defense warfighting demonstrations and experiments have also been conducted using DIS-based synthetic performance environments.

The current situation, vis-à-vis OT&E and synthetic performance settings, is just the beginning. In the not-too-distant future, it will theoretically be possible to assemble nearly any combination of constructive (models), virtual (simulators), and live (instrument actual equipment) simulations needed and to link them in real time using a satellite network such as the Defense Simulation Internet (DSI) creating a distributed, dynamic, and realistic synthetic theater of war (STOW) environment. The attraction of this approach is that constructive, virtual, and live components can serve as building blocks for the development of complex performance environments (STOW) which are more controlled and precise than field settings but cost less to set up and operate.

The new simulation technologies hold tremendous potential for the future of testing, although in many respects, they represent an idea whose time has not yet fully arrived. Currently, DIS-based artificial performance environments suffer from a variety of limitations. Line and node "crashes," lack of protocols defining full DIS compliance, and inadequate constructive models can combine to make a synthetic performance environment unsuitable for realistic testing. Further, the technology needed to integrate live tactical battle command systems essential to the technical or operational evaluation of battle command concepts or systems has lagged behind our ability to link constructive simulations and virtual simulators, although pioneering efforts by ARL and the D&SA Battle Lab (Bouwens, Ching, & Pierce, 1996; Copenhaver, Ching, & Pierce, 1996) have allowed for the integration of fire support command and control systems and DIS-compatible simulations. The problem still exists, however, and before committing to the use of a simulation test bed, test planners and system developers must assess proposed test-bed-based applications to determine whether they can deliver what they promise and what is required for effective testing.

The acquisition and use of information age technologies will challenge our ability to evaluate new or modified systems or concepts in operationally sound environments. The use of simulation-based testing provides our best means to conduct early and frequent user testing to ensure that the systems acquired to support the 21st century soldier provide the technological edge needed to fight and win the information war. Relatively small scale, controlled research efforts, such as the one described in the current report, are necessary to establish the methodology for testing procedures to meet the demands of a rapidly changing environment as evidenced by the information age.

REFERENCES

Army Research Laboratory (1992a). Combat information processor. U.S. Army Research Laboratory, Advanced Sensor Systems Branch, Adelphi, MD: Author.

Army Research Laboratory (1992b). Software user's manual for the combat information processor (CIP), version 1.1. U.S. Army Research Laboratory, Advanced Sensor System Branch, Adelphi, MD: Author.

Bouwens, C.L., Ching, H.L.F., & Pierce, L.G. (1996). Development and engineering of a distributed interactive simulation system (ARL-CR-293). Adelphi, MD: U.S. Army Research Laboratory.

Copenhaver, M.M., Ching, H.L.F., & Pierce, L.G. (1996). Training effectiveness evaluation of an MLRS fire control panel trainer using distributed interactive simulation system (ARL-CR-29r). Adelphi, MD: U.S. Army Research Laboratory.

Headquarters, Department of the Army (1985). Operational terms and symbols (Field manual 101-5-1). Washington, DC: U.S. Government Printing Office.

APPENDIX A

USING SIMULATION TEST BEDS IN OT&E

USING SIMULATION TEST BEDS IN OT&E

Overview

Meister (1986) remarks that the fundamental issue underlying human (i.e., soldier) performance measurement in OT&E is, "How does human performance, which is an intermediate output for the system as a whole, influence total system performance?" In military OT&E, the essence of this issue is captured in the related question:

> Can this soldier, in this organization, with this training, perform the following tasks on this system to established standards and, if not, why not?

The central problem in OT&E is the conduct of an experiment or series of experiments to answer this question. During these experiments, the MANPRINT team's objective is to obtain quality measures of soldier performance of the tasks in question that generalize beyond the test setting to the broader arena of military operations. Quality, in present usage, refers to the military operations. Quality, in present usage, refers to the reliability and validity of performance measures. We define a reliable measure as one that does not contain excessive measurement error. A valid observation is one that does not contain systematic irrelevant variation.

Anyone familiar with field OT&E is aware of the special problems of conducting rigorous performance measurement in an operational setting. To begin, there is never enough money, never enough time, and never enough test subjects to do it "right." Moreover, a field environment is different from a laboratory setting in several ways that affect the ease with which measurement can be conducted and the validity and reliability of the resulting data. These differences are summarized in Table A-1.

Simulation test beds represent a hybrid performance setting--somewhere between a field setting and a laboratory environment. Therefore, the conditions encountered in a simulation-based performance setting will often fall somewhere between the extremes represented by laboratory and field settings. Planning for OT&E using a simulation test bed must be sensitive to the impact of the factors listed in Table A-1 on test conduct and the eventual utility of the test results.

Table A-1

Field Testing Versus Laboratory Experimentation

Characteristic	Laboratory	Field setting
Experimental error	Multiple replications	One or a few trials
Experimental control	Matched or control groups	Usually a single group of test subjects
Experimental protocol (mission scenarios)	Well defined	Less well defined--sometimes approximates "free play"
Intrusive factors	Eliminated or controlled	Often uncontrolled
Physical environment	Controlled or artificial	Natural or operational
Time units measured	Short	Continuous mission segments
Experimental variables dictated by	Experimenter interest	Test limitations or system considerations
Subject source	Varied--determined by experimenter	User population
Subject attitude toward test	Positive or neutral	Neutral to negative

Source: Adapted from Johnson and Baker (1974)

In spite of the difficulties traditionally associated with field testing (many of which will carry over into a simulation-based test setting), there is a set of basic testing principles that should guide OT&E practitioners in any testing situation. These principles are summarized in Table A-2 and represent a foundation for good testing in any situation. These principles are a solid point of departure for a discussion of potential improvements in the theory, technology, and practice of OT&E.

Early User Tests

In this report, we refer to a class of OT&E called early user tests (EUTs). One objective of the effort is to explore the use of simulation test beds such as the JBSC as vehicles for conducting EUTs of complex, information age systems. Therefore, we judge it necessary to define the terms "user test" and "early user test."

Table A-2

Basic Testing Principles

1. Testing is measurement--the assignment of numbers to systems or subsystems to represent properties of interest.

2. Testing represents a compromise between experimental and methodological rigor and operational reality.

Corollary: In testing, the old adage KISS (keep it simple, straightforward) should rule the day. Complex designs and procedures often spell trouble for a testing program.

3. Testing is not glamorous and has only a few basic principles. These principles must be observed rigorously.

Corollary: When assessing the cost effectiveness of any test design, one must consider the costs of doing it wrong. A "bad" test often costs as much to conduct as a "good" test.

4. The degree to which basic principles of experimental design, measurement, and statistical analysis are enabled or observed during testing determine the limits of inference and generalizability of any testing situation.

Corollary: Clever statistics will not compensate for an ill-conceived test design, poor measurement methods, or faulty test execution.

5. User and operational tests will span a continuum ranging from simple demonstrations to rigorous experimental tests. Tests are positioned on this continuum by the extent to which requirements for inference and generalizability are met.

6. Soldier performance measurement in a complex human-machine setting is not trivial. Further, soldier performance measurement will not get any easier in the emerging era of information age systems.

7. The technology of good testing is well known. The problem is observing good testing practices within a dynamic, realistic, cost-constrained, and sometimes perverse setting.

Source: Adapted from Hawley and Frederickson (1990a)

"User testing" is a generic term that refers to OT&E conducted with user-representative troops during Early Tests and Experimentation (EUTEs), Force Development Tests and Experimentation (FDTEs), Innovative Tests, Concept Evaluation Programs (CEPs), Initial Operational Tests and Evaluation (IOTEs), and Follow-On Tests and Evaluation (FOTEs). In general, user tests address (a) a system's operational effectiveness and (b) supportability considerations. The specific objectives of a user test can include any of the following:

- Assess tactical concepts
- Assess initial manpower, personnel, and training concepts for a system
- Develop an initial set of tactics, techniques, and procedures (TTPs)
- Identify interoperability problems
- Identify future testing requirements
- Refine procedures for providing test player personnel

"Early user testing" is defined as (p. 12):

> Operational testing conducted during the Proof-of-Principle phase of the Army Streamlined Acquisition Process to support a Milestone I or II decision. Early user tests are conducted to prove out both the technical approach (primarily from the standpoint of the man-machine interface) and the operational concept. Early user testing provides an opportunity for early involvement of soldiers in the OT&E process. New systems may be configured as breadboards, brassboards, or early prototypes. **Early user testing may also involve the testing of components as surrogates.**

The final sentence of the previous paragraph is bold to emphasize that an early user test conducted using a simulation test bed is an example of "testing of components as surrogates." That subject is the topic of the present report.

After reviewing the various documents addressing user testing concepts, Hawley and Frederickson (1990a, p. 12) proposed the following umbrella definition for an early user test:

> An early user test is any OT&E-like activity involving user-representative personnel conducted between Program Initiation (Milestone 0) and Milestone I or II to support an initial assessment of a system concept.

In the present effort, we will use this description as a working definition of an early user test.

Test and Evaluation Preliminaries

In the following subsections, seven issues that must be addressed when using a simulation test bed in OT&E are listed and discussed. Several issues are peculiar to a synthetic performance setting such as that provided by a simulation test bed. Others apply to testing overall but are of particular significance in a simulation test bed because of the greater opportunities for precision and experimental control. We include this latter group of topics here because they are critical to a successful test. In present usage, a successful test is one that (a) fully addresses all test issues and (b) produces results that generalize beyond the test setting to the broader arena of field military operations.

The System or Concept to be Tested

In field OT&E, many systems and concepts brought to test are not fully developed or are incomplete when they arrive at the test site (see Hawley & Frederickson, 1990a). As a result, anticipated system capabilities do not work and breakdowns are frequent. Nonfunctional or unreliable equipment can wreak havoc on an operational test. System training cannot take place if equipment does not work as documented. Moreover, frequent equipment breakdowns or software failures disrupt testing and have a ripple effect on test player performance that extends well beyond the interrupted task sequence. These same problems will affect OT&E conducted in a simulation test bed. In fact, since many systems brought to test in a synthetic performance setting will not be systems in a traditional sense, the problems associated with an incomplete or immature concept may be compounded with even more damaging consequences. An immature system or concept can sink the test before it begins.

Every test is based upon a system concept, even if that concept is not fully articulated. Test planners must not go to a test with an ill-defined system concept. A clear and comprehensive definition of the system or concept to undergo testing must be prepared. In addition, the human operators' role in the system must be defined, and human functions must be identified and described. These role and function descriptions are the basis for developing doctrine, organization, and operational procedures.

Doctrine, Tactics, and Organization

During many operational tests, we have observed that test results are often compromised by (a) ill-defined or invalidated doctrine and tactics and (b) lack of regard for the performance-shaking effects of organizational factors. Ill-defined or invalidated doctrine and tactics can lead to a situation where TTPs are debugged "on the fly" during record trials. Consequently, changing TTPs also means that test player performance does not have a chance to stabilize.

With respect to organization, test planners often do not consider the impact of unit structure and command and control relationships on test player performance. An organization is more than the sum of its parts, and melding the disparate parts into a functioning unit takes time. If this "gelling" process has not occurred, it will take place during the initial stages of the test and will differentially affect test player performance.

The recent biological integrated detection system (BIDS) test provides a case in point regarding the effects of organizational factors (see Hawley, Dawdy, Rozmaryn, & Wilkinson, 1995). Test players had been trained to operate individual BIDS components, but they had minimal training in using the components as a suite to meet mission requirements. A significant aspect of these mission requirements involved several BIDS teams operating together as a platoon; however, many structures (standing operating procedures [SOPs], TTPs, command and control relationships, coordination patterns, etc.) that define a BIDS platoon had not been specified. Confusion regarding organizational relationships persisted throughout the test. To further complicate data interpretation, individual BIDS teams developed their own nonstandard procedures in an attempt to cope with the situation. Many soldier performance problems recorded during the test reflected this lack of organizational definition.

There is an old adage that nature abhors a vacuum. This idea can be extended to OT&E conducted with loose doctrine and TTPs. If workable doctrine and TTPs are not provided, soldiers will develop their own in an attempt to handle test demands. The result will be a nonstandard "mishmash" of procedures and techniques across soldiers, teams, and units. Obtaining meaningful performance data during such conditions is nearly impossible.

Defining doctrine, TTPs, and organization is outside the scope of most OT&E support efforts. Their development is the responsibility of the Directorate of Combat Developments or Battle Lab. Test planners must be sensitive to the potential impact of these factors on test player performance. Further, if unit or team performance is critical, planners must bear in mind that stable performance usually cannot be developed in a week or two.

The Synthetic Performance Environment

If a simulation test bed is to be used in OT&E, the test proponent must be able to assemble and support the technical infrastructure necessary to create a suitable performance setting. This issue is absolutely critical to the success of simulation-based OT&E. If the performance setting is unstable or otherwise unsuitable, test results may be compromised beyond recovery. The worst case outcome in this respect is that the test may never get off the ground.

When considering the suitability of the synthetic performance setting, two issues must first be addressed: (a) technical feasibility and (b) physical and functional fidelity. Key aspects of these issues are summarized in the following subsections.

Technical Feasibility

A prerequisite for using a simulation test bed in OT&E is the proponent's ability to assemble and support the technical infrastructure necessary to conduct the test. Since most OT&E applications using a test bed will involve DIS, the primary factors defining technical feasibility pertain to the ability of potential constructive simulations, virtual simulators, and live players to maintain a suitable performance network. Another way of characterizing this factor is the "DIS-compliance" of proposed players. The major variables defining DIS-compliancy are as follow (Institute for Simulation and Training, 1993).

- Interoperability. The ability of entities to register their interactions within the synthetic performance environment.

- Network access and capacity. The availability of and access to networks with sufficient capacity to handle real-time data transmission requirements.

- Correlation of environments, entity models, and outcomes. The network's ability to create and maintain essential space, time, and entity correlations within the synthetic environment.

To this list, we add a fourth consideration.

- Demonstrated capacity. The network's ability to successfully demonstrate the integration of the necessary constructive, virtual, and live simulations.

The following questions must be satisfactorily resolved. Are the test bed and any ancillary equipment and supporting software and models ready now? Has the proposed capability been proven in exercises similar to the proposed test? Was the performance network stable, or were line or node crashes frequent? The test officer must be tough with respect to these and related questions; "fly before buying" and do not be deceived by unproved claims. The time and resources allocated to most user tests will not permit extensive debugging of test bed capabilities after the exercise begins. If the test bed must be modified or enhanced before the exercises can be conducted, verify that modifications and enhancements work before starting the test.

The Physical and Functional Fidelity of the Performance Setting

Beyond superficial technical feasibility (i.e., the test bed apparently "works"), the test bed must also provide adequate physical and functional fidelity for the exercise of human functions. Fidelity is the similarity between the synthetic performance environment and the operational situation simulated (Hayes & Singer, 1989). It is standard practice to characterize fidelity in terms of two dimensions: (a) physical fidelity and (b) functional fidelity. Physical fidelity refers to the congruence between physical aspects of the synthetic and operational environments. Functional fidelity is defined in terms of the similarity of task demands (e.g., performances, initiating and terminating cues, information flow and availability, stimulus and response timing, etc.) between the synthetic and operational environments. Physical fidelity can often be reduced without significant impact on the validity of the simulation. However, most aspects of functional fidelity must be preserved if the synthetic performance setting is to provide meaningful results.

Simulations and Test Scenarios

Supporting Simulations

In a test-bed-based OT&E exercise, the enabling simulation models--software embedded within virtual nodes or supporting constructive models--will have a significant impact on the utility of test results. The world view and level of detail of supporting simulation models will determine the validity and thus the generalizability of test data. World view, in present usage, refers to a model's assumed conditions of use and how a model's designers chose to treat the various entities and phenomena of interest. For example, simulation models used to support training (e.g., Janus) run in real time or faster. Such models sacrifice simulation detail to maintain the time fidelity of an exercise. Other models such as the Combined Arms Task Force Engagement Model (CASTFOREM) provide a high level of simulation detail but were not intended to support training or the running of real-time exercises.

Every simulation model has limitations that reflect trade-offs among (a) the real-world situation being modeled, (b) design goals, and (c) usage realities. No model provides an exact simulation of all aspects of the real world. Test planners must be aware of the world view and limitations of proposed simulation models and the potential impact of these constraints on test objectives.

Test Scenarios

Test scenarios are another determinant of a test's utility or value. Scenarios provide the stimuli necessary to drive the system according to doctrine and exercise-essential soldier functions. They also are an important factor in a test's external validity. Campbell and Stanley (1966) define external validity as the certainty with which test conclusions can be generalized across different persons, settings, and times. Practically speaking, external validity depends on the operational fidelity of the test situation--simulation validity, system features, user personnel, threat, and range of operating environments.

In a field test, scenario content is often restricted by considerations such as cost, range availability, safety, and the like. Many of these factors will no longer be important in a simulation-based test. Eliminating such factors as determinants of scenario content is a primary consideration when deciding to use a synthetic versus a field test setting. Whatever the setting, test scenarios must be reviewed to determine which measures of performance (MOPs) and data requirements (DRs) are supported by the test design. If test scenarios do not provide the stimuli for a particular response, data about that performance cannot be obtained. Similarly, limits on the operational fidelity of a test situation can affect the generalizability of MANPRINT-related conclusions. For example, performance times and error rates for many types of soldier tasks increase significantly during extreme or boundary conditions. If test scenarios are too benign, documenting the performance impact of extreme conditions will not be possible.

Test Player Selection and Training

Test Player Selection

Test player selection and training can have a significant impact on test results. Considering, first, test player selection, an all-too-common practice in OT&E is "creaming," or packing the test player sample to include only soldiers in the upper portion of the military occupational specialty (MOS) aptitude distribution. Creaming can produce test results that do not generalize to the target MOS population. In cases of excessive creaming, test performance will overestimate later operational performance.

Test Player Training

One of the most serious and recurring problems in OT&E is inadequate test player training and preparation. Following an in-depth review of a cross section of early user tests, Hawley and Frederickson (1990a) reported that inadequate test player preparation was one of the most frequent reasons for test "failure." The issue of test player training assumes even more importance with the arrival of the new generation of information age systems (e.g., computer-based decision support systems, command and control systems, communications systems, and the like). Information age systems are complex and require a high level of expertise for effective use. Moreover, the progression from novice to journeyman to master performer in such systems takes time and appropriately structured experiences (e.g., see Salas, Prince, Baker & Shrestha, 1995). Test planners cannot expect that anything useful will come from an exercise in which relatively unskilled test players are thrown together with a complex but loose system concept and minimal doctrine and TTPs. If test player performance capabilities are uncertain, test results are likely to be compromised. The most likely form of compromise is confounding between test performance and pretest proficiency levels. It will not be possible to state unambiguously that the test outcome reflects system capabilities, test player proficiency levels, or some combination of the two.

As with test player selection, test personnel often have little say about the design or conduct of test player training. We can remind test managers of the importance of adequate training; we can document the training that takes place before testing begins; we can attempt to measure test player proficiency levels at the end of training; and we can try to relate pretest proficiency levels to later test performance. Nevertheless, if the past is any indication of the future, we will rarely succeed in delaying testing because of test player training deficiencies.

Soldier Performance Measurement

In the overview section, we noted that the fundamental question underlying human performance testing is the impact of human performance on total system performance. Soldier performance and its effect on overall system performance should be the primary concern of MANPRINT during OT&E. Yet, the MANPRINT chapters in many test and evaluation plans (TEPs) do not even mention human performance. This is a shortcoming that must be corrected if MANPRINT OT&E is to produce its intended result of characterizing the relationship between soldier performance and system capability.

Eddy (1989) lists six steps in the development of an effective performance measurement process:

- Perform a comprehensive system and job analysis
- Identify critical soldier tasks
- Determine performance requirements for critical tasks
- Select measures appropriate to the behaviors to be evaluated
- Determine the conditions under which to measure performance on critical tasks
- Choose techniques for recording measurement data and for combining individual MOPs into aggregate measures of effectiveness

For the most part, conducting this process is not difficult, but three recurring problems must be overcome for quality (i.e., reliable and valid) performance measurement to take place. These problems are lack of (a) a detailed task list during early user tests, (b) operational definitions of MANPRINT domains for OT&E, and (c) standard human performance measures and moderators. Each of these topics is discussed in the following subsections.

Task Identification

Three general sources of soldier performance data are available during OT&E: (a) subject matter expert (SME) observations and ratings, (b) test player interview and questionnaire results, and (c) task time and error data. The preferred type of soldier performance data is task time and error results. In many early user tests, obtaining reliable and valid time and error data is difficult because task lists and supporting task analysis results (i.e., task steps and enabling skills) do not exist. The root cause of this deficiency is the lack of a comprehensive system and job analysis early during the system development process. System immaturity contributes to this problem as does a lack of emphasis on early training products by combat and materiel developers.

With no validated task list, MANPRINT practitioners are often forced to develop one before testing begins. These "seat-of-the-pants" task lists are often not comprehensive and lack essential detail. Consequently, they are not a satisfactory basis for rigorous task time and error analyses. The problem of missing or inadequate job analysis results has been aggravated by the elimination of the Directorates of Training and Development (DOTDs) in many TRADOC schools. Nobody within the system development community has assumed responsibility for previous DOTD tasks.

Operational Definition of MANPRINT Domains for OT&E

We noted previously that the MANPRINT chapters in many TEPs do not mention soldier performance. The usual emphasis is on SME observations and ratings and test player reports. Without an explicit requirement in the TEP for the task time and error data, justifying expending test resources to obtain such data is difficult. Test managers generally hold to the position that if something is not specified in the TEP, it will not be done.

Part of the problem here is unfamiliarity with MANPRINT methods by the test cadre who write TEPs. We think that a contributing factor is that the MANPRINT technical domains (i.e., manpower, personnel, training, human engineering, system safety, health hazards, and soldier survivability) are not operationally defined in terms of soldier performance measures, moderators, and shaping factors. In present usage, a performance moderator is a variable that affects performance differentially for distinct subgroups (cf. Lord & Novick, 1968). Operator workload (OWL) and situation awareness (SA) are examples of performance moderators. Performance-shaping factors are intervening variables between human performance and system performance. For example, organization at various levels within a unit is a performance-shaping factor for both manpower and personnel.

The best way to operationally define the MANPRINT technical domains for OT&E would be to develop a set of exemplary domain (sometimes called criterion) definitions along with subordinate MOPs and Drs. Test personnel charged with developing a TEP could tailor these exemplary criteria, MOPs, and DRs to suit the system undergoing consideration. This baseline structure for MANPRINT OT&E would improve the testing process by (a) focusing MANPRINT data collection on soldier performance; (b) defining the MANPRINT technical domains in terms of observable soldier performance measures, moderators, and shaping factors; and (c) providing more commonality and consistency across tests. Above all, operational definition would ensure that soldier performance issues get into the TEP.

Standard Human Performance Measures and Moderators

The use of standard performance measures and moderators in OT&E should be encouraged. When possible, the tendency of MANPRINT practitioners to constantly "re-invent the wheel" with respect to performance measures should be avoided. Standard measures, moderators, and assessment procedures will increase the reliability and validity of test results and be useful in root cause analyses of observed performance failures. The catalog of battle staff performance measures contained in Lowry (1955) is an example of standard performance measures for command and control applications. Lysaught et al. (1989) and Endsley (1995) list various indices of OWL and SA, respectively, that can be used as performance moderators. MANPRINT practitioners must bear in mind that OWL, SA, and other constructs are performance moderators and not measures of performance per se.

Despite their utility, standard measures and moderators should not be prescribed blindly. They must be screened with respect to their suitability in a given situation. In the BIDS test, for example, the TEP prescribed the use of NASA's Task Load Index (TLX) OWL metric (see Hart & Staveland, 1988). The TLX was not particularly suitable for use with a system such as BIDS. TLX results are most meaningful when used comparatively--OWL with respect to a baseline condition or predecessor system. In BIDS, no reference point for OWL ratings was available. Consequently, the L results had little utility. A preferred approach would have been to identify OWL as a significant MANPRINT issue and then let the test team determine the most appropriate measurement approach.

Experimental Design and Research Methodology

For a variety of reasons, the exercises performed during military OT&E usually become quasi-experiments (see Hawley & Frederickson, 1990a). A quasi-experiment has treatments, outcome measures, and experimental units but does not use random assignment to create the comparisons from which treatment-caused change can be inferred (Cook & Campbell, 1979). Quasi-experimental designs are not as robust as classical experimental procedures based on randomization. These designs are often subject to a range of threats to valid statistical inference such as history, maturation, selection, testing, reliability of treatment implementation, and statistical regression. Chapter 2 in Cook and Campbell (1979) defines and discusses these and other threats to valid inference in quasi-experimentation. Threats to valid inference are often subtle and go unnoticed. Therefore, if statistical comparisons are planned, care must be taken to ensure that the potential range of threats is identified and addressed.

There is an interaction among resources, design complexity, and the level of inference possible in a test setting (Miles & Hawley, 1991). Several of these relationships are illustrated in Table A-3, and definitions are given in Table A-4. Miles and Hawley argue that the level of inference possible in an OT&E exercise depends on design factors such as the use of randomization, number of replications, level of measurement, and the like.

In OT&E, these factors are usually constrained by resources, test subject availability, and other limitations. Hawley and Frederickson (1990a) found that the level of inference planned for a test typically exceeded what was possible given the constraints imposed by the test design. That is, test planners often attempted to "get more out of" a test than design constraints would permit. Consequently, the useful results obtained from many tests cost more than would have been necessary had more thoughtful experimental procedures been used.

Table A-3

The Relationship Between Test Design Factors and Levels of Permissible Inference and Generalizability

	Demonstration		Tests	
Type:	Simple	Complex	Simple	Complex
Resources required	Scant	Modest	Extensive	Abundant
Nature of results	Narrative descriptive criterion-referenced measurement	Descriptive statistics	Limited statistical modeling and comparisons	Extensive statistical modeling and comparisons
Level of permissible statistical inference	Restricted	Low	Moderate	High
Generalizability of test results	Restricted	Low	Moderate	High

Source: Adapted from Miles and Hawley (1991)

Table A-4

A Taxonomy of OT&E Exercises

- ♦ Demonstration: A practical showing of how something works or is used. Requirements for statistical inference and generalizability of results are not met.
- Simple Demonstration: Single or a few replications; restricted environmental fidelity; criterion-referenced (e.g., Go/No-Go) measurement.
- Complex Demonstration: Multiple replications; possible multiple environments; descriptive statistics or criterion-referenced measurement.
- ♦ Tests: An exercise in which the requirements for statistical inference and generalization of results are met.
- Simple Test: A test conducted during conditions of restricted operational fidelity.
- Complex Test: The test environment approximates that of the operational environment.

Source: Adapted from Hawley and Frederickson (1990a)

Based on their critique of a representative cross section of user tests, Hawley and Frederickson (1990b) argue for simplicity in the experimental designs used in OT&E. Their earlier (i.e., 1990a) results showed that design complexity was a significant factor in the failure of many tests to satisfy their objectives. Besides simplicity, two other design-related considerations are important in cost-effective OT&E:

- Do not let design complexity exceed the limits of inference. The limits of inference can be determined from a review of test plans. These limits can often be used to reduce test complexity and divert scarce resources into more pressing areas.
- If statistical comparisons among groups are planned, verify (a) the groups' initial quasi-comparability and (b) that any differential treatments are reliably performed.

Regarding Point 2, quasi-comparable means that control and experimental groups are statistically equivalent on pretest characteristics of interest. In addition, one of the most significant threats to valid inference in military operations research is reliability of treatment implementation. The treatments defining the levels of an experimental condition must be conducted rigorously and equally. A situation must not be allowed to exist where one treatment (usually the "preferred" one) is rigorously applied while others are conducted haphazardly. If this situation occurs, it will not be possible to state unambiguously that the experimental condition and not something else is the cause of observed performance differences.

A Final Comment on Test Design

In any test situation, data analysts must be able to "peel the onion" with respect to the impact of (a) hardware capabilities, (b) test player aptitudes and performance capabilities, and (c) environmental conditions on system performance in the test setting. The test design must permit the effects of each of these factors to be estimated independently. A situation must not be allowed to occur in which some or all of these factors are confounded with the result that the root causes of system performance deficiencies cannot be determined. If that happens, much of the effort devoted to the test will effectively have been wasted.

Methodological (i.e., experimental design, measurement, or statistical analysis) failures are often the Achilles' heel of military OT&E. Poor planning, overly complex or otherwise unsuitable test designs, lack of test control, and the "fog and friction" of a free-play environment frequently combine to produce flawed and uninterpretable results. From a methodological point of view, military OT&E (field or simulation-based) will always be fraught with difficulties, but these problems need not be fatal. With a modicum of proper planning and methodology, many of these problems can be avoided. Moreover, smart planning in the design and methodology area is one way of doing more with less, and doing more with less is an important consideration in the present resource-constrained era.

DISCUSSION

To place the previous material in perspective, we think it worthwhile to conclude by considering the factors that make a MANPRINT OT&E program useful. What comprises MANPRINT value added? What must MANPRINT practitioners do to be viewed as useful and productive members of the OT&E community? From experience, we think that any MANPRINT OT&E program must, at a minimum, provide data relevant to two uses:

- Shooting performance bugs. Provide data about the root causes of observed performance failures: What happened? Is it important? Why did it happen? What can be done about it?
- Validating or debugging a system's personnel and usage concepts. Develop recommendations for changes in a system's (a) staffing concept, (b) soldier-machine function allocation concept, (c) crew or team division of labor, (d) aptitude prerequisites, (e) proposed training regimen, (f) design, (g) work flow, or (h) operating procedures.

Data concerning these issues can be obtained at various levels ranging from observations and opinions by MANPRINT SMEs to rigorous time and error analyses of test player performance data, often supported by extensive modeling excursions. Currently, the standard approach is a combination of SME observations and user opinion data supported by limited time and error results.

Purists often criticize the conventional approach to MANPRINT OT&E. They argue that rigorous time and error analyses must be part of all MANPRINT OT&E exercises. In an ideal world, we would agree with the view that rigorous time and error analyses should be the standard. As one ascends the scale of MANPRINT data rigor from SME observations to a mix of observational and user opinion data to rigorous time and error analyses, the data collection cost function increases at a nonlinear rate. Questioning whether rigorous analyses of time and error data are always warranted is legitimate. In early user tests, for example, the usual situation is that doctrine and TTPs are not well defined, and test players are not well trained in the application of whatever exists. A rigorous analysis of highly variable data will still yield inexact conclusions.

In MANPRINT OT&E, we have observed that a variation of the well-known Pareto principle often applies: An 80% "solution" for the two uses cited previously can be developed from 20% of a standard data collection effort, particularly if that effort is well structured. In the present cost-conscious times, analytical rigor must be balanced against cost and the potential utility of results. Rigorous and comprehensive analyses are important and have their place, but such analyses are not necessary or justified in all situations.

OT&E is often an exercise in "satisficing"--doing the best job possible within the time and resources available. An excessive concern for the "best" methods and approach can produce less in the way of usable results than a carefully crafted program using simpler methods. As noted previously, real-world OT&E is not glamorous and has only a few basic principles. These principles must be observed rigorously, and there are no silver bullets. Success comes to those who best manage the issues addressed in the previous section. One or more of these issues will be a problem in every test; they can be managed but not eliminated.

Table A-5 presents a summary of the issues discussed in the previous section. Test planners can use the points listed in Table A-5 as a checklist against which to assess their readiness to proceed with operational testing using a simulation test bed.

Table A-5

Test Readiness Checklist

Issue	Test readiness criteria
1. System or concept to be tested	1. Has a clear and comprehensive description of the system or concept to be tested been prepared? 2. Have the humans' roles in the system been defined? 3. Have human functions been identified and described?
2. Doctrine, tactics, and organization	1. Have preliminary doctrine and tactics been defined? 2. Have TTPs been drafted and reviewed? 3. Has unit structure been defined and have unit SOPs been drafted and reviewed?
3. The synthetic performance environment	1. Is the proposed performance network (nodes and communications capabilities) technically feasible? 2. Have the network and its components been tested in a configuration similar to that planned for the OT&E exercise? 3. Have physical and functional fidelity requirements been defined? 4. Will the synthetic performance environment provide the necessary physical and functional fidelity?
4. Simulations and test scenarios	1. Have enabling simulation models been reviewed with respect to their ability to (a) adequately model phenomena of interest and (b) provide the necessary level of detail in test results? 2. Have test scenarios been reviewed to ensure that they will drive essential soldier functions? 3. Have test scenarios been calibrated with respect to difficulty?
5. Test player selection and training	1. Is the test player sample representative of the target MOS population? 2. Have test player training plans been reviewed with respect to their adequacy? 3. Was test player training conducted according to plan? 4. Were test player proficiency levels measured before the start of testing?
6. Soldier performance measurement	1. Are MOPs and DRs based on the results of a comprehensive front end analysis of soldier performance requirements? 2. When possible, are MOPs and DRs based on observable soldier performance measures, moderators, and shaping factors? 3. When possible, will standard human performance measures and moderators be used during the test?
7. Experimental design and test methodology	1. Does the level of planned inference or generalizability exceed that supported by the test design? 2. Have threats to valid inference been identified and provisions made for their control? 3. Were experimental groups quasi-comparable before the start of testing? 4. Were any differential treatment conditions reliably and equally applied? 5. Are plans for test administrative control adequate?

REFERENCES

Campbell, D.T., & Stanley, J.C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston:

Eddy, D.R. (1989). Selected team performance measures in a C3 environment: An annotated bibliography (USAFSAM-TM-87-25). Brooks Air Force Base, TX: U.S. Air Force School of Aerospace Medicine.

Endsley, M.R. (1995). Measurement of situation awareness in dynamic systems. Human Factors, 37(1), pp. 65-84.

Hart, S., & Staveland, K. (1988). Development of NASA-TLC (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), Human mental workload. Amsterdam: Elsevier.

Hawley, J.K., & Frederickson, E.W. (1990a). MANPRINT guidelines for soldier performance reliability in early user test and experimentation (EUTE): A formal EUTE framework. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hawley, J.K., & Frederickson, E.W. (1990b). MANPRINT guidelines for soldier performance reliability in early user test and experimentation (EUTE): A primer. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hawley, J.K., Dawdy, E.D., Rozmaryn, S.T., & Wilkinson, T.D. (1995). MANPRINT test of the biological integrated detection system (BIDS). Ft. Hood, TX: U.S. Army Research Laboratory, Human Research and Engineering Directorate, TEXCOM Field Element.

Hayes, R.T., & Singer, M.J. (1989). Simulation fidelity in training system design. New York: Springer-Verland, Inc.

Institute for Simulation and Training (1993). The DIS vision. Orlando, FL: Author.

Johnson, E.M., & Baker, J.D. (1974). Field testing: The delicate compromise. Human Factors, 16(2), pp. 203-214.

Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Lowry, J.C. (1995). Battle staff competency model and measures. Fort Sill, OK: U.S. Army Research Laboratory, Human Research and Engineering Directorate, USAFAS Field Element.

Lysaught, R., Hill, S., Dick, O., Plamondon, B., Linton, P., Wierwille, W., Zaklad, A., Birrner, A., & Wherry, R. (1989). Operator workload: Comprehensive review and evaluation of operator workload methodologies (ARI Technical Report 851). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Meister, D. (1986). Human factors testing and evaluation. Amsterdam: Elsevier.

Miles, J.M., Jr., & Hawley, J.K. (1991). Planning early operational assessments of complex manned systems in cost-conscious times. ITEA Journal, XXII(1), pp.34-37.

Salas, E., Prince, C., Baker, D.P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. Human Factors, 37(1), pp.123-136.

NO. OF COPIES	ORGANIZATION
2	ADMINISTRATOR DEFENSE TECHNICAL INFO CENTER ATTN DTIC DDA 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CS AL TA RECORDS MANAGEMENT 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CI LL TECHNICAL LIBRARY 2800 POWDER MILL RD ADELPHI MD 207830-1197
1	DIRECTOR US ARMY RESEARCH LABORATORY ATTN AMSRL CS AL TP TECH PUBLISHING BRANCH 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTORATE FOR MANPRINT ATTN DAPE MR DEPUTY CHIEF OF STAFF PERSONNEL 300 ARMY PENTAGON WASHINGTON DC 20310-0300
1	DIRECTOR ARMY AUDIOLOGY & SPEECH CENTER WALTER REED ARMY MED CENTER WASHINGTON DC 20307-5001
1	OUSD(A)/DDDR&E(R&A)/E&LS PENTAGON ROOM 3D129 WASHINGTON DC 20301-3080
1	CODE 1142PS OFFICE OF NAVAL RESEARCH 800 N QUINCY STREET ARLINGTON VA 22217-5000
1	WALTER REED ARMY INST OF RSCH ATTN SGRD UWI C (COL REDMOND) WASHINGTON DC 20307-5100

NO. OF COPIES	ORGANIZATION
1	COMMANDER US ARMY RESEARCH INSTITUTE ATTN PERI ZT (DR E M JOHNSON) 5001 EISENHOWER AVENUE ALEXANDRIA VA 22333-5600
1	DEPUTY COMMANDING GENERAL ATTN EXS (Q) MARINE CORPS RD&A COMMAND QUANTICO VA 22134
1	HEADQUARTERS USATRADOC ATTN ATCD SP FORT MONROE VA 23651
1	COMMANDER US ARMY MATERIEL COMMAND ATTN AMCAM 5001 EISENHOWER AVENUE ALEXANDRIA VA 22333-0001
1	COMMANDER USA OPERATIONAL TEST & EVAL AGENCY ATTN CSTE TSM 4501 FORD AVE ALEXANDRIA VA 22302-1458
1	USA BIOMEDICAL R&D LABORATORY ATTN LIBRARY FORT DETRICK BUILDING 568 FREDERICK MD 21702-5010
1	COMMANDER USA AEROMEDICAL RESEARCH LAB ATTN LIBRARY FORT RUCKER AL 36362-5292
1	AAMRL/HE WRIGHT PATTERSON AFB OH 45433-6573
1	US ARMY TROOP SUPPORT CMD NATICK RD&E CENTER ATTN BEHAVIORAL SCI DIV SSD NATICK MA 01760-5020
1	US ARMY TROOP SUPPORT CMD NATICK RD&E CENTER ATTN TECH LIBRARY (STRNC MIL) NATICK MA 01760-5040

NO. OF COPIES	ORGANIZATION
1	MEDICAL LIBRARY BLDG 148 NAVAL SUBMARINE MEDICAL RSCH LAB BOX 900 SUBMARINE BASE NEW LONDON GROTON CT 06340
1	AFHRL/PRTS BROOKS AFB TX 78235-5601
1	DR JON FALLESEN ARI FIELD UNIT PO BOX 3407 FORT LEAVENWORTH KS 66027-0347
1	ARI FIELD UNIT FORT KNOX BUILDING 2423 PERI IK FORT KNOX KY 40121-5620
1	COMMANDANT USA ARTILLERY & MISSILE SCHOOL ATTN USAAMS TECH LIBRARY FORT SILL OK 73503
1	COMMANDER WHITE SANDS MISSILE RANGE ATTN TECHNICAL LIBRARY WHITE SANDS MISSILE RANGE NM 88002
1	COMMANDER USA TANK-AUTOMOTIVE R&D CENTER ATTN AMSTA TSL (TECH LIBRARY) WARREN MI 48397-5000
1	PEO COMMUNICATIONS ATTN SFAE CM RE FT MONMOUTH NJ 07703-5000
1	COMMANDANT US ARMY ARMOR SCHOOL ATTN ATSB CDS (MR LIPSCOMB) FT KNOX KY 40121-5215
1	CECOM SP & TERRESTRIAL COM DIV ATTN AMSEL RD ST MC M H SOICHER FT MONMOUTH NJ 07703-5203
1	PRIN DPTY FOR TECH GY HDQ US ARMY MATL CMND ATTN AMCDCG T M FISETTE 5001 EISENHOWER AVE ALEXANDRIA VA 22333-0001
1	PRIN DPTY FOR ACQTN HDQ US ARMY MATL CMND ATTN AMCDCG A D ADAMS 5001 EISENHOWER AVE ALEXANDRIA VA 22333-0001
1	DPTY CG FOR RDE HDQ US ARMY MATL CMND ATTN AMCRD BG BEAUCHAMP 5001 EISENHOWER AVE ALEXANDRIA VA 22333-0001
1	DPTY ASST SCY FOR RSRCH & TECH SARD-TT F MILTON RM 3E479 THE PENTAGON WASHINGTON DC 20310-0103
1	DPTY ASST SCY FOR RSRCH & TECH SARD-TT D CHAIT THE PENTAGON WASHINGTON DC 20310-0103
1	DPTY ASST SCY FOR RSRCH & TECH SARD-TT K KOMINOS THE PENTAGON WASHINGTON DC 20310-0103
1	DPTY ASST SCY FOR RSRCH & TECH SARD-TT B REISMAN THE PENTAGON WASHINGTON DC 20310-0103
1	DPTY ASST SCY FOR RSRCH & TECH SARD-TT T KILLION THE PENTAGON WASHINGTON DC 20310-0103
1	ODCSOPS D SCHMIDT WASHINGTON DC 20310-1001
1	OSD OUSD(A&T)/ODDDR&E(R) J LUPO THE PENTAGON WASHINGTON DC 20301-7100
1	ARL ELECTROMAG GROUP CAMPUS MAIL CODE F0250 A TUCKER UNIVERSITY OF TEXAS AUSTIN TX 78712
1	DUSD SPACE 1E765 J G MCNEFF 3900 DEFENSE PENTAGON WASHINGTON DC 20301-3900

NO. OF COPIES	ORGANIZATION
1	USAASA MOAS-AI W PARRON 9325 GUNSTON RD STE N319 FT BELVOIR VA 22060-5582
1	CECOM PM GPS COL S YOUNG FT MONMOUTH NJ 07703
1	GPS JOINT PROG OFC DIR COL J CLAY 2435 VELA WAY STE 1613 LOS ANGELES AFB CA 90245-5500
1	ELECTRONIC SYSTEMS DIV DIR CECOM RDEC J NIEMELA FT MONMOUTH NJ 07703
3	DARPA L STOTTS J PENNELLA B KASPAR 3701 N FAIRFAX DR ARLINGTON VA 22203-1714
1	SPECIAL ASST TO THE WING CDR 50SW/CCX CAPT P H BERNSTEIN 300 O'MALLEY AVE STE 20 FALCON AFB CO 80912-3020
1	USAF SMC/CED DMA/JPO M ISON 2435 VELA WAY STE 1613 LOS ANGELES AFB CA 90245-5500
1	ARL HRED AVNC FIELD ELEMENT ATTN AMSRL HR MJ (R ARMSTRONG) PO BOX 620716 BLDG 514 FT RUCKER AL 36362-0716
1	ARL HRED MICOM FIELD ELEMENT ATTN AMSRL HR MO (T COOK) BUILDING 5400 ROOM C242 REDSTONE ARSENAL AL 35898-7290
1	ARL HRED USAADASCH FLD ELEMENT ATTN AMSRL HR ME (K REYNOLDS) ATTN ATSA CD 5800 CARTER ROAD FORT BLISS TX 79916-3802

NO. OF COPIES	ORGANIZATION
1	ARL HRED ARDEC FIELD ELEMENT ATTN AMSRL HR MG (R SPINE) BUILDING 333 PICATINNY ARSENAL NJ 07806-5000
1	ARL HRED ARMC FIELD ELEMENT ATTN AMSRL HR MH (J JOHNSON) BLDG 1109B 3RD FLOOR FT KNOX KY 40121-5215
1	ARL HRED CECOM FIELD ELEMENT ATTN AMSRL HR ML (J MARTIN) MYER CENTER RM 3C214 FT MONMOUTH NJ 07703-5630
1	ARL HRED FT BELVOIR FIELD ELEMENT ATTN AMSRL HR MK (P SCHOOL) 10115 GRIDLEY ROAD SUITE 114 FORT BELVOIR VA 22060-5846
1	ARL HRED FT HOOD FIELD ELEMENT ATTN AMSRL HR MV (E SMOOTZ) HQ TEXCOM BLDG 91012 RM 111 FT HOOD TX 76544-5065
2	ARL HRED NATICK FIELD ELEMENT ATTN AMSRL HR MQ (M FLETCHER) ATTN SSCNC A (D SEARS) USASSCOM NRDEC BLDG 3 RM R-140 NATICK MA 01760-5015
1	ARL HRED FT HUACHUCA FLD ELEMENT ATTN AMSRL HR MY (B KNAPP) GREELY HALL (BLDG 61801 RM 2631) FORT HUACHUCA AZ 85613-5000
1	ARL HRED FT LEAVENWORTH FLD ELE ATTN AMSRL HR MP (D UNGVARSKY) TPIO ABCS 415 SHERMAN AVE RM 327 FT LEAVENWORTH KS 66027-1344
1	ARL HRED FLW FIELD ELEMENT ATTN AMSRL HR MZ (A DAVISON) 320 ENGINEER LOOP STE 166 FT LEONARD WOOD MO 65473-8929
1	ARL HRED OPTEC FIELD ELEMENT ATTN AMSRL HR MR (D HEADLEY) PARK CENTER IV RM 1450 4501 FORD AVENUE ALEXANDRIA VA 22302-1458

NO. OF
COPIES ORGANIZATION

1 ARL HRED SC&FG FIELD ELEMENT
ATTN AMSRL HR MS (L BUCKALEW)
SIGNAL TOWERS RM 207
FORT GORDON GA 30905-5233

1 ARL HRED STRICOM FIELD ELEMENT
ATTN AMSRL HR MT (A GALBAVY)
12350 RESEARCH PARKWAY
ORLANDO FL 32826-3276

1 ARL HRED TACOM FIELD ELEMENT
ATTN AMSRL HR MU (M SINGAPORE)
BLDG 200A 2ND FLOOR
WARREN MI 48397-5000

1 ARL HRED USAFAS FIELD ELEMENT
ATTN AMSRL HR MF (L PIERCE)
BLDG 3040 RM 220
FORT SILL OK 73503-5600

1 ARL HRED USAIC FIELD ELEMENT
ATTN AMSRL HR MW (E REDDEN)
BLDG 4 ROOM 332
FT BENNING GA 31905-5400

1 ARL HRED USASOC FIELD ELEMENT
ATTN AMSRL HR MN (F MALKIN)
HQ USASOC BLDG E2929
FORT BRAGG NC 28307-5000

1 US ARMY RSCH DEV STDZN GP-UK
ATTN DR MICHAEL H STRUB
PSC 802 BOX 15
FPO AE 09499-1500

ABERDEEN PROVING GROUND

2 DIRECTOR
US ARMY RESEARCH LABORATORY
ATTN AMSRL CI LP (TECH LIB)
BLDG 305 APG AA

1 LIBRARY
ARL BLDG 459
APG-AA

1 ARL HRED ERDEC FIELD ELEMENT
ATTN AMSRL HR MM (R MCMAHON)
BLDG 459
APG-AA

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE December 1997	3. REPORT TYPE AND DATES COVERED Final

4. TITLE AND SUBTITLE

JANUS Digitization Test Bed: Assessing Performance in a Simulated Battlefield Environment

5. FUNDING NUMBERS

AMS: 622716.H700011
PR: 1L162716AH70
PE: 6.27.16

6. AUTHOR(S)

Hawley, J.K.; Flanagan, J.P.; Wilkinson, T.D. (Hughes Training, Inc.); Pierce, L.G. (ARL)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
Human Research & Engineering Directorate
Fort Sill, OK 73503-5600

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
Human Research & Engineering Directorate
Aberdeen Proving Ground, MD 21005-5425

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ARL-TR-1437

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

An evaluation of the combat information processor (CIP) was conducted by the Fort Sill Field Element, Human Research and Engineering Directorate (HRED) of the U.S. Army Research Laboratory (ARL), and Hughes Training, Inc., (HTI) in support of the Depth and Simultaneous Attack Battle Lab at Fort Sill, Oklahoma. The Training and Doctrine Command, through the Concept Evaluation Program, funded the research. Our objectives were to (a) evaluate the CIP in a simulation test bed to provide user feedback to the system developers and (b) develop guidance for using simulation test beds such as the Janus Battle Simulation Center (JBSC) in operational test and evaluation, particularly early user tests. More specifically, the CIP user test was conducted to (a) evaluate the battle command utility of the CIP software, (b) provide formative feedback to the CIP system developer, and (c) recommend technological and procedural enhancements to improve information management during the conduct of battle exercises within the JBSC. It was hypothesized that by establishing a test bed in conjunction with a multi-user simulation center, developmental or prototype hardware and software products could be evaluated in relationship to fielded systems, allowing system developers to receive early user feedback regarding the suitability of the product for user application. This process would result in the enhancement of the acquisition cycle by providing a high fidelity, low cost environment to support early and frequent user testing. The first phase of the test, a functional review of the CIP software, revealed a number of deficiencies that would limit the usability of the CIP by a battle command staff, whether in a simulation test bed or in the field. The second phase of the test involved a limited user evaluation during two Janus battle simulations. A number of deficiencies were identified in the use of the CIP in an operational environment, especially in the use of software control measures. Deficiencies and our observations are included in the report, along with recommended solutions to aid in the design of the next generation software. The current research program was initiated to address the use of simulation test beds to support the acquisition of battle command systems. Although the current simulation test bed was adequate for conducting a limited user evaluation, it was suggested that future simulations-based testing be developed using distributed interactive simulation (DIS) technology. The use of a DIS environment will allow for immersion of the test systems and operator into the synthetic environment to increase the realism of the training and ensure the validity of the user assessment.

14. SUBJECT TERMS

battle staff performance	testing
command and control	training

15. NUMBER OF PAGES

60

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102