

STAR: A unique embedded performance assessment technique

CHARLES GRAHAM, MARY R. COOK, HARVEY D. COHEN, JAMES W. PHELPS, and
MARY M. GERKOVICH

Midwest Research Institute, Kansas City, Missouri

STAR (The Strategic and Tactical Assessment Record) is a versatile research tool designed to evaluate the effects of sustained performance and other stressors on integrated, complex cognitive functioning. This completely automated task is presented in the form of a highly motivating computer game. However, unlike typical computer games, successful performance of STAR depends on an individual's ability to rapidly and accurately assess risk/benefit ratios in a variety of situations, and on the skilled use of an array of multipurpose control systems. STAR is unique in that approximately 80 performance measures are unobtrusively embedded in the operations required to "play the game"; no obvious performance assessment interferes with task presentation. STAR provides multiple measures of psychomotor function, attention, memory, information processing, decision making, risk-taking behavior, subjective state, errors, and error paths. In this report, we describe STAR, and summarize two preliminary experiments designed to separately evaluate the effects of task difficulty and task stress level on complex performance.

STAR (The Strategic and Tactical Assessment Record) grew out of our attempts to better understand the effects of sustained operations on human performance capabilities (Graham, 1983). We were specifically interested in studying (1) the ability to assimilate high rates of multisource, variable-priority information; (2) the ability to integrate and use this information on a real-time basis; and (3) the ability to accurately assess risks and make complex decisions under the pressure of time, ambiguity, and shifting priorities.

There are two basic approaches to studying how a stressor, such as sustained operations, affects the underlying processes involved in these cognitive activities. The first is to study each function separately, as if it were an isolated unit of behavior, and make inferences about overall complex performance from the segmental information gained. Previous research has typically followed this approach, and much valuable information has been generated concerning the components of complex performance (e.g., Fleishman, 1972; Graham et al., 1984; Shingledecker, 1984; Sternberg, 1975; Thorne, Genser, Sing, & Hegge, 1983).

Efficient performance on complex tasks, however, often appears to depend more on the interactive, integrated, and simultaneous use of many abilities than on simple additions or combinations of separate abilities (Carroll, 1974). Thus, another approach is to attempt to study the "whole"

person functioning as an integrated unit. The use of simulations (Fischetti & Truxal, 1985) and synthetic work methodologies (Alluisi, 1969; Chiles, 1982; Chiles, Alluisi, & Adams, 1968) fit into this category.

Although complete simulation studies focus on integrated human function, their utility can be questioned, because the measures obtained often depend upon the subjective judgment of participants and observers. The synthetic work approach typically provides multiple measures of performance, allows strict experimental control, and is relatively cost effective. However, due to the nature of the tasks performed, it often is difficult to sustain high levels of motivation over long periods of time.

A promising variant of these two approaches to the measurement of complex performance is to combine aspects of synthetic work methodology with embedded performance assessment techniques. This combined approach would allow individuals to function in their normal integrated fashion without interference, and would provide unobtrusive measures of task performance. Ideally, this combined approach also should allow the detailed analysis of human error, the tracking of error paths, and, ultimately, the development of testable research models of integrated human performance. The purpose of this report is to describe our preliminary work in developing STAR as a viable means of implementing this integrated approach to the assessment of complex human performance.

DESCRIPTION OF STAR

STAR is a computerized, individual, cognitive performance task patterned after and based upon a variety of widely available computer games. The major difference between STAR and existing computer games is that STAR

This research was supported by USAMRDC Contract DAMD17-80-C-0075. Findings are not to be construed as reflecting an official Department of the Army position. We wish to thank Frederick W. Hegge, Walter Reed Army Institute of Research, Division of Neuropsychiatry, for his helpful comments and contributions during the course of the project. Please send requests for reprints to: C. Graham, Midwest Research Institute, 425 Volker Blvd., Kansas City, MO 64110.

includes a comprehensive set of operationally defined measurement parameters relevant both to specific subareas of human function and to complex integrated functions. Many of the parameters are variants of common laboratory measurements. All measurement procedures are unobtrusively embedded within the operations required to "play" the computer game.

The use of embedded performance assessment technology has two important advantages. First, the unobtrusiveness of the measurement procedure reduces interference effects. Second, the technique allows the measurement of traditional performance variables (e.g., reaction time, memory), not as the typical isolated units of behavior, but rather as components involved in the integrated performance of a complex task of direct interest to the subject.

STAR was designed specifically to have the strong motivational properties necessary in studies of sustained operations. It is a game of skill, not luck. Successful performance depends on the rapid and effective use of cognitive and psychomotor skills, and the individual is allowed as much freedom to act as is feasible within the constraints of measurement. Finally, STAR was designed to be an experimental tool. Beneath its facade of apparently random situations and unlimited freedom to act, it is a highly determined and controlled experimental task, with optimal strategies against which actual performance can be compared.

Task Scenario

STAR is set in the context of a futuristic war. The Federation is being overwhelmed by an advanced and particularly nasty alien force (Xenoids). Fortunately, an ultrasophisticated galactic battle cruiser, the Venture, has been developed. Only a few of these small, highly mobile, individually operated cruisers are available. A few candidates who possess that rare blend of skill and daring have been notified to report for command training. The subject is among those selected.

Due to the grave situation facing the Federation, candidates are sent on missions after receiving minimal basic training. Each candidate is supplied a new Venture cruiser, and is beamed down from the orbiting training command cruiser to a hostile galaxy to fight alone against the superior alien force. Due to the dangers involved, the command cruiser can remain in orbit for only a limited time. Thus, the mission must be completed quickly and efficiently.

The STAR Task

The subject assumes the role of a Federation Venture captain, and is sent repeatedly on missions that vary in difficulty and duration. The subject's task is to locate and destroy a specified number of alien cruisers within the mission time limit, and in the most energy-efficient manner possible.

The successful accomplishment of a mission is dependent on the knowledge and skills of the subject in using Venture's on-board control systems: short- and long-range scanners for use in locating the Xenoids and navigating through the galaxy; a scan history display for use in strategic planning; two types of offensive weapons (phasers and photon torpedoes); main defensive energy shields; quick-response energy shields; navigations systems; manual docking system; and an onboard command and control system that (1) provides damage control capabilities and information on the status of various system energy levels, (2) allows the reallocation of energy resources from onboard reserves to various systems, and (3) provides warnings and messages from the cruiser's automated sensor system. Figure 1 shows the computer display seen by the subject.

Since all operations of the Venture require energy and subtract time from the allowable mission duration, the subject must develop an efficient search strategy and continuously balance mission goals against energy, armament supplies, and time. Supplies can be replenished by docking at the command cruiser. The docking procedure in-

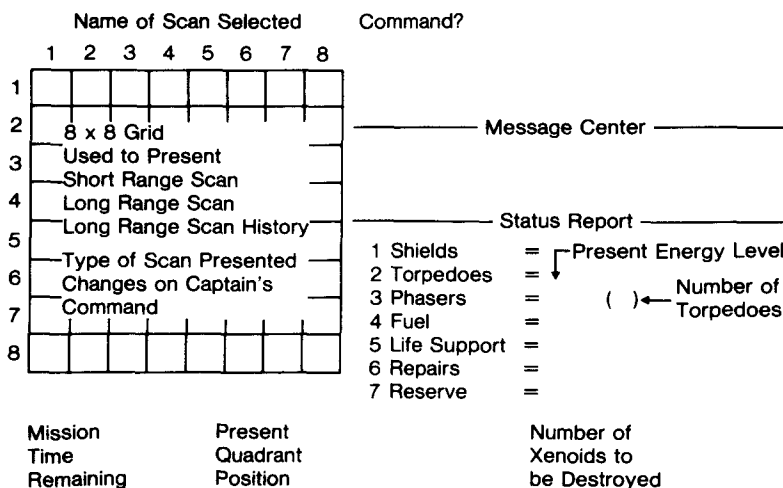


Figure 1. Computer display format for STAR.

corporates a standardized tracking task and requires a mission status report that provides measures of subjective state and memory.

The Captain faces various hazards during the mission. For example, he or she encounters variable numbers of more sophisticated enemy cruisers. Phantom Xenoids, which are not detectable by Venture scanners, appear suddenly and attack without warning. Unless immediate defensive action is taken (reaction time task), Venture can be damaged or destroyed. Super Xenoids are detected by Venture's scanners, but are camouflaged to appear as typical alien cruisers. On entry into a quadrant containing a Super Xenoid, the enemy begins to drain the life support system of the Venture at a rapid rate (internal task stressor). The subject must then make a number of tactical decisions, while simultaneously maintaining close watch on life-support energy levels (multiple measures of information processing, decision making, and risk-taking behavior).

The subject begins each mission by piloting a shuttlecraft to the command cruiser (manual tracking task). He or she makes an initial status report (baseline memory and subjective measures), is assigned a Venture, and enters the galaxy at quadrant 1,1 (the galaxy is divided into 64 quadrants arranged as an 8×8 grid, and each quadrant is further divided into 64 sectors). The location of Xenoids and stars is unknown to the subject, who must search the galaxy to find and destroy the enemy (multiple measures of information processing and decision making).

The Venture can be destroyed if an enemy attack overwhelms the energy level in the defensive shields, if the life support system is drained, or if faulty navigation results in a collision with a star (measures of operational memory, perception, and risk-taking). If the Venture is damaged, the subject must decide whether to effect repairs using onboard supplies, or to return to the command cruiser for repairs and replenishment of supplies (risk/benefit measures).

Prior to any movement of the Venture, the subject must use the terminal to enter into the mission log the purpose of the movement (search, attack, maneuver, evade) and the desired destination (e.g., quadrant 4,6). This allows comparisons to be made of intentions and executions, as well as of the timing and sequencing of information processing and decision making variables. At the beginning and end of each mission and during any within-mission docking procedures, the subject must make a status report to the training commander. These and other features of the task allow analysis of errors and error paths, and provide subjective ratings of levels of fatigue, workload, confidence, efficiency, and stress. The primary measures of performance and decision adequacy are derived from the timing and sequencing of interrogatory, response, and action commands issued by the subject during the mission.

Well-practiced subjects can complete a mission requiring the destruction of 20 Xenoids within 10-15 min, and

missions can be presented in selected orders of difficulty and stress levels (see TRAINING section).

Task Control and Operation

STAR is controlled and operated using a set of five computer programs. The GALAXY program is used by the experimenter prior to a study to set up the parameters of a mission, or of a series of missions. GALAXY determines the distribution and concentration of enemy forces, the difficulty level of the mission, mission time, and the resources available to the subject. The STAR program presents the mission. It requires the experimenter to enter initial subject- and file-identification information. Subsequently, STAR interacts only with the subject. Raw data are collected on the timing and sequence of commands issued by the subject, and are stored in categories related to particular mission situations.

SHORT is an optional program that is automatically called after each mission when performance feedback is desired. This program presents a mission debriefing report to the subject. It presents summary information on selected performance variables in various command areas, and indicates areas where improvement is needed. The report also tells the subject whether his or her performance during the mission will result in promotion or demotion, and presents the criterion to be met to achieve the next higher level of command rank. Subjects begin at the neophyte rank of Venture Captain (Level 10), and strive to achieve and stay at the rank of Venture Captain (Level 1).

The OUTPUT program is used only by the experimenter. It reduces the raw performance data collected by STAR during the mission into the measurement variables, or combinations of measurement variables, to be used for analysis purposes. Data can be stored on disk or tape, printed, or displayed on the terminal. The program provides approximately 80 measures divided into eight major performance categories.

The final program, PATH, is an automated visual-motor tracking/teaching program. It is used only during training to allow subjects to reach the tracking task criterion set in STAR.

Operating Environment

The STAR programs are configured to function on DEC PDP/11 computers. The operating system is DEC RT-11 F/B, version 4, and the programs are written in DEC FORTRAN IV version 2.5. Hardware peripherals include a programmable real-time clock, and flexible or hard disk. DECSCOPE VT-52 or VT-100 series terminals using VT-52 cursor control codes (4800 baud rate) are required. The programs require the use of FORTRAN library calls, including loading and inspection of addresses, and timed interrupt routines controlled by a programmable clock.

PERFORMANCE MEASURES

STAR was designed to examine higher order functioning under sustained performance conditions, as well as

Table 1
Selected Measures Obtained During Each
Mission of the STAR Task

Overall Performance Efficiency
Mean time and energy per Xenoid; mean time per command; number of ventures lost, Xenoid destroyed.
Psychomotor Skills
Tracking Task: % time-on-target; root mean square and mean absolute error. Reaction Time (RT) Task: mean RT; % detection; number errors.
Perceptual Accuracy and Speed
% Torpedos on target; mean absolute course error; % correct quadrant entries; mean time per course calculation.
Memory Function
Recall Task: Mean RT; % correct.
Recent Memory Task: mean RT; % correct; pre- and post-mission RT.
Long-Term Memory Task: mean RT; % correct.
Operational Memory: % commands issued without adequate resources; number of redundant scans.
Information Processing
Accuracy: mean phaser payload calculation error; % successful phaser attacks.
Duration: mean payload calculation time (1, 2, and 3 Xenoids); payload calculation time (Super Xenoid present vs. absent).
Efficiency: % commands and energy used for attack.
Information Resource Usage
Amount: % long-range scans (LRS), short-range scans (SRS), history scans (LRSH); % LRS+LRSH; Number LRS; Number LRSH.
Frequency: mean time LRS to LRS, LRSH to LRSH.
Duration: mean viewing time per LRS and LRSH.
Efficiency: % redundancy, noncontiguity, and new information per LRS; number commands and fuel units for search.
Decision Making
Breakdown: number commands per mission; % in eight categories.
Efficiency: mean time per command; search adequacy as a function of search strategy (algorithm); number critical search decision nodes per mission; number maneuvers during attack; % correct weapons choice (algorithm); number of within-mission dockings and time to first docking.
Risk-Taking Behavior
Mean shield level per enemy quadrant and per weapon command; % weapons commands with inadequate shield levels; mean difference actual versus minimum shield level required during attack; mean life support level at resupply; mean energy allocated to life-support system; onboard energy level at each docking; mean shield level at docking.
Subjective Ratings
Fatigue; confidence; efficiency; stress; workload (1-10 scales).

to assess the effects of various types of environmental stressors. A number of measurement parameters are embedded in the task. Some measures no doubt will prove to be more useful under specific conditions than others. Table 1 presents a listing of the major measures currently obtained during each mission.

TRAINING

The goal of the training protocol is to enable subjects to achieve a high, stable level of performance. Performance levels are operationalized in terms of ranks. Subjects begin at the rank of Venture captain, Level 10, and strive to achieve Level 1 over the course of training. Ad-

vancement is based on progressively more difficult mission time and energy usage criteria. For example, the first promotion is awarded simply for the ability to complete a mission successfully (i.e., destroy all enemy cruisers within the mission time limit, and not lose the Venture in the process). In contrast, Level 1 performance requires the ability to destroy all enemy cruisers at a mean rate below 45 sec/Xenoid and a mean energy expenditure below 225 units/Xenoid. Subjects who achieve Level 1 status can typically complete a 30-min maximum duration mission in approximately 10 min at very high levels of performance efficiency.

The training protocol is composed of a sequence of three types of training sessions. The first session is used to introduce multiple subjects to the task. A training manual presents the scenario and explains the operation of each of Venture's systems. Subjects view a demonstration mission, and then work paper and pencil exercise problems concerned with various command activities. The next session involves one-on-one coaching of the subject during practice missions. During the final sessions, the subject performs missions alone, with the experimenter monitoring mission performance from another area of the laboratory.

Twenty-two male and female volunteers were trained using the above procedures. These subjects required an average of 22 missions to reach Level 1 status. The training time involved was approximately 5 h. However, not all subjects could reach Level 1. We found that if a subject could not achieve the first promotion within the first 8 training missions, he or she could not achieve Level 1 status in the total training time allowed. Figures 2 and 3 show group skill acquisition curves for two of the performance measures. Other measures showed similar acquisition curves.

EXPERIMENT 1: EFFECTS OF TASK DIFFICULTY ON STAR PERFORMANCE

For studies of sustained operations, it is valuable to be able to evaluate the effects of variations in workload. This is often accomplished by changing task difficulty levels. Consequently, the purpose of Experiment 1 was to first identify those parameters that controlled task difficulty level, and then evaluate the impact of parameter variation on task performance.

Based on our experience with STAR, we hypothesized that task difficulty was controlled primarily by two factors: the density of stars encountered in the galaxy during a mission, and the dispersion of enemy units within the galaxy. Density was expected to have the greatest impact on performance variables related to navigation, and enemy dispersion was expected to primarily affect variables related to search strategy and information seeking activities.

Three levels of star density were operationally defined by multiplying the density used in the training galaxies by 1, 2, or 3. Enemy dispersion scores were calculated for 30 randomly generated galaxies, and three levels of

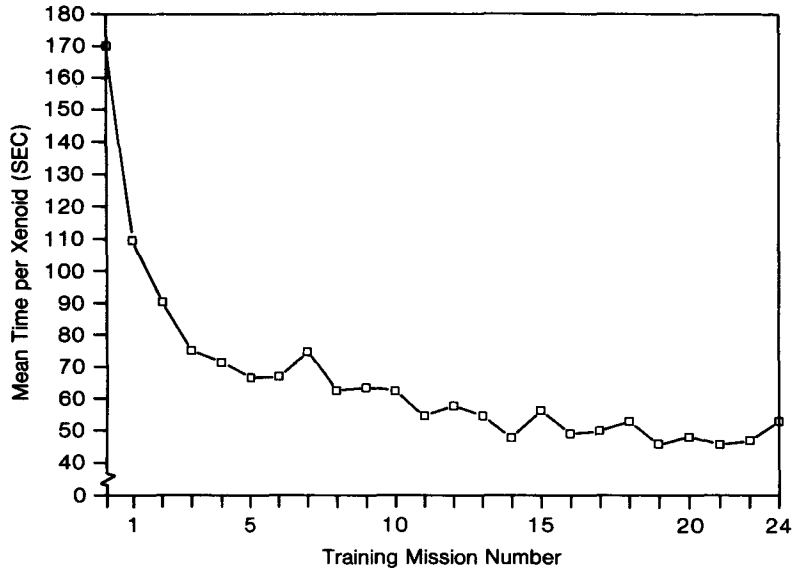


Figure 2. Changes in mean time per Xenoid destroyed as a function of training.

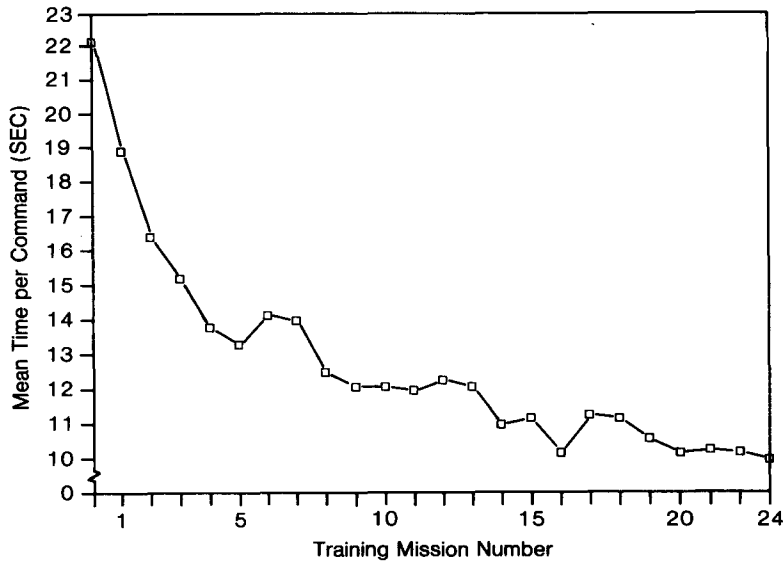


Figure 3. Changes in mean time per command as a function of training.

dispersion were operationally defined. Thus, mission difficulty ranged along a continuum from “easy” (few stars and enemy units in close proximity) to “hard” (triple star density and widely dispersed enemy units). Nine galaxies were selected to provide complete factorial testing of the contribution of star density and enemy dispersion to overall task difficulty, and to examine the particular performance variables affected by changes in these factors.

Procedures

After providing informed consent, 6 male and 6 female subjects (mean age=23 years; range=21-29 years) were

paid to participate in the three, 2-h sessions required for the study. These subjects were recruited from the pool of subjects previously trained on STAR. All subjects had achieved Venture captain Level 1 performance status during training, and each was able to meet United States Army enlistment standards.

For a given subject, each session was conducted at the same time of day to control for circadian effects, and each began with a warm-up mission. In the first session, subjects performed a set of four training missions to assure that they had retained Level 1 performance skills. Over the next two sessions, the nine test missions were performed (five missions in Session 2, and four missions in

Session 3). Each test mission had a set maximum duration of 30 min; however, for these highly practiced subjects, the mean mission duration actually ranged from 10 min for easy missions to 17 min for difficult missions. After each mission, subjects rated their levels of subjective fatigue and workload.

Analysis

Each performance variable was submitted to $2 \times 3 \times 3$ (sex \times star density \times enemy dispersion) ANOVA with repeated measures on the last two variables. The Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was used to correct for inflated degrees of freedom due to correlations among repeated measures. Post hoc tests employed Bonferroni's *t* test (Neter & Wasserman, 1974). Alpha was set at $p < .05$ for all tests. No significant three-way interactions were observed.

Results and Discussion

The effects of star density and Xenoid dispersion on subjective measures and overall performance measures were examined first. Analysis of the fine structure of mission performance was then used to provide an in-depth picture of the effects of increasing difficulty, and to examine the emergence of common error patterns. Variables related to navigation, search strategy, offensive and defensive strategy, resource allocation, information processing, and memory were included in these analyses.

Subjective measures. As star density increased, ratings of workload [$F(1,11) = 7.37, p < .05$] and stress [$F(1,11) = 7.45, p < .05$] also increased, while ratings of confidence [$F(1,11) = 11.60, p < .01$] and efficiency [$F(1,11) = 6.65, p < .05$] decreased. In contrast, Xenoid dispersion had no effect on any of the subjective measures obtained. Thus, star density is the major factor influencing measures of subjective workload.

Overall performance measures. Three overall performance measures are obtained from STAR: time per Xe-

noid (mission time/number of Xenoids destroyed); energy per Xenoid (total mission energy/number of Xenoids destroyed); and time per command (total mission time/number of commands issued). Time per Xenoid increased as star density increased [$F(1,11) = 54.18, p < .01$], and as enemy dispersion increased [$F(1,11) = 12.46, p < .01$]. Similarly, energy per Xenoid increased as a function of both factors {star density [$F(1,11) = 17.90, p < .01$]; Xenoid dispersion [$F(1,11) = 5.11, p < .05$]}. Overall time per command increased as star density increased [$F(1,11) = 17.18, p < .01$], but this effect was not significant for Xenoid dispersion.

Detailed performance measures. It is not possible in the space available to describe in full the results of analyses of the fine structure of STAR performance. Findings will be described briefly, using several examples to demonstrate the richness of the information that can be obtained from STAR.

Theoretically, star density should have a direct effect on variables related to navigation, and Xenoid dispersion should affect primarily those navigation variables related to searching for the enemy. As shown in Table 2, these hypotheses were confirmed; all effects were in the expected direction.

In addition to navigation variables related to searching for the enemy, STAR provides several measures of information seeking, information redundancy, processing time, and resource allocation which are also related to search strategy. The expected results were obtained for this class of variables. Star density significantly affected only 6 of the 16 variables examined, and Xenoid dispersion affected 14. Table 3 summarizes these results. Both star density and Xenoid dispersion increased the amount of information seeking activity, as measured by the number of times information was requested by the subject. Star density also affected the percent of commands used for information seeking, but further analysis indicated that

Table 2
Effects of Task Difficulty on Variables Related to Navigation

Variable	Star Density		Xenoid Dispersion	
	F	p <	F	p <
Percent of all commands that were used for navigation	67.32	.001	2.11	ns
Number of navigation commands	75.07	.001	45.20	.001
Percent of all navigation commands used to:				
Search	1.90	ns	95.41	.001
Attack	77.30	.001	68.56	.001
Maneuver	118.60	.001	5.19	.05
Number of navigation commands for:				
Search	12.76	.01	111.46	.001
Attack	0.70	ns	8.34	.05
Maneuver	106.27	.001	4.28	ns
Evade	1.42	ns	2.51	ns
Fuel used in navigation to:				
Search	9.56	.05	156.18	.001
Attack	5.90	.05	3.54	ns
Maneuver	45.38	.001	3.50	ns
Evade	0.88	ns	1.89	ns

Table 3
Effects of Task Difficulty on Variables Related to Search Strategy

Variable	Star Density		Xenoid Dispersion	
	F	p <	F	p <
Percent of total commands used for:				
Short-range scans (SRS)	6.41	.05	1.26	ns
Long-range scans (LRS)	24.20	.001	0.38	ns
Long-range scan histories (LRSH)	0.07	ns	24.44	.001
LRS + LRSH	6.84	.05	37.31	.001
Number of:				
SRS	18.33	.01	61.59	.001
LRS	12.96	.01	43.28	.001
LRSH	5.42	.05	41.06	.001
Number of LRS with no new information	2.78	ns	21.69	.001
Number of LRSH with no new information	1.62	ns	4.98	.05
Time between LRSH commands	2.10	ns	6.44	.05
Number of LRSH/Number of LRS × 100	2.56	ns	4.73	.05
Percent resource commands, fuel	0.74	ns	20.81	.005
Energy units allocated to fuel	2.88	ns	52.74	.001
Percent energy allocated, fuel	0.35	ns	29.04	.001
Percent quadrants of new information per LRS	2.17	ns	10.93	.01
Mean time to extract information per LRS	0.83	ns	12.50	.01

this was due to the general increase in the number of commands required when star density increases.

Offensive strategy variables were examined next. As star density increased, the time from entering a quadrant to destroying all enemy increased [$F(1,11) = 7.50$, $p < .05$], regardless of the type of enemy configuration encountered. Analysis of other offensive strategy variables helped to explain this change in performance. During attack sequences the number of moves to a better battle position increased, reflecting the tactical advantages obtained by "hiding" behind stars while attacking the enemy. Star density also affected the choice of weapon systems. As density increased, torpedo use decreased and phaser use increased, consistent with the advantages of using the "phaser" system when the enemy is hiding behind stars. This interpretation is supported by the finding that star density did not affect measures of correct weapon choice.

Xenoid dispersion was not expected to affect offensive strategy, and this was confirmed. Dispersion is important only when attempting to find the enemy; once the subject enters a hostile quadrant, the dispersion of enemy over other quadrants is irrelevant.

The number of commands used to reallocate resources increased with increases in both star density [$F(1,11) = 6.97$, $p < .05$] and Xenoid dispersion [$F(1,11) = 13.01$, $p < .01$]. The amount of energy reallocated was greater for double and triple star densities than for single densities [$F(1,11) = 6.86$, $p < .05$], and increased as Xenoid dispersion increased [$F(1,11) = 10.81$, $p < .01$].

The calculation of torpedo courses, navigation courses, and phaser payloads provide excellent measures of information processing. The percent of torpedoes on target declined as star density increased [$F(1,11) = 4.60$, $p < .05$], as did the accuracy of phaser payload calculations

[$F(1,11) = 9.73$, $p < .05$]. Navigational course calculations were unaffected by star density, and none of the information processing variables was affected by Xenoid dispersion.

Errors of memory during command operations increased as Xenoid dispersion increased. During the docking phases of the missions, however, neither recent nor long-term memory was affected by the independent variables.

Although men and women did not differ on major performance variables, there were striking differences in the processes each sex followed in performing the task. For example, women made greater use of information sources, attended more to resupplying their cruisers, issued more commands, and performed more navigations than did men. In contrast, men used a greater proportion of their commands for attack purposes, and attended less to maintaining their defensive supplies.

The goal of this initial study was to establish three levels of task difficulty, and to test specific hypotheses about the effects of variations in star density and enemy dispersion on mission performance variables. The hypothesized effects were obtained. Moreover, the findings were robust, internally consistent, and readily interpretable. These characteristics of the data are highly encouraging, particularly in light of the complexity of the task performed, the relatively small number of subjects tested, and the large number of variables assessed.

EXPERIMENT 2: EFFECTS OF TIME AND RESOURCE LIMITATIONS ON STAR PERFORMANCE

The usefulness of STAR would be enhanced if variations in task parameters could be used to produce a stress

or crisis condition. The task then could be used to evaluate the impact of sustained performance demands, environmental stressors, and so forth on cognitive function under both crisis and noncrisis conditions. The second experiment using STAR addressed this issue.

Our analysis of the training data and the data from the difficulty study suggested that two major task parameters might be used to induce a crisis: overall mission time allowed and the amount of resources available on-board the Venture to complete the mission.

Two difficulty levels were employed to examine interactions between crisis conditions and task difficulty. The easy level was defined as a galaxy with standard star density and clustered enemy. The difficult level was defined as a galaxy with triple star density and widely dispersed enemy. Four missions were performed at each level of difficulty.

Data from the previous study were used to define the time and energy restrictions employed to induce crisis conditions. Both were set at values equal to the mean plus one standard deviation. If these values had been in effect in the previous study, 12% of the missions would have been unsuccessful. Of the four missions performed at each difficulty level, one served as a control (no new limitations on either time or energy), one had reduced mission time, one had restricted energy resources, and one restricted both the time and the energy allowed.

Procedures

The 12 volunteers who participated in the previous study served as subjects. Each was informed that "various changes have been made in STAR that might make the game more interesting and challenging," and that we wished to evaluate the impact of these changes on performance. The inclusion of time and energy restrictions made it necessary to add a surrender command to STAR; if a subject ran out of on-board energy by allocating supplies incorrectly, and could not dock to obtain additional supplies, the only option would be to surrender. The operation of this command was explained to subjects at the start of the study.

Eight test missions were performed over two, 2-h sessions. Sessions were conducted at the same time of day for each individual, and each session began with a practice mission. The eight missions were presented in counterbalanced order across subjects.

Analysis

Data were submitted to a $2 \times 2 \times 2$ ANOVA for repeated measures (difficulty \times energy \times time). The Greenhouse-Geisser technique was used to correct for inflated degrees of freedom due to repeated measures. Bonferroni's *t* was used for post hoc comparisons, and Student's *t* was used for planned comparisons between control mission and missions involving restrictions on both time and energy.

Results and Discussion

Subjective effects. If the experimental paradigm was successful in producing a crisis situation, we would ex-

pect higher subjective rating of stress and workload for missions with time and energy limitations than for control missions. This was confirmed for both stress [$t(1,11) = 5.86, p < .001$] and workload [$t(1,11) = 6.01, p < .001$]. No other significant subjective effects were found.

Overall performance measures. Manipulation of time and energy resources had marked consequences in terms of the number of enemy destroyed and the number of Venture cruisers lost. In the previous experiment, these highly practiced subjects destroyed all enemy cruisers on all missions. Under crisis conditions, they failed to destroy 32 enemy cruisers (1 in time-restricted missions, 2 in energy-restricted missions, and 29 in missions with limitations on both time and energy). They also lost their own cruisers on approximately one-third of the missions conducted.

Despite these marked changes in mission success, no main effects for energy restriction or time restriction were found for the three major overall performance variables (time/Xenoid, energy/Xenoid, time/command). Planned comparisons between control galaxies and galaxies with combined time and energy restrictions helped to explain these findings. Contrary to our original hypothesis, effects of combined time and energy restrictions on overall performance were greater when subjects performed "easy" missions than when they performed "hard" missions. In the easy missions, significantly more energy/Xenoid [$t(1,11) = 1.94, 1\text{-tail}, p < .05$] and time/Xenoid [$t(1,11) = 2.00, 1\text{-tail}, p < .05$] were expended when both energy and time were restricted than when no restrictions were in effect. No such difference was found for difficult missions. Further examination of the unsuccessful missions discussed above revealed the same pattern.

Detailed performance measures. Analysis of the fine structure of STAR performance suggested that the overall decrement described above was due to an increase in risk-taking behavior under conditions of restricted time and energy, particularly during easy missions. For example, Figure 4 shows one of these measures of risk taking: the percent of weapon commands issued with inadequate shield protection.

Time and energy restrictions also affected information seeking and processing variables. Unlike risk-taking behavior, which was greater in easy galaxies, changes in information seeking and processing were significant only in the hard galaxies. Under combined time and energy restrictions during difficult missions, subjects increased their ability to obtain new information efficiently, but the strategies they selected to use this information were less effective. For example, subjects under these conditions were able to obtain a greater percent of new information from each scan request [$t(1,11) = 2.89, p < .02$], and the redundancy of information seeking decreased [$t(1,11) = 4.59, p < .001$].

However, despite having gained more information, subjects used a less efficient search strategy under restricted conditions compared to control conditions. The contiguity of the search patterns generated was less during restricted

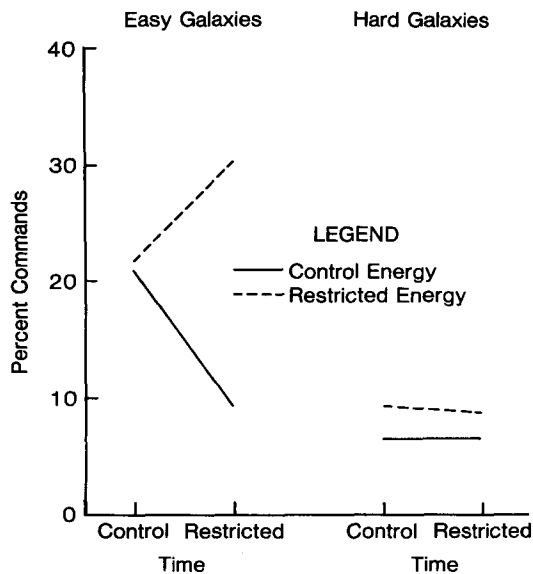


Figure 4. Effects of task stressors on percent weapons commands issued without adequate shield protection.

missions [$t(1,11) = 4.15, p < .01$]; operationally, this resulted in subjects leaving "holes" which had to be searched later.

Restrictions of time and energy had little effect on psychomotor task performance, memory, or the ability to accurately perform arithmetic operations. However, the effects of task difficulty on such variables replicated the effects observed in the previous study.

In general, restrictions of time and energy resources affected overall performance measures more than detailed performance measures, while changes in level of difficulty affected detailed performance measures without significantly altering overall performance. The marked decline in mission success (destroying all the enemy without destroying the Venture) appeared to be due to increased risk-taking behavior. Subjects significantly decreased their margin for error; thus, a simple miscalculation made during attack sequences could result in the loss of the Venture.

The finding that restrictions in time and energy had greater effects for easy than for hard missions is particularly interesting. It may be due to the fact that easy and hard missions were easily discriminated by an experienced subject, and that subjects were more motivated to be careful and reduce risks during difficult missions.

The effects noted here were obtained with highly practiced subjects under minimal stress of short duration. The results suggest that STAR might be very sensitive to continuous performance demands or other experimental conditions resulting in increased stress of even longer durations.

DISCUSSION

The primary goal of the research reported here was to create a complex, cognitive performance task patterned

after existing computer games, and incorporating unobtrusive multiple measures of information processing, decision making, and risk-taking behavior. This goal was accomplished. In its current form, STAR measures multiple aspects of an individual's performance, and it does so without interfering with the ongoing performance activity. The fact that the measurement process is embedded in the task itself allows the analysis of integrated, rather than isolated, performance. Thus, one of the major advantages of STAR is its ability to measure and evaluate integrated components of real-time, ongoing behavior without, at the same time, influencing or interfering with the behavior being measured.

The development of STAR constituted a necessary first step in a planned three-phase study of the effects of sustained operations on the efficiency of complex cognitive functioning. Once the development phase was completed, subjects were to perform STAR continuously for 24-48 h, using chained sequences of missions systematically varied for stress and difficulty levels. Our goal was to produce a controlled "hours of boredom, moments of panic" situation, and then to use this situation as a test bed to develop techniques to effectively aid individuals in maintaining cognitive efficiency under such adverse conditions. However, the planned study using STAR under sustained operation conditions has not been conducted as yet.

The experiments briefly reported here provide basic information on task training parameters, as well as the data required to systematically vary task difficulty and stress levels in studies of longer duration. The training time required for subjects to stabilize at the highest performance level is approximately 5 h, a duration that compares favorably with that involved in learning to perform the Multiple Task Performance Battery (Chiles, 1982). While the data on difficulty and stress variations are promising, it is important to realize that these data were collected on only 22 subjects. More subjects need to be evaluated, and more statistical analyses of the measures need to be conducted. Analysis of larger data sets would provide the means to empirically evaluate the adequacy of the measurement variables and constructs currently available in STAR.

STAR was designed as a basic research tool to address certain specific questions. Our experience, however, suggests that both the task context and the particular measures evaluated could be tailored to fit the needs of a wide variety of research questions, especially those related to sustained operations. For future research purposes, the concept of unobtrusive embedded performance assessment may be more valuable than the specifics of any particular task. Our findings support the feasibility of using this concept in future studies of the effects of sustained operations on integrated cognitive function.

REFERENCES

- ALLUISI, E. A. (1969). Sustained performance. In E. A. McD. Bilodeau & I. McD. Bilodeau (Eds.), *Principles of skill acquisition*. New York: Academic Press.

- CARROLL, J. H. (1974). *Psychometric tests as cognitive tasks: A new structure of the intellect*. Princeton, NJ: Educational Testing Service.
- CHILES, W. D. (1982). Workload, task and situational factors as modifiers of complex human performance. In E. A. Alluisi & E. A. Fleishman (Eds.), *Human performance and productivity: Vol. 3, Stress and performance effectiveness* (pp. 11-56). Hillsdale, NJ: Erlbaum.
- CHILES, W. D., ALLUISI, E. A., & ADAMS, O. S. (1968). Work schedules and performance during confinement. *Human Factors*, **10**, 143-196.
- FISCHETTI, M. A., & TRUXAL, C. (1985). Computers simulating the "right stuff." *IEEE Spectrum*, **22**(3), 38-47.
- FLEISHMAN, E. A. (1972). On the relation between abilities, learning, and human performance. *American Psychologist*, **27**, 1017-1032.
- GRAHAM, C. (1983). *Task validation for studies on fragmented sleep and cognitive efficiency under stress*. (US Army Medical Research and Development Command Final Report, Contract No. DAMD17-80-C-0075). Kansas City, MO: Midwest Research Institute. (Defense Technical Information Center No. AD:A130260)
- GRAHAM, C., COOK, M. R., GERKOVICH, M. M., COHEN, H. D., PHELPS, J. W., & FOTOPOULOS, S. S. (1984, June). Oral dose effects of pyridostigmine on human performance. In *Proceedings of the USAMRDS Fourth Annual Chemical Defense Bioscience Review* (pp. 121-130). Aberdeen Proving Ground, MD: US Army Medical Research Institute of Chemical Defense.
- GREENHOUSE, S. N., & GEISSER, S. (1959). Methods in the analysis of profile data. *Psychometrika*, **24**, 95-112.
- NETER, J., & WASSERMAN, W. (1974). *Applied linear statistical models*. Homewood, IL: Irwin.
- SHINGLEDECKER, C. A. (1984). *A task battery for applied human performance assessment research*. (USAF AMRL Technical Report No. TR-84-71). Wright-Patterson, AFB, OH: U.S. Air Force Aerospace Medical Research Laboratory.
- STERNBERG, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, **27**, 1-32.
- THORNE, D., GENSER, S., SING, H., & HEGGE, F. (1983, May). Plumbing human performance limits during 72 hours of high task load. In *Defense Group Proceedings: The human as a limiting element in military systems* [Vol. 1, No. DS/A/DR(83)170]. Toronto: NATO Defense Research Group.