# FOREWORD

This report was prepared by Dr. Earl A. Alluisi and Mr. Thomas J. Hall of the Lockheed-Georgia Company and by Dr. W. Dean Chiles of the Behavioral Sciences Laboratory. The research summarized in this report was conducted by the Human Factors Research Department of the Operations Research Division, Lockheed-Georgia Company, under the direction of Dr. Earl A. Alluisi, Project Director, and the general supervision of Dr. R. B. Levine, Department Manager, with the technical assistance of Thomas J. Hall, George D. Hayes, James N. Howard, Ann C. Kellogg, O. Edmund Martin, Jr., Thomas W. Meighan, Jr., Richard P. Smith, and Herman E. Williams.

The work was supported by the Behavioral Sciences Laboratory, 6570th Aerospace Medical Research Laboratories, Aerospace Medical Division, under Contract No. AF 33(616)-7607, Project No. 1710, "Training Personnel and Psychological Stress Aspects of Bioastronautics," and Task No. 171002, "Performance Effects of Environmental Stress." Dr. W. Dean Chiles, Chief, Environmental Stress Section, Training Research Branch, served as Task Scientist and consultant.

Major Richard T. Cave, Training Research Branch, Dr. Paul D. Hood, HumRRO Leadership Research Unit, and Dr. James T. Ray, Trust Company of Georgia, all assisted greatly in the development of the group-performance tasks and the general plan of research. The AFROTC Detachment Commanders at the Georgia Institute of Technology and at Emory University, Colonel Donald S. Dunlap and Lieutenant Colonel Allen C. Hart, respectively, gave generously of their time and efforts to assist in the selection of the 25 AFROTC students who served as subjects in this study. To each of these the authors wish to acknowledge their indebtedness and to express their gratitude. Finally, the authors wish to thank Dr. Oscar S. Adams, Operations Research Division, for his substantial contributions to every phase of the research.

# ABSTRACT

This study was carried out to determine the test-retest reliability of a battery of six performance tasks. Four of these tasks were designed to assess individual performance, and the remaining two tasks were designed to measure crew or group-dependent performance. In each of these group tasks, successful performance required interactions among crewmembers in the form of exchanges of information, cooperation, and temporal coordination.

Each of 5, 5-man crews was tested for 4 consecutive hours a day on each of 4 days after a 3-day training period. All testing was conducted with the crewmembers seated at work stations in an advanced-system crew compartment mock-up. All of the measures taken with the six tests exhibit satisfactorily high reliability coefficients.

# PUBLICATION REVIEW

This technical documentary report has been reviewed and is approved.

WALTER F. GRETHER
Technical Director
Behavioral Sciences Laboratory

# TABLE OF CONTENTS

# INTRODUCTION

Among the most important operational requirements of both current and planned weapon systems are those relating to crew or small-group proficiency in performing assigned tasks. The manning requirements for aircraft and spacecraft, and for many military installations (e.g., radar picket stations such as the DEW-line and BMEWS sites) place heavy demands on both individual and crew proficiencies if the necessary over-all system capabilities are to be achieved. Crew proficiency is known to contribute in many ways to operational effectiveness, but its contribution to ultimate system performance is still not exactly specifiable. The research reported here was undertaken principally to establish reliabilities of several individual and crew (small-group) performance tasks similar to those found in the complex aerospace systems of today.

It has been concluded on the basis of previous studies (refs. 1, 2, 3, 4, and 5) that highly motivated subjects, who can be found with a minimum of selection, will maintain acceptable performance levels on a highly demanding schedule of work for periods at least as long as 15 days. The most extensive study in the series was one (ref. 4) in which operational personnel (two combat-ready B-52 crews) followed a work-rest schedule that required 4-hours-on-duty and 2-hours-off-duty for a period of 15 days while confined to the small space provided in a simulated advanced-system crew compartment. During this experimentation it was observed (a) that the testing conditions had fewer adverse effects on the performances of those crewmembers who were judged by the experimenters to have the highest levels of motivation, and (b) that 7 of the 11 subjects improved significantly in performing a pattern-discrimination task containing major learning components. It was also found (c) that the majority of the crewmembers in post-test interviews said they could have continued the test for an additional 15 days were it necessary and important, and (d) that in comparison to the data obtained from control subjects the major effects of the 15-day, 4/2 schedule were related to depressions of performances rather than learning efficiencies.

These observations and conclusions were based on the data derived from performance of five tasks, each of which was an "individual" rather than a "group" task in the sense that one individual's performance was independent of any other individual's performance. Because the test battery used in these previous studies consisted entirely of such individual tasks, estimates of group or crew performances were necessarily restricted either to simple summations of individual performances or to qualitative judgments made by the experimenters.

Small-group performances have been investigated directly in the present study. Specifically, two group-performance tasks were incorporated into the previously used basic battery of tests. In each of these two tasks, successful performance required interactions among crewmembers in the form of exchanges of information, cooperation, and coordination. One of the tasks stressed individual proficiency as a principal determiner of the crew "score," whereas the other stressed crew coordination; both required whole-crew participation and interaction for successful performance, and both provided quantitative indices of crew performance. The purpose of the study was to evaluate these newly added tasks and to determine in what ways they might be improved through further modification; it was decided also to analyze several criterion measures of performance with each of these tasks in order to determine which measures could be most efficiently adopted in future studies.

# METHOD

## SUBJECTS

A total of five groups of AFROTC students from two universities in the Atlanta, Georgia, area volunteered to participate in this study as subjects. Each subject was paid on an hourly basis for his 28 hours of duty. The ranking man in each group had been previously designated as the group leader by his AFROTC Detachment Commander. All members within any given group were enrolled at the same school, knew all other members of the group, and had been together in AFROTC training for at least one academic year prior to testing. The subjects had completed a median of 6 quarters of college work, with a range from 3 to 10 quarters. They ranged in age from 19 to 22 years, with a median of 20. The median height was 5 feet 10 1/2 inches, with a range from 5 feet 6 inches to 6 feet 3 1/2 inches; weight ranged from 125 to 200 pounds, with a median of 165.

## TEST FACILITY

Crews were tested in the five-man, advanced-system crew-compartment mock-up described in previous reports (refs. 1, 2, 4, and 5). An artist's view of the compartment in its present configuration is shown in figure 1. The compartment is divided into three sections: (a) a five-station work area, (b) a leisure area, and (c) a sleeping area. Only the work-station and leisure areas were used during this study, and use of the leisure area was limited to use of the toilet facility during the 10-minute "low-performance" periods each hour.

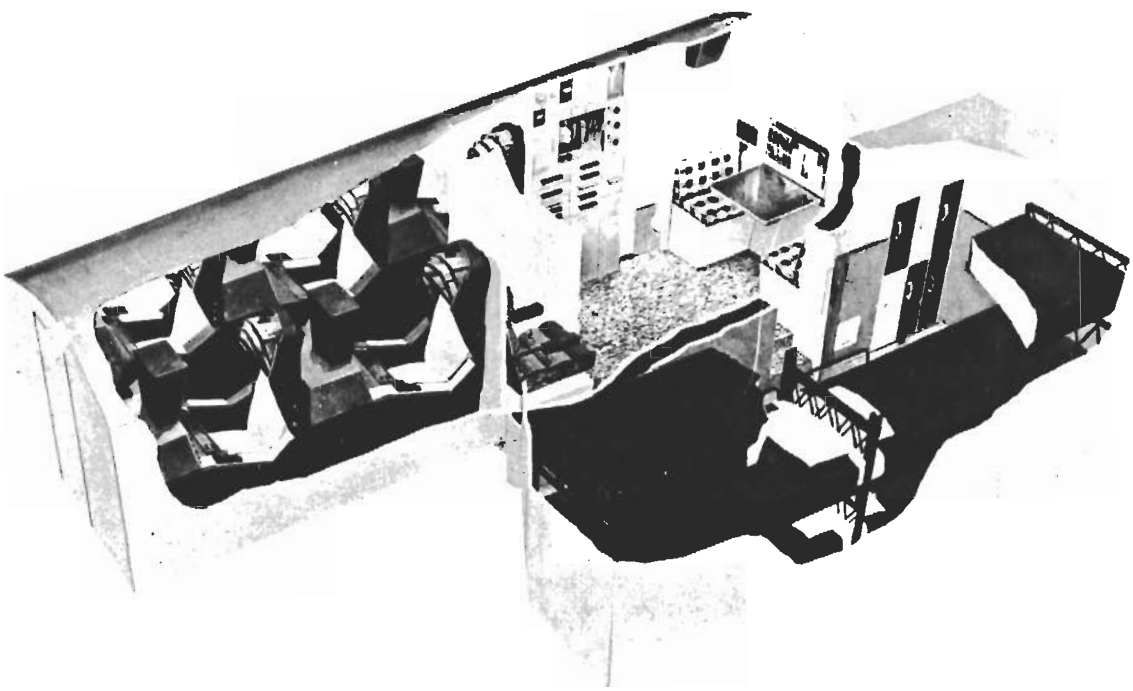

**Figure 1 Cutaway View of the Advanced-System Crew-Compartment Mock-up**

The work-station and leisure areas were adequately and comfortably illuminated at all times. Broad-band noise at a level of approximately 85 decibels (Sound Pressure Level) was presented during all testing periods through a system of small speakers located at different points in the compartment. The noise was presented both for purposes of realistic simulation and to mask all extraneous outside sounds.

## PERFORMANCE MEASURES

A battery of six performance tasks designed to test the psychological functioning of both individuals and small groups was used in this study. The tasks required performance that sampled fundamental abilities such as mental computation, pattern discrimination, monitoring, and vigilance; in addition, the two group-performance tasks required information exchange, cooperation, and coordination among the members of the crew. The tasks were displayed on identical 11-by-28-inch panels (figure 2) mounted at the work stations in the crew compartment. Since the individual tasks have been discussed in detail in previous reports (refs. 1, 2, 4, and 5) only the two group-performance tasks are described here.

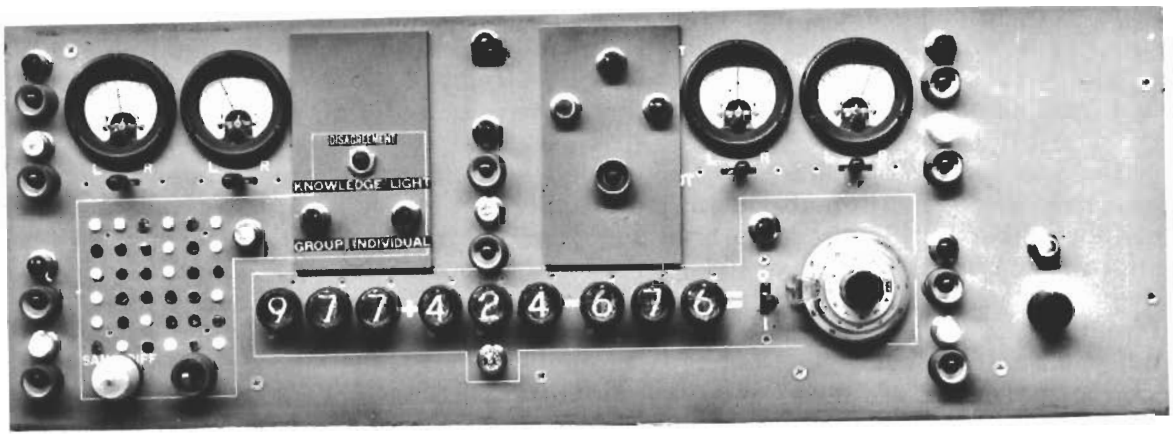


Figure 2 Front View of the Performance-Task Panel Mounted at Each of the Five Work Stations

### Target Identification

The first of the group performance tasks was based on what was previously called the "pattern-discrimination" task (ref. 1, pp. 27-28; ref. 5, pp. 5-6). In the original version of this task, the patterns shown on the 6x6 matrix of lights were identical at all five work stations. Each subject was presented the same pair of patterns on any specific trial, and each was required to decide independently of the other subjects whether the second pattern of the pair was the same as (or different from) the first.

In attempting to modify this task so that it would provide an estimate of group performance (as well as individual performance), it was apparent that a group effort could be obtained merely by demanding that the crew commander report the concensus of individual judgments. However, it was also apparent that if this were done, the probability of a group error would be a function strictly of the probability of there being three or more individuals in error. That is to say, if the task were modified only in this way, then group errors would occur only infrequently as the subjects became proficient at their individual tasks. The pattern-discrimination task obviously had to be modified further so that it would continue to present a challenge to the group even at asymptotic levels of individual performance.

The desired characteristic was achieved by modifying the individual displays so that when a given pair of patterns was routed to all five displays there would be differences from one display to the next in what was actually presented. This modification was accomplished simply--by making a different set of nine lights inoperative at each work station. Thus, as in the original version of the task, each member of a pair of patterns was displayed for 5 seconds; they were separated by a 5-second off-period and were followed by a 15-second off-period. The second member of the pair was rotated 0, 90, 180, or 270 degrees from the orientation of the first. The subject had to make an individual judgment as to whether the second pattern displayed at his work station was the same as (or different from) the first; if the pattern was the same except for rotation by 90, 180, or 270 degrees, he was to report that the patterns were the "same." The subject also had to report his individual judgment to the crew commander, using the intercom system, and the crew commander, in turn, was instructed to submit not only his own individual judgment but also a "group response" based on the majority report of all members of the crew including himself. All members of the crew were informed that if each was correct in his individual judgment and in his report of that judgment to the crew commander, and if the crew commander reported the majority judgment, then the group response would be correct. Thus, it was possible to have as many as five, or as few as zero, correct individual responses of "same" (or "different") on a given trial, depending upon the patterns actually displayed. The correct group response was uniquely determined for each trial by the patterns actually displayed.

The presentation of patterns was controlled with a punched-tape programmer and a tape containing 192 different pairs of patterns divided into four groups of 48 pairs each. Each group of 48 pairs of patterns had been prepared so that there were eight pairs for which correct individual judgments would be unanimous (four pairs in which all five operators should respond "same," and four pairs in which all should respond "different"). Of the remaining 40 pairs of patterns in each group, 20 called for three correct judgments of "same" and two of "different," whereas the other 20 called for two correct judgments of "same" and three of "different." A set of 40 pairs, drawn from one of the four, 48-pair groups, was presented during the 20-minute period of target-identification performance called for during each hour of testing.

On a given presentation, the first member of the pair of patterns was presented for 5 seconds. This was followed by a 5-second off-period, and then the second member of the pair was presented for 5 seconds. Following the display of the second member of the pair of patterns, there was no display for 15 seconds. It was during the last 20 seconds of each 30-second problem that the subject had to indicate whether his individual judgment was "same" or "different." He did this by throwing a 3-position spring-centered toggle switch in the appropriate direction. He also had to report this over the intercom to the crew commander during this interval. On the basis of the reports from the four crewmembers, as well as the individual judgment of his own display, the crew commander announced a group or "crew" decision to the crew over the intercom and activated a "group-response" switch. The group response, like the individual responses, had to be entered within the last 20 seconds of each 30-second problem in order to be scored as correct. Just before the beginning of a new target-identification problem, a blue light marked "group" was illuminated on each panel if the group report submitted by the crew commander had been correct; this provided knowledge of group results to each member of the crew. A second blue light, marked "individual," was illuminated at the appropriate work stations to provide each subject with knowledge of results concerning his individual response. The percentages of both individual and group responses correctly made during each period were used as criteria of performance.

Since the group response was based on the majority of individual responses, it seemed reasonable to expect that a high level of individual performance would be a prerequisite to high levels of group performance. However, in the course of the study it became apparent that such was not the case; instead, the group performance of a crew judged poorest in individual levels of performance seemed to equal or exceed that of a crew judged to be individually better. The distribution of correct same-different splits among the problems used (8/48 or 1/6 being 5-0 or 0-5 splits, and 40/48 or 5/6 being 3-2 or 2-3 splits) deviated from a binomial distribution rather radically. On the assumption that this deviation might have given rise to an atypical relation between the probability of an erroneous group response and the number of individual responses in error, a theoretical analysis of the actual distribution was made. This analysis was made as follows:

Theoretically, if there is no individual in error the group response entered by the crew commander on the basis of the majority judgment cannot be in error. If one individual is in error, the 8/48 problems calling for 5-0 splits in judgments will not be affected and all of these group responses will be correct. However, of the remaining 40/48 problems that call for 3-2 splits in individual judgments, the group response will be in error whenever the erroneous individual judgment should have been on the 3-side of the split; it will not be in error whenever that erroneous individual judgment should have been on the 2-side. Since by chance there is a probability of 3/5 that a single erroneous judgment will fall on the 3-side of a 3-2 split, and a probability of 2/5 that it will fall on the 2-side, there is a total probability of (3/5)(40/48) or 24/48 that the group response will be in error when there is one erroneous individual judgment. The results shown in table 1 were obtained by continuing this analysis to include all possible numbers of individual judgments in error. It is apparent that a non-monotonic relation exists; e.g., the probability of an erroneous group response when only a single subject makes an individual error is greater than when two subjects make errors.

Table 1

Probability of a "Group Error" as a Function of the Number of Erroneous Individual Judgments

| Number of Individual Judgments In Error | Probability of Group Response Being In Error |
|---|---|
| 0 | .00 |
| 1 | .50 |
| 2 | .25 |
| 3 | .75 |
| 4 | .50 |
| 5 | 1.00 |

## Code-Lock Solving

The second group-performance task, the code-lock task, required the crew to discover the proper sequential order for depressing five push-buttons -- one for each crewmember. Three jewel indicator lights (red, amber, and green) and a push-button were placed on each subject's panel for the performance of this task. Illumination of the red light was the signal that an unsolved problem was present. The amber light, which was illuminated during the time in which any subject kept his push-button depressed, served as an indication to all subjects that some member of the crew was responding. The problem would be solved only when each of the five subjects had depressed his push-button at the point in the sequential order that was correct for a given problem. The red light was extinguished when the correct first subject in a sequence depressed his push-button; it remained extinguished until an incorrect response was made, at which time it was re-illuminated (and the programming apparatus automatically reset the problem to the beginning of the sequence). Once the correct first subject had been identified by the group, that subject had to remember to push his button any time the red light came on again (i.e., whenever an error had been made) in order to permit the search for the correct second subject to proceed. When the correct second subject was found, the first and second subjects had to remember their positions while the search continued for the third correct subject, etc.

When all five crewmembers had responded in the correct order, the green light was illuminated to signal that the problem had been solved. Following a between-problem pause of 30 seconds, the programmer moved to a new problem, the green light went off, the red light came on, and the crew was presented with a new sequence or "code" to solve.

The code-lock task was the only task in the battery that was paced by the crew (self-paced); all other tasks were forced-paced at specified rates. Each code-lock problem was scored in terms of the time required for solution, the total number of responses made, and the number of errors (or programmer-resetting responses) made. In addition, five information measures were derived and analyzed for this task. These are described in Appendix II.

## PROCEDURE

### EXPERIMENTAL DESIGN

Each of the crews worked four hours on each of seven days. Although each crew worked alone, all crews followed the schedule of hourly activities that is shown in table 2. The first three days of work were used for orientation and training, and the last four days were used for experimental testing. All seven days of work for a given group were consecutive except that no training or testing was performed on Saturdays or Sundays. For purposes of identification the groups were assigned the letters A, B, C, D, and E as they reported to the laboratory for participation in the study. Group A reported for work at 1300 hours, starting on a Wednesday. Groups B and C both started on a Monday, with group B reporting at 0800 hours and group C at 1300 hours. Groups D and E followed the same schedules as B and C, respectively, but both started on a Thursday following the completion of all testing of the other groups.

### TASK PROGRAM

During each of the four hours of testing on the fourth through the seventh days, each group worked according to the task program shown in table 3. Performance was scored only during the middle 50-minute (high-performance) section of the basic 1-hour program; however, the subjects were not told this and the program was repeated cyclically without interruption. Responses to the warning lights and to the auditory-vigilance signals were scored immediately after the occurrence of each signal. Responses to the probability monitoring task were cumulated automatically and recorded at the end of each 5-minute period. The target-identification and arithmetic computation tasks were scored and recorded at both the 10-minute midpoint and the 20-minute endpoint of their hourly presentations. Finally, performance on the code-lock task was recorded following the solution of each problem.

Table 2

Schedule of Hourly Activities

| Day* | Hour 1 | Hour 2 | Hour 3 | Hour 4 |
|---|---|---|---|---|
| 1 | Orientation | Instructions | Arith. Comp. & Monitoring | Arith. Comp. & Monitoring |
| 2 | Patterns & Monitoring | Arith. Comp. & Monitoring | Target ID & Monitoring | Target ID & Monitoring |
| 3 | Code Lock | Code Lock & Monitoring | Arith. Comp. Code Lock & Monitoring | Target ID Code Lock & Monitoring |
| 4-7 | Program** | Program** | Program** | Program** |

* Since days 1-3 were used for orientation and training, only the data obtained during the actual experimental testing on days 4-7 are presented in the results section, and there these last four days are referred to as days 1-4 rather than days 4-7.

** A single 1-hour task program was used cyclically over the four hours of performance on each day of experimental testing; the program is shown in table 3.

Table 3

Basic Hourly Task Program

| TASK | MINUTES 00 | 05 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Warning-Lights Monitoring | x | x | x | x | x | x | x | x | x | x | x | x | x |
| Auditory Vigilance | x | x | x | x | x | x | x | x | x | x | x | x | x |
| Probability Monitoring | | x | x | x | x | x | x | x | x | x | x | x | |
| Target Identification | | x | x | x | x | x | | | | | | | |
| Code-Lock Solving | | | | x | x | x | x | x | x | x | | | |
| Arithmetic Computation | | | | | | | | x | x | x | x | x | |

ORIENTATION AND TRAINING

During the first two hours of duty (on day 1), each crew was given orientation and general instructions pertaining to the study. This included general descriptions of the training and testing routines, the ground rules to be observed during the tests, and an explanation of the importance of the study and the need for cooperation. The role of the crew commander with respect to his responsibilities and the importance of both individual cooperation and crew coordination in performing the group tasks were emphasized. None of the groups was given specific instructions for handling the group tasks; rather, they were told that it was one of the responsibilities of the crew commander to develop efficient procedures and to standardize these. The commander, in turn, was encouraged to discuss alternative procedures with the members of his crew.

As shown in table 2 (p. 7), each crew spent about 10 hours in being trained on various combinations of the six tasks used. Except for the first hour on day 2 (when only individual performance was required on the target-identification task) and the first hour on day 3 (when the code-lock task was presented with no other activities), the training periods consisted of combinations of tasks that were used later during parts of the basic hourly testing program.

## RESULTS AND DISCUSSION

The results are presented and discussed in the four major subdivisions of this section. Because of the relatively large number of variates and tests involved in this study summaries of the analyses of variance computed are presented in Appendices I and II and referred to where necessary in the text. All the data presented are based on the last four days of experimental testing, and no data are presented for the three days each group spent in training. For this reason, and to avoid confusion in interpreting the data, the four days of testing are referred to subsequently as days 1-4.

RELIABILITY OF MEASURES

For each criterion measure a Pearson product-moment coefficient of correlation was computed between the appropriate combined mean scores of days 1 and 2 and the associated combined means of days 3 and 4. On all individual tasks, and on the individual-response parts of the target-identification task, these split-half reliability coefficients are based on the total number (N) of subjects, i.e., on an N of 25. No reasonably accurate coefficient could be computed for the group-performance measures of the target-identification task because of the small number of groups (5). Estimates of the reliabilities of measures used to evaluate code-lock solving were computed by using separately the performance scores obtained by each of the five groups under each of the three different program conditions (i.e., code-lock solving with target-identification, "alone," and with arithmetic computations). The split-half reliability coefficients obtained and the coefficients predicted for the entire four-day test period on the basis of the Spearman-Brown Prophecy Formula (ref. 6, pp. 174-177) are presented in table 4.

It is evident from the data of table 4 that all the measures are highly reliable; only one of the split-half reliability coefficients is not statistically significant, and even in the case of this measure (proportion erroneous code-lock responses) the predicted coefficient is high ($r = .629$, $P$ less than .05). Of the remaining thirteen measures, twelve of the associated split-half reliability coefficients are statistically significant at the .001 level of confidence; all thirteen of the predicted coefficients were significant at the .001 level.

Table 4

Reliabilities of the Various Measures of Performance

| Task | Measure of Performance | Reliability Coefficient ≠ | |
|---|---|---|---|
| | | Split-Half | Predicted |
| Arithmetic Computations | Percentage Correct Solutions | .957** | .978** |
| Auditory Vigilance | Percentage Correct Signal Detections | .777** | .874** |
| Warning-Lights Monitoring | Response Latency to Red Lights | .802** | .890** |
| | Response Latency to Green Lights | .702** | .825** |
| Probability Monitoring | Mean Detection Time | .746** | .854** |
| | Percentage Correct Signal Detections | .451* | .621** |
| Target Identification | Percentage Correct Individual Responses (Pattern Perception) | .751** | .858** |
| Code-Lock Solving | Mean Time per Individual Response | .959** | .979** |
| | Proportion Erroneous Responses | .458 | .629* |
| | Rate of Information Transmission | | |
| | Per Performance Period | .963** | .981** |
| | While Responding | .961** | .980** |
| | Relative Information-Transmission Rate | | |
| | Per Performance Period | .964** | .982** |
| | While Responding | .960** | .980** |
| | Response Equivocation Rate | | |
| | While Responding | .969** | .984** |

≠ Each reliability coefficient is based on an N of 25, except those for Code-Lock Solving where N = 15. The predicted reliability coefficients were obtained by application of the Spearman-Brown Prophecy Formula to the appropriate split-half coefficients.

*P less than .05; **P less than .001; all levels of significance are based on use of Fisher's z' transformation (ref. 6, pp. 305-307).

In general, these data are interpreted as corroborating the previously published data regarding the reliability of these performance measures (ref. 5, pp. 11-12.) Apparently, neither the use of the new group-performance tasks nor the attendant changes required in the basic schedule of activities had any degrading effect on the generally high reliability of each task and measure used. Indeed, the reliabilities reported here are higher than those reported previously, except in the case of the percentage correct measure of probability monitoring signal detections. The reliabilities of the measures used with the two group-performance tasks are as high as, or higher than, those of the individual tasks and on this basis it is concluded that all measures are suitable for use.

## INDIVIDUALLY PERFORMED TASKS

Four of the six tasks used in the test battery were individual-performance tasks in the sense that one individual's performance would not affect (nor be affected by) any other individual's performance, and also in the sense that where a subject was provided knowledge of results it was solely with regard to his own performance (in the auditory-vigilance task, no knowledge of results was provided any subject). The pertinent results obtained with these individual tasks are summarized below with particular emphasis on their effects on (and how they were affected by) the group-performance tasks. Summaries of the analyses applied are presented in Appendix I.

### Arithmetic Computations

The mean percentages of arithmetic problems correctly computed during each period (one-hour program presentation) and during each day (group of four periods) are shown in figure 3; analyses of these data are summarized in table 5, Appendix I. The data are treated separately for each of the two different conditions under which arithmetic computations were made, i.e., for the solutions computed with and without concurrent presentation of code-lock problems.

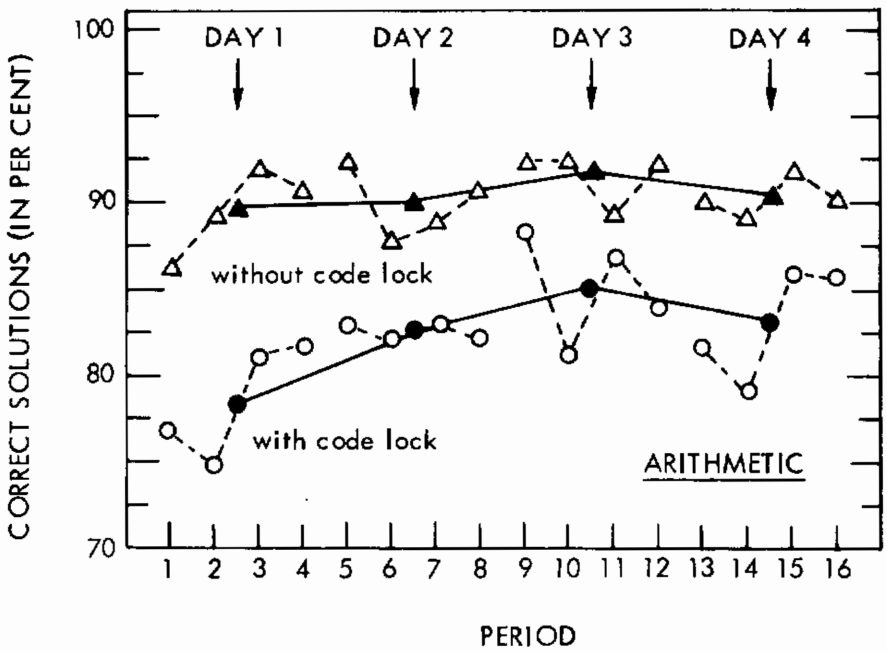


Figure 3 Mean Percentage of Arithmetic Problems Correctly Computed During Each Day and Each Program Period.

It is apparent from figure 3 that, as would be expected, the operator's performance on the arithmetic task was lowered on the average by requiring him to attend concurrently to the code-lock task. This may be interpreted as definite indication that the code-lock task introduces a "performance stress"; it is "stressful" in the sense that it makes the operator use up more of his available (mental-work) "channel capacity" than he can do without reducing the "bandwidth" that he subjectively assigns to arithmetic computations (cf. Garner and Taylor, ref. 7, for similar findings with tracking performances). Since two distinctly different levels of performance were obtained, it appears best to continue to analyze these data separately for the two conditions of performing arithmetic computations. The results of the analyses of variance support this decision further in showing that there were statistically significant differences among the four daily means obtained with concurrent presentation of the code-lock task, but not without.

## Auditory Vigilance

Separate analyses of auditory-vigilance performance as a function of the concurrent tasks were not considered feasible because of the small number of signals presented by the task program (e.g., there would have been many 10- and 20-minute subperiods in which no auditory-vigilance signal was presented). The mean percentage of auditory-vigilance signals correctly detected during each period and each day of testing is shown in figure 4, and the analysis is summarized in table 6, Appendix I. It is apparent that performance was relatively stable over the four days of testing. The differences among periods, though marked, were not statistically significant over and above a significant interaction of groups with periods.

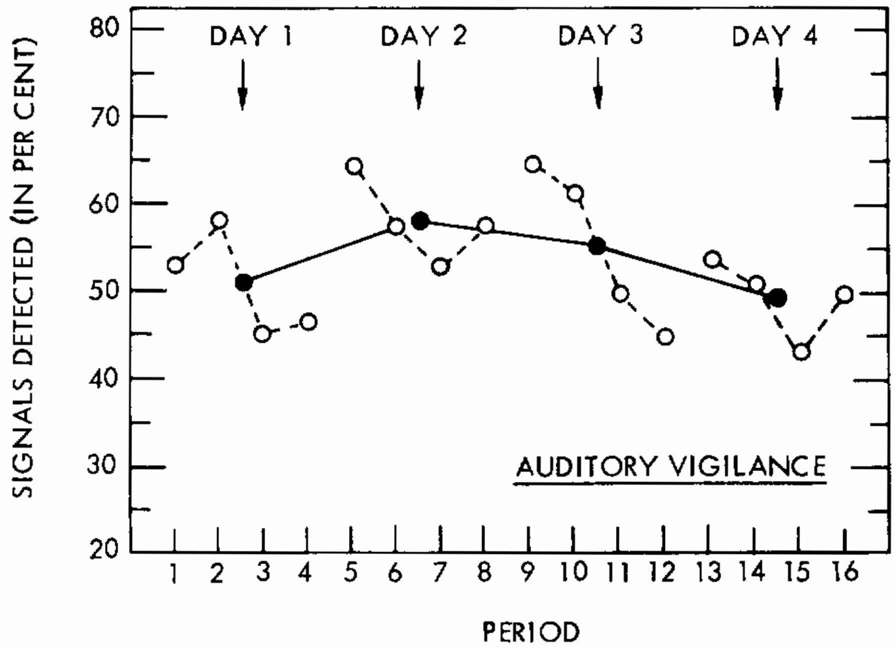


Figure 4 Mean Percentage of Auditory-Vigilance Signals Correctly Detected During Each Day and Each Program Period

## Warning-Lights Monitoring

The response latencies on this task were transformed using the scale shown in table 7, Appendix I. The mean response latency (normalized scale) in detecting warning-light signals during each period and each day is presented in figure 5 separately for the green and red lights; the analysis is summarized in table 8, Appendix I. As was the case with auditory vigilance, and for the same reasons, it was not feasible to break-down this performance into separate subperiod intervals. Essentially identical results were obtained with the two lights except for the longer response latencies to green lights and the specifics of the interactions of groups with periods. These two exceptions are sufficiently great, however, to warrant continuation of separate analyses for the two differently colored lights.

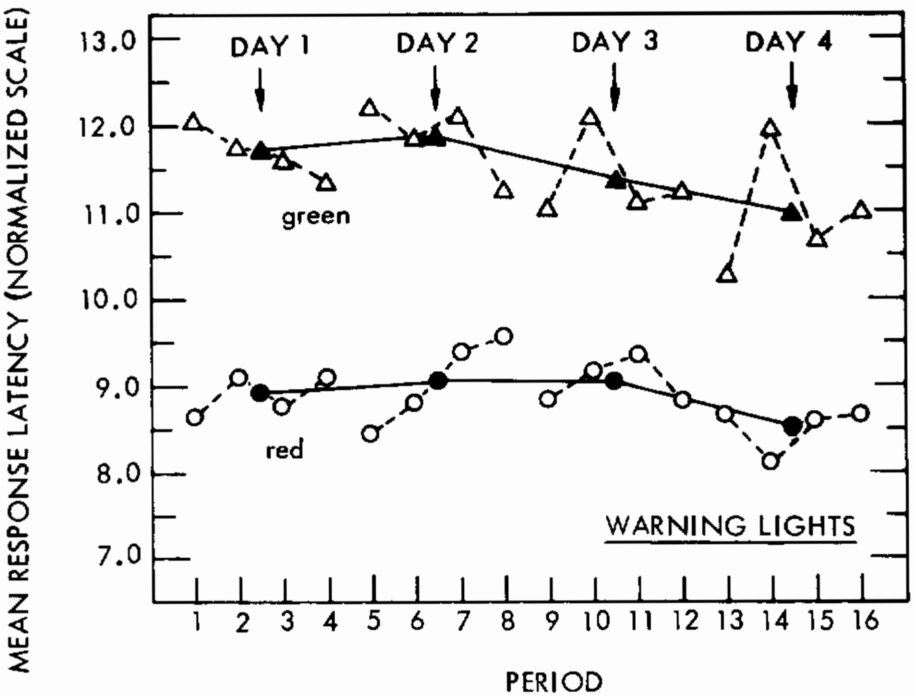


Figure 5 Mean Response Latency (Normalized Scale) in Detecting Red and Green Warning-Light Signals During Each Day and Each Program Period.

## Probability Monitoring

As was the case with the other vigilance and monitoring tasks, probability-monitoring performance was not partitioned into subsections of the basic hourly program. The two measures of performance taken are shown in figure 6; the mean percentage of correct detections of probability-monitoring signals is read off the ordinate on the right, and the mean detection time is read off the ordinate on the left, for each period and each day of testing. The analyses of these data are summarized in table 9, Appendix I.

Although the performance of this task was quite stable with respect to the percentages of detections (no statistically significant differences), there were significant changes in detection time. These changes were principally interactions of groups with days and with hours that seem to be correlated with the actual time of day during which the groups were tested. The data obtained seem to support the conclusion that the afternoon groups improved slightly in performance over the four days, whereas the morning groups declined markedly; also, within the four-hour daily work period, the two morning groups showed improvements, whereas the three afternoon groups did not. These findings are consistent with previously reported work (e.g., ref. 11, fig. 4, p. 12). Apparently, the probability-monitoring task is more likely to show consistent hourly and daily trends than the other individual-performance tasks used here.

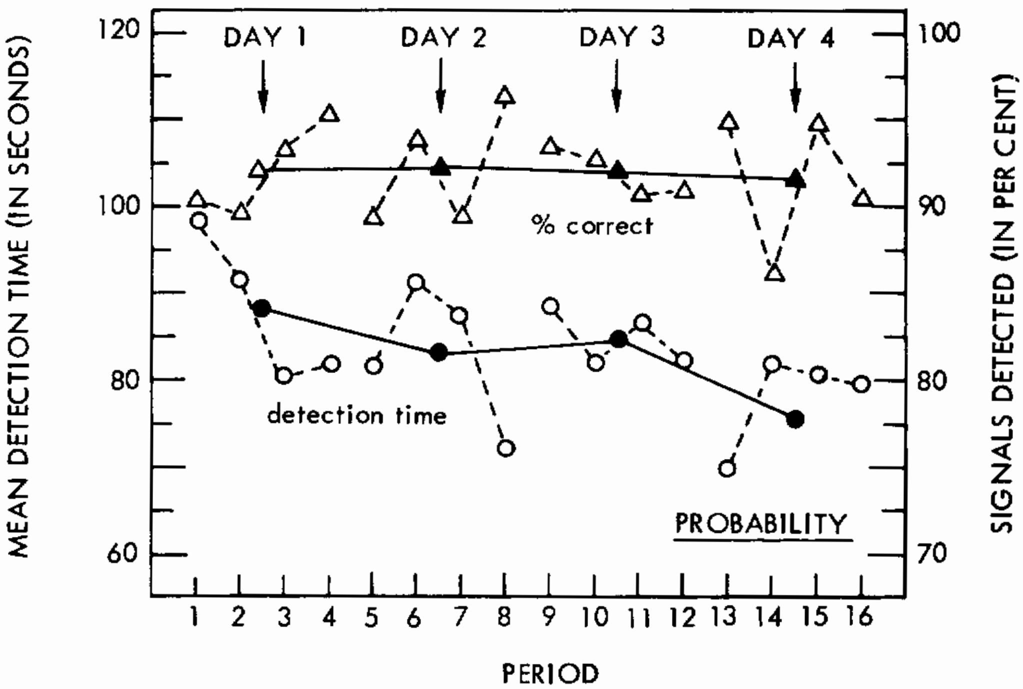


Figure 6 Mean Detection Time (Left Ordinate) and Mean Percentage of Correct Detections (Right Ordinate) of Probability-Monitoring Signals During Each Day and Each Program Period

## TARGET IDENTIFICATION

Although the group-performance measure of this task is dependent on individual proficiency in pattern discrimination, the relation is non-monotonic as indicated in table 1 (p. 5). Because of this, it was decided to use both group and individual measures of performance in the analysis of the results. In addition, since this task was presented both with and without concurrent presentation of the code-lock task, a separate analysis was carried out for each of these conditions.

### Individual Performance

The mean percentage of targets identified correctly by the individual operators during each period of testing is shown in figure 7; analyses of these data are summarized in table 10, Appendix II. A separate function is presented for performances with and without concurrent presentation of code-lock problems.

During the early periods of testing, performance without the code-lock task was apparently better than performance with the code-lock task. Although performance under both conditions improved with practice, improvement with the code-lock task was more rapid. Thus, on the fourth day of testing there was essentially no difference between the performances obtained under the two conditions of task presentation. This is interpreted as a demonstration that the time-sharing required by the code-lock task was sufficiently disruptive to produce initial decrements in performance, but that as the target-identification task became better learned, the load stress had less effect in producing decrements in performance. This interpretation is consistent with previous findings that the "pattern-discrimination" task was the one task in the battery that showed major improvement with time (ref. 4). A possible inference to be made from this improvement in performance is that continued practice results in a decrease in the subjective "perceptual bandwidth" demanded for identification of the targets. A further inference is that the effects of stressful tasks and performance or load stresses can be measured only when the operator is being asked to exceed his "channel capacity." This was apparently the case with target identifications during early periods and with arithmetic computations throughout (see also Garvey and Taylor, ref. 7).

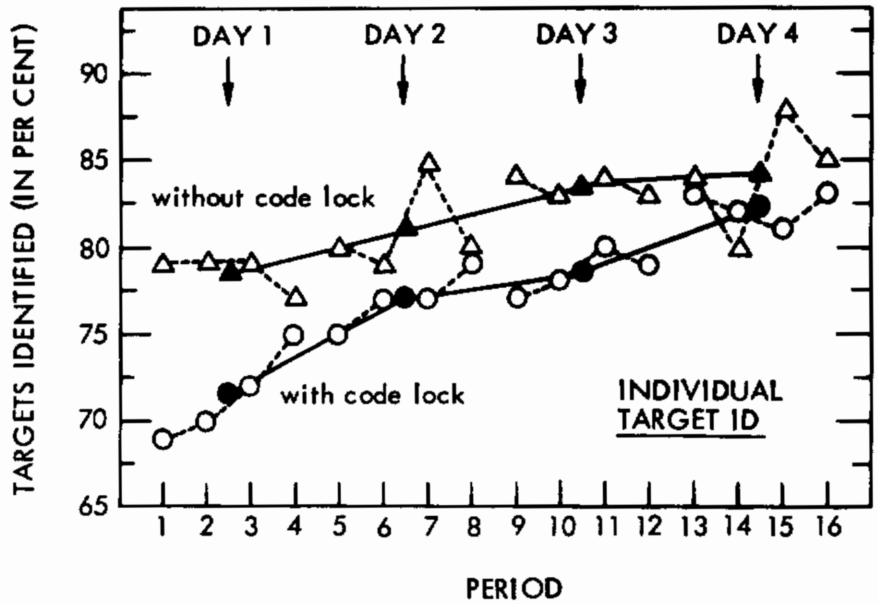


Figure 7 Mean Percentage of Targets Identified Correctly by Individual Operators (With and Without Concurrent Presentation of Code-Lock Problems) During Each Day and Each Program Period.

## Group Performance

The mean percentage of targets identified correctly by the five groups during each period of testing is shown in figure 8; data are presented separately for the two conditions of performance, i.e., with and without concurrent presentation of code-lock problems. The analyses of these data are summarized in table 11, Appendix II.

As would be expected from the dependence of the group-performance scores on the individual-performance scores in this task, the group performances are essentially similar to the individual performances. The hourly group scores are somewhat more variable than the individual scores. However, this is expected in that on those problems for which the correct individual answers are either three "sames" and two "differents" or two "sames" and three "differents," an error by one subject can make the group solution incorrect. Thus, in these cases, whereas a given subject contributes only one score in five with respect to the average of the individual performances, he has a "go-no-go" effect on the group score. There are apparent increases in performance over the four days of testing, but these are statistically significant only for the condition associated with concurrent presentation of code-lock problems; after the first day, performance under the two conditions was essentially the same.

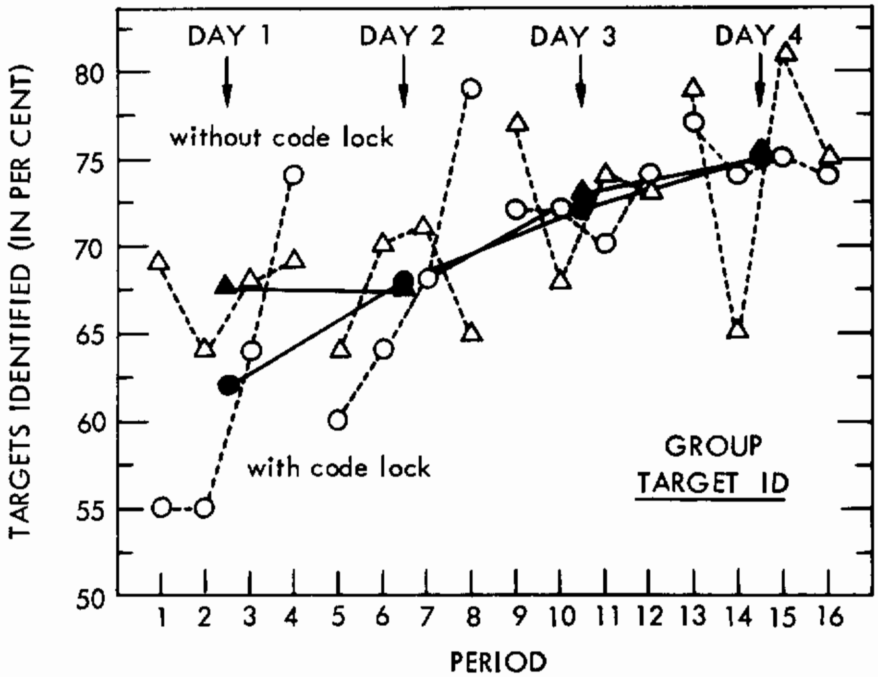


Figure 8 Mean Percentage of Targets Identified Correctly by the Five Groups (With and Without Concurrent Presentation of Code-Lock Problems) During Each Day and Each Program Period

## CODE-LOCK SOLVING

Seven different criteria, including five measures of various information-handling rates, were used to provide a rather complete analysis of the performances obtained in code-lock solving. As was noted from the data of table 4 (p. 9) all of these measures are of high reliability with the exception of the proportion of erroneous responses (split-half reliability coefficient of .458). A discussion of each of these measures, as well as the derivations of the information measures, is given in Appendix II. Also presented in Appendix II are the summary tables of analyses relating to these measures. A brief discussion of the most pertinent results is given in the present section.

### Time per Response

The mean time per individual key-pressing response in code-lock solving is shown in figure 9 for each performance period and each of the three performance conditions. It is apparent that two distinctly different levels of performance were attained. The average time required per response when solving code-lock problems concurrently with only the monitoring and vigilance tasks was lower than when solving these problems with either arithmetic computations or target identification in addition to the monitoring and vigilance tasks. Since both the target-identification and arithmetic tasks were programmed on a fixed schedule, the subjects had to time-share the code-lock task with the particular one of these tasks being presented concurrently at a given time. Thus, since the code-lock task was subject-paced, it was neglected during those times when problems were appearing on the concurrent task. However, as indicated in figure 9 and in the analyses summarized in table 12, Appendix II, there was a significant improvement in performance (lowering of mean time per response) for all three modes of presentation over the four days of testing.

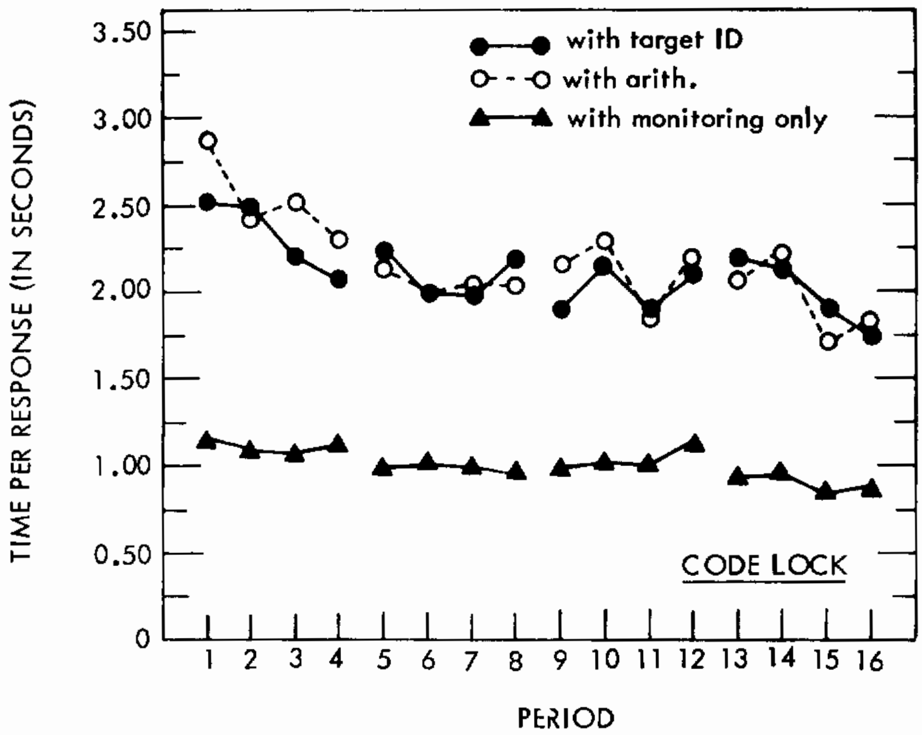


Figure 9 Mean Time Per Response While Performing the Code-Lock Task During Each Program Period.

## Proportion Erroneous Responses

The mean proportion of erroneous code-lock responses made under the three performance conditions is shown in figure 10, and the analyses of variance of these data are summarized in table 13, Appendix II. These data are based on the performances of only three groups (A, D, and E); because of apparatus problems, the error data of groups B and C were inaccurate and had to be discarded. An interaction of days with hours of testing is apparent for the performance of this task with the target-identification task. This interaction reflects not only the slight improvements over the four-hour test periods on days 3 and 4 versus slight declines on days 1 and 2, but also the wide variations that took place on day 1. Except for specific interactions with the groups, and group differences, the only other significant effect was the variation in the proportion of errors over the four days when the code-lock task was presented concurrently with arithmetic computations. This was also found with a composite score based on all data for a given hour of testing without regard to the concurrently presented tasks. No clear-cut interpretation of this effect can be offered. The mean proportion of errors on day 4 is lower than that on day 3 which suggests that there is no trend toward decrement, but there is apparently no definite trend toward improvement. The effect would appear to be best interpreted as a result of some unnoticed shift in the approach to the task taken by the groups.

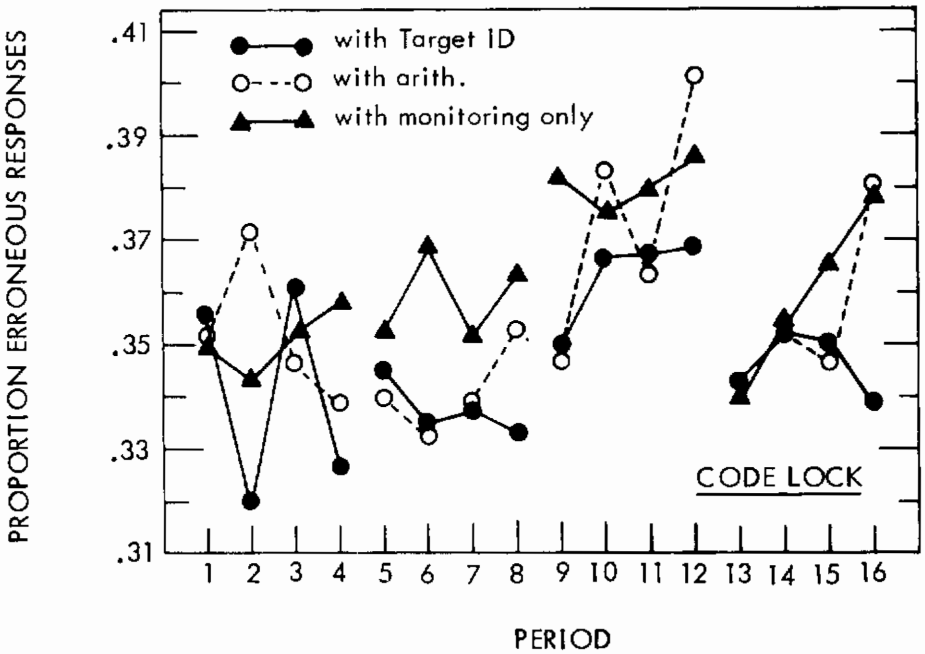


Figure 10 Proportion of Erroneous Code-Lock Responses During Each Program Period

## Information-Handling Rates

Five different information-handling-rate scores were used in a relatively complete informational analysis of the performances attained in code-lock solving. Each of these measures is defined, and the analyses of the results obtained with each are presented, in Appendix II. A factor analysis of the five information measures led to identification of two factors: one of these was highly loaded with the relative information rate while responding, or R(respdg), and the other was highly loaded with the equivocation rate while responding, or E(respdg), as shown in table 16, Appendix II (see also p.29f, Appendix II). The results obtained with these two measures are presented here; they are the only two of the five measures that will be used in future studies.

Relative Information Rate While Responding.--R(respdg) is shown as a function of the different periods of testing and the three performance conditions in figure 11; analyses of these data are summarized in table 14, Appendix II. There were no statistically significant effects attributable to the different periods. In general, however, it may be noted that performance of the code-lock task concurrently with the monitoring tasks only was better than its performance with arithmetic computations; performance with target identification was intermediate. Presumably, this occurred because the target identification task placed smaller time-sharing demands on the subjects than did the arithmetic task, but greater demands than the "monitoring-tasks-only" condition.

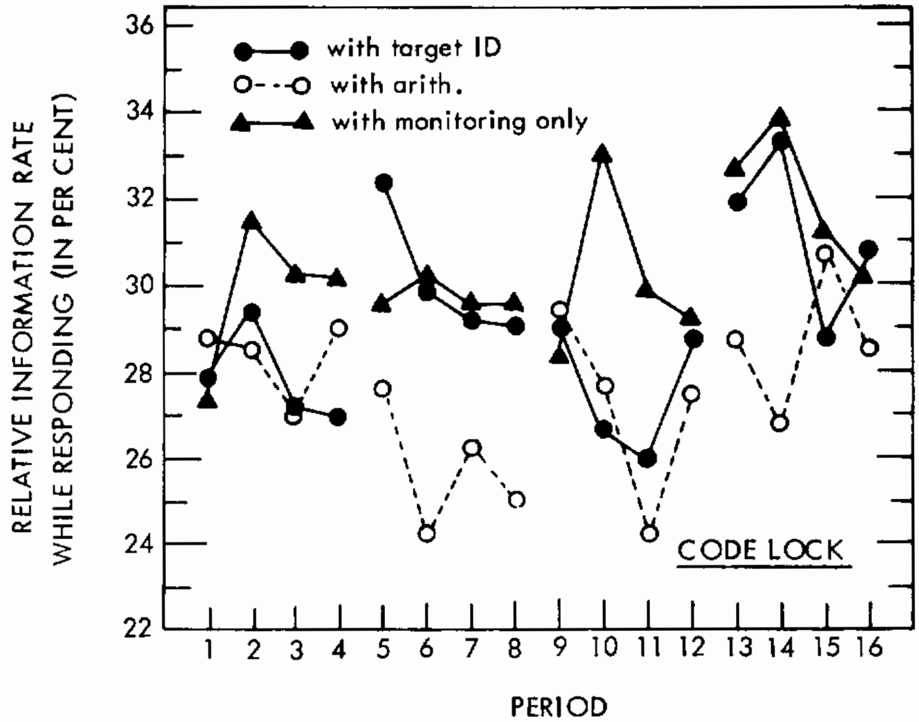


Figure 11 Mean Relative Information Rate While Responding to Code-Lock Problems During Each Program Period

Equivocation Rate While Responding.--E(respdg) is shown as a function of the different periods of testing and the different performance conditions in figure 12; analyses of these data are summarized in table 15, Appendix II. Although it appears from figure 12 that different levels of performance are obtained when the code-lock task is performed concurrently with only monitoring tasks as opposed to its being combined with the other major tasks, this difference is related principally to the "time" aspects of performance. The denominator of the equation that defines this measure includes the time spent in solving each problem. In the two combined presentation conditions, the subjects have to time-share this task with the concurrently presented task, and as was shown in figure 9 (p. 16) a larger mean time per response (and, thus, a larger denominator) results. The difference could reflect this change in the response times, therefore, without implying a change in the absolute equivocation. That is to say, the equivocation rate may be higher simply because the response rate is higher. Similarly, the gradual increase in equivocation rate over periods (and days) is also at least partially a reflection of the decrease in the mean time per response (see figure 9, p. 16), or increased rate of responding. Unfortunately, the lack of a clear-cut change in errors over periods makes impossible a more precise interpretation of the equivocation-rate data.

## SUMMARY

The performances of five-man groups on a test battery of six tasks were investigated in the present study. Four of these tasks were selected initially on the basis of an analysis of individual operator requirements for long-range, long-endurance weapon systems (ref. 1).

The remaining two tasks were designed and added as possible measures of crew, or group-dependent performances. In each of these two tasks, successful performance required interactions among crewmembers in the form of exchanges of information, cooperation, and coordination. One of the two tasks (target identifications) stressed individual proficiency as a principal determiner of the crew score, whereas the other (code-lock solving) stressed crew coordination. Both required whole-crew participation and interaction for successful performance, and both provided quantitative indices of crew performance.

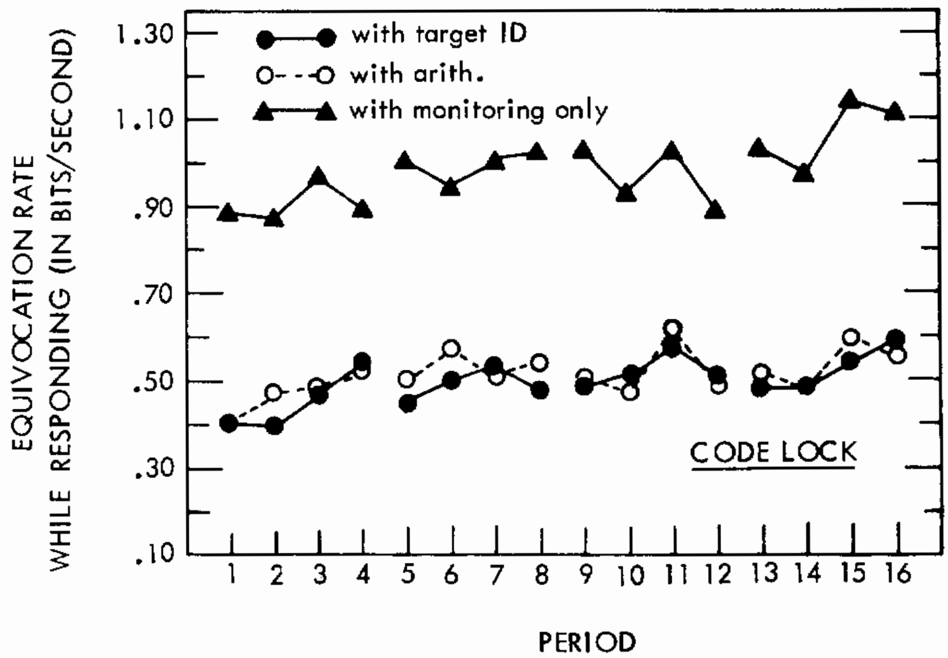


Figure 12 Mean Equivocation Rate While Responding to Code-Lock Problems During Each Program Period

The principal purpose of the study was to evaluate these newly added group-performance tasks and to determine in what ways they might be improved through further modification. Accordingly, the results obtained with these tasks were analyzed rather thoroughly. On the basis of these results the following conclusions seem justified:

1. All six tests in the battery and all of the criteria of performance used with these tests are of sufficiently high reliability to warrant selection of any of the measures for future use (with the possible sole exception of the proportion of erroneous responses in the code-lock task).

2. In the target-identification task, the function relating the probability of a group error with the number of individuals in error proved to be non-monotonic. In view of this, and of the difficulties it produces in interpreting the group-performance results, the sampling distribution of target-identification problems (as well as the display) will be changed in future uses of this task.

3. Seven different criteria, including five measures of various information-handling rates, were used in analyzing code-lock performances. A factor analysis of these latter five measures was computed and two orthogonal multiple-group factors were identified. The

relative information rate while responding had a loading of + .99 on one of these factors, and the equivocation rate while responding had a loading of + .98 on the others. Therefore, of the five information measures examined, only these two will be used in future studies.

4. In nearly all cases of scoring the code-lock task, the level of performance obtained when the task was performed with only the monitoring and vigilance tasks differed from the levels obtained when it was performed concurrently with either arithmetic computations or target identifications. Also, two different levels of performance were obtained with both arithmetic computations and target identifications when these tasks were performed under the two experimental conditions, i.e., either with or without concurrent presentation of code-lock problems. This was interpreted as illustrating performance or load stress whenever presentation of the code-lock task coincided with presentation of one of the other two major tasks. Accordingly, these tasks will be analyzed separately for different performance conditions in future studies.

In general, the data of this study demonstrate that it is feasible to develop and use crew tasks to measure group performance quantitatively in activities that require interactions among crewmembers, exchanges of information, cooperation, and coordination. Two such tasks have been used here with individual-performance tasks in the setting of an advanced-system crew-compartment mock-up. Multiple-task performances such as these can be used to produce the performance stresses that occur when man's limited channel capacity is overtaxed by the demands of his assigned duties. The study of such stresses, particularly in terms of their effects on (and from) group-dependent behavior, is of major importance to the designer of all crew-performance jobs and to development of the appropriate training requirements.

## REFERENCES


1. Adams, O. S. Aircrew fatigue problems during extended endurance flight, Phase 1: Planning. USAF WADC Technical Report 57-510, May 1958.

2. Adams, O. S., and Chiles, W. D. Human performance as a function of the work-rest cycle. USAF WADD Technical Report 60-248, March 1960.

3. Adams, O. S., and Chiles, W. D. Prolonged human performance as a function of the work-rest cycle. Marietta, Georgia, Lockheed Aircraft Corporation, ORD 273, May 1961.

4. Adams, O. S., and Chiles, W. D. Human performance as a function of the work-rest ratio during prolonged confinement. USAF ASD Technical Report 61-720, November 1961.

5. Adams, O. S., Levine, R. B., and Chiles, W. D. Research to investigate factors affecting multiple-task psychomotor performance. USAF WADC Technical Report 59-120, March 1959.

6. Edwards, A. L. Statistical methods for the behavioral sciences. New York : Rinehart, 1954.

7. Garvey, W. D., and Taylor, F. V. Interactions among operator variables, system dynamics, and task-induced stress. Journal of Applied Psychology, 1959, 43, 79-85.

8. Hicks, C. R. Fundamentals of analysis of variance: Part I -- The analysis of variance (ANOVA) model. Industrial Quality Control., 1956, 13(2), 17-20.

9. Hicks, C. R. ____________________: Part II -- The components of variance model and the mixed model. Ibid., 13(3), 5-8.

10. Hicks, C. R. ____________________: Part III -- Nested designs in analysis of variance. Ibid., 13(4), 13-16.

11. Ray, J. T., Martin, O. E., Jr., and Alluisi, E. A. Human performance as a function of the work-rest cycle: A review of selected studies. Washington, D. C.: National Academy of Science -- National Research Council Publication No. 882, 1961.

APPENDIX I

# SUMMARY TABLES OF ANALYSES OF VARIANCE OF INDIVIDUAL PERFORMANCE TASKS AND MEASURES

Table 5

Summaries of Analyses of Variance of the Mean Percentages of Arithmetic Computations Correctly Made With and Without Concurrent Presentation of Code Lock

| Source of Variation | df | With Code Lock | | Without Code Lock | |
|---|---|---|---|---|---|
| | | Mean Square | F | Mean Square | F |
| Groups (G) | 4 | 3346 | 1.79 | 998 | 2.31 |
| Periods | 15 | 323 | 2.01* | 86 | ---- |
| (Days) | (3) | (768) | (3.49*) | --- | ---- |
| (Hours) | (3) | (486) | (1.88) | --- | ---- |
| (D x H) | (9) | (121) | (1.12) | --- | ---- |
| Periods x Groups | 60 | 161 | 3.10*** | 91 | 1.63** |
| (D x G) | (12) | (220) | (4.23***) | (101) | (1.81*) |
| (H x G) | (12) | (259) | (4.98***) | (100) | (1.79*) |
| (D x H x G) | (36) | (108) | (2.08***) | (84) | (1.52*) |
| Subjects Nested in Groups | 20 | 1871 | n.t. | 433 | n.t. |
| S(G) x Periods | 300 | 52 | --- | 56 | ---- |
| Total (Sum of Squares) | 399 | (80815) | --- | (36073) | ---- |

* P less than .05; ** P less than .01; *** P less than .001; n.t., no test possible.

Table 6

Summary of an Analysis of Variance of the Mean Percentages of Auditory-Vigilance Signals Correctly Detected

| Source of Variation | df | Mean Square | F |
|---|---|---|---|
| Groups (G) | 4 | 4248.4 | --- |
| Periods | 15 | 1072.1 | 1.02 |
| Periods x Groups | 60 | 1052.5 | 4.39*** |
| (Days x G) | (12) | (1329.7) | (5.55***) |
| (Hours x G) | (12) | (1000.5) | (4.17***) |
| (D x H x G) | (36) | (977.5) | (4.08***) |
| Subjects Nested in Groups, S(G) | 20 | 5578.1 | n.t. |
| S(G) x Periods | 300 | 239.8 | --- |
| Total (Sum of Squares) | 399 | (279735) | --- |

*** P less than .001;
n.t., no test possible.

Table 7

Normalized Scale Transformation Used
With Response Latencies to Warning Lights*

| Response Latency (in thousandths of a minute) | Assigned Normalized Scale Value | Relative Frequencies Theoretical | Empirical |
|---|---|---|---|
| 15 or less | 1 | .25 | .25 |
| 16 | 2 | .61 | .25 |
| 17 | 3 | 1.28 | 1.23 |
| 18 | 4 | 2.48 | 1.97 |
| 19 | 5 | 4.43 | 2.96 |
| 20-21 | 6 | 6.69 | 8.87 |
| 22-23 | 7 | 9.27 | 9.36 |
| 24-25 | 8 | 11.93 | 9.36 |
| 26-29 | 9 | 12.93 | 15.52 |
| 30-34 | 10 | 12.93 | 12.32 |
| 35-43 | 11 | 11.93 | 12.81 |
| 44-57 | 12 | 9.27 | 9.61 |
| 58-80 | 13 | 6.69 | 6.16 |
| 81-127 | 14 | 4.43 | 4.43 |
| 128-400 | 15 | 2.48 | 2.46 |
| 401-785 | 16 | 1.28 | 1.23 |
| 786-1200 | 17 | .61 | .25 |
| 1201-2000 | 18 | .25 | .98 |

* Response latencies typically yield positively skewed distributions. This has been the result obtained with response latencies to warning lights in the several studies completed (refs. 2, 3, 4, and 5). In order to normalize the distribution it has been the practice to use a reciprocal transformation (refs. 2 and 3) or a specially derived normalized-scale transformation (refs. 4 and 5). The normalized scale used is presented in this table; it is based on a random sample of over 500 scores drawn from data previously collected. The theoretical (normal curve) and empirical (obtained in 500-item sample) relative frequencies are included to show the goodness of fit.

Table 8

Summaries of Analyses of Variance of Response Latencies (Normalized Scale) to Red and to Green Warning Lights

| Source of Variation | df | Red Mean Square | Red F | Green Mean Square | Green F |
|---|---|---|---|---|---|
| Groups (G) | 4 | 54.363 | 2.25 | 28.182 | --- |
| Periods | 15 | 3.684 | 1.07 | 8.746 | 1.50 |
| Periods x Groups | 60 | 3.430 | 2.18*** | 5.812 | 1.72** |
| (Days x G) | (12) | (6.950) | (4.42***) | (9.999) | (2.96**) |
| (Hours x G) | (12) | (2.745) | (1.75) | (4.034) | (1.20) |
| (D x H x G) | (36) | (2.484) | (1.58*) | (5.009) | (1.49*) |
| Subjects Nested in Groups, S(G) | 20 | 24.170 | n.t. | 35.005 | n.t. |
| S(G) x Periods | 300 | 1.571 | ---- | 3.373 | ---- |
| Total (Sum of Squares) | 399 | (1433.070) | --- | (2264.638) | ---- |

* P less than .05; ** P less than .01; *** P less than .001;
n.t., no test possible.

Table 9

Summaries of Analyses of Variance of Mean Detection Times and Mean Percentages of Probability-Monitoring Signals Detected

| Source of Variation | df | Time: Mean Square | Time: F | Percentage Detected: Mean Square | Percentage Detected: F |
|---|---|---|---|---|---|
| Groups (G) | 4 | 19922 | --- | 2174.20 | --- |
| Periods | 15 | 1286 | --- | 190.51 | 1.30 |
| Periods x Groups | 60 | 1866 | 1.82*** | 183.67 | 1.29 |
| (Days x G) | (12) | (2921) | (2.85**) | ( ---) | (---) |
| (Hours x G) | (12) | (1847) | (2.20*) | ( ---) | (---) |
| (D x H x G) | (36) | (1521) | (1.82**) | ( ---) | (---) |
| Subjects Nested in Groups, S(G) | 20 | 20407 | n.t. | 1315.82 | n.t. |
| S(G) x Periods | 300 | 1024 | --- | 138.99 | --- |
| Total (Sum of Squares) | 399 | (926201) | --- | (90588.9) | --- |

*P less than .05; ** P less than .01; *** P less than .001;

n.t., no test possible.

# APPENDIX II

## DESCRIPTIONS OF MEASURES USED WITH THE CODE-LOCK TASK AND ANALYSES OF RESULTS OBTAINED WITH BOTH GROUP TASKS

The code-lock task, one of the two group-performance tasks added to the test battery, had been untried previously. Since the major purpose of this study was the detailed evaluation of each of the new group-performance tasks, a variety of measures of the performance obtained in code-lock solving was employed. This kind of approach was not necessary with the target-identification task because of the background of experience already gained in its use as an individual-performance, "pattern-discrimination" task.

## CRITERIA OF PERFORMANCE

Seven different criteria were selected for use. Of these, two were more-or-less traditional measures of speed and accuracy, whereas the remaining five were derived measures of information-handling rates. All of these measures were based on the basic data recorded at the time of solution of each code-lock problem -- i.e., the total number of responses made, the total number of "resets" or erroneous responses made, and the total time required to solve the problem. These three items were then cumulated for each 10-minute subperiod of the interval spent in code-lock solving during each hour. The two basic measures were the mean _time per response_ and the _proportion of erroneous responses_. The information measures are described more fully below.

Assuming that each of the 120 possible code-lock sequences is used with the same frequency as every other sequence, then the sequences may be said to be equiprobable, and the amount of information (or uncertainty) per sequence is: $H(seq) = -\log p = -\log 1/120 = 6.90689$ bits/sequence, where the logarithms are taken to the base _2_. The _rate of information transmission while responding_ is defined as:

$$H(respdg) = N(seq)H(seq)/T(respdg), \text{ in bits/second}$$

where N(seq) is the number of sequences solved during a given subperiod, H(seq) is the average uncertainty per sequence or 6.907 bits/sequence, and T(respdg) is the cumulative time in seconds spent in solving the problems within the subperiod. This was one of the five informational measures used.

A second measure used was the _rate of equivocation while responding_. This was defined as:

$$E(respdg) = H_{max}(respdg) - H(respdg), \text{ in bits/second}$$

or, for computation,

$$E(respdg) = \frac{H(seq)\left[N(resps) - 5N(seq)\right]}{5T(respdg)}, \text{ in bits/second}$$

where N(resps) is the total number of responses made during a given subperiod, 5N(seq) is the minimum number of responses that could have been made to solve the same number of problems during the subperiod, and the other terms are as previously defined. The term, $H_{max}(respdg)$, is the maximum rate of information transmission possible with the rate of responding held constant.

The *relative rate of information transmission while responding* was the third informational measure used. This was defined as:

$$R(respdg) = (100)H(respdg)/H_{max}(respdg), \text{ in percentages}$$

or, computationally,

$$R(respdg) = (100)5N(seq)/N(resps), \text{ in percentages}$$

where all terms are as previously defined.

The remaining two of the five informational measures were based on the total performance attained during the entire subperiod -- i.e., including the 30-second pause between successive problem presentations. Thus, the *rate of information transmission per period* was defined as:

$$H(period) = N(seq)H(seq)/T(period), \text{ in bits/minute}$$

or, computationally,

$$H(period) = N(seq)/(8.687)T(period), \text{ in bits/second}$$

where T(period) is the total duration of the subperiod in minutes, and the other terms are as previously defined. It was not necessary to use the equivocation rate per period, E(period), since under the conditions of this study and with the definitions given here that term would correlate perfectly (but negatively) with H(period).

The *relative rate of information transmission per period* was the fifth, and final, informational measure used. This was defined as:

$$R(period) = 100H(period)/H_{max}(period), \text{ in percentages}$$

or, for computations,

$$R(period) = 50N(seq)/T(period), \text{ in percentages}$$

where $H_{max}$(period) is the theoretical maximum rate of information transmission in bits/minute during each period, i.e., it is based on the theoretical expectancy that two code-lock problems could be solved per minute (if all responses were made instantaneously), and if all the other terms are as previously defined.

Table 10

Summaries of Analyses of Variance of Mean Percentages of Targets Correctly Identified With and Without Concurrent Presentation of the Code-Lock Task (Individual Responses)

| Source of Variation | df | With Code Lock: Mean Square | With Code Lock: F | Without Code Lock: Mean Square | Without Code Lock: F |
|---|---|---|---|---|---|
| Groups (G) | 4 | 16955 | 1.48 | 9793 | 1.15 |
| Period | 15 | 4577 | 3.81*** | 2203 | 2.56** |
| (Days) | (3) | (20063) | (11.58***) | (6633) | (7.68***) |
| (Hours) | (3) | (1553) | (1.30) | (2117) | (2.45) |
| (D x H) | (9) | (423) | (----) | (756) | (----) |
| Periods x Groups | 60 | 1200 | 1.41* | 864 | --- |
| (D x G) | (12) | (1732) | (2.03*) | --- | --- |
| (H x G) | (12) | (1356) | (1.59) | --- | --- |
| (D x H x G) | (36) | (971) | (1.14) | --- | --- |
| Subjects Nested in Groups, S(G) | 20 | 11430 | n.t. | 8550 | n.t. |
| S(G) x Periods | 300 | 852 | --- | 997 | --- |
| Total (Sum of Squares≠) | 399 | (692660) | --- | (594180) | --- |

≠ Mean Squares and Sums of Squares have been multiplied by 10.
* P less than .05; ** P less than .01; *** P less than .001;
n.t., no test possible.

Table 11

Summaries of Analyses of Variance of Mean Percentages of Targets Correctly Identified With and Without Concurrent Presentation of the Code-Lock Task (Group Responses)

| | | With Code Lock | | Without Code Lock | |
|---|---|---|---|---|---|
| Source of Variation | df | Mean Square | F | Mean Square | F |
| Groups (G) | 4 | 7405 | n.t. | 21900 | n.t. |
| Periods | 15 | 27887 | 3.19*** | 13296 | 1.31 |
| (Days) | (3) | (64207) | (9.34***) | --- | --- |
| (Hours) | (3) | (36657) | (2.25) | --- | --- |
| (D x H) | (9) | (12858) | (1.87) | --- | --- |
| Periods x Groups | 60 | 8754 | n.t. | 10161 | n.t. |
| (D x G) | (12) | (10895) | n.t. | (16906) | n.t. |
| (H x G) | (12) | (16258) | n.t. | (26707) | n.t. |
| (D x H x G) | (36) | (5539) | n.t. | (2398) | n.t. |
| Total (Sum of Squares ≠) | 79 | (973150) | --- | (896720) | --- |

≠ Mean squares and sums of squares have been multiplied by 100.
*** P less than .001; n.t., no test possible.

## Table 12

Summaries of Analyses of Variance
of Mean Times Per Code-Lock Response

| Source of Variation | df | With Target ID: Mean Square≠ | With Target ID: F | Alone: Mean Square≠ | Alone: F | With Arith. Comp.: Mean Square≠ | With Arith. Comp.: F | Composite: Mean Square≠ | Composite: F |
|---|---|---|---|---|---|---|---|---|---|
| Groups (G) | 4 | 29365 | n.t. | 630 | n.t. | 46387 | n.t. | 12038 | n.t. |
| Periods | 15 | 2202 | 1.86* | 429 | 3.46*** | 4171 | 3.05** | 1130 | 4.91*** |
| (Days) | (3) | (4762) | (4.02*) | (1680) | (6.64**) | (13279) | (3.85*) | (3493) | (5.76*) |
| (Hours) | (3) | (2333) | (1.97) | (112) | (1.22) | (3477) | (4.10*) | (1202) | (8.84***) |
| (D x H) | (9) | (1304) | (1.10) | (118) | (1.28) | (1367) | (1.61) | (318) | (2.34*) |
| Periods x Groups | 60 | 1185 | n.t. | 124 | n.t. | 1369 | n.t. | 230 | n.t. |
| (D x G) | (12) | (1499) | (n.t.) | (253) | (n.t.) | (3451) | (n.t.) | (606) | (n.t.) |
| (H x G) | (12) | (931) | (n.t.) | (88) | (n.t.) | (874) | (n.t.) | (157) | (n.t.) |
| (D x H x G) | (36) | (1164) | (n.t.) | (93) | (n.t.) | (840) | (n.t.) | (129) | (n.t.) |
| Total (Sum of Squares ) | 79 | (221563) | --- | (16381) | --- | (330240) | --- | (78914) | --- |

≠Mean Squares have been multiplied by $10^4$.
*P less than .05; ** P less than .01; *** P less than .001;
n.t., no test possible.

Table 13

Summaries of Analyses of Variance of Proportions Erroneous Code-Lock Responses Made by Groups A, D, and E Over All Four Days of Testing

| Source of Variation | df | With Target ID | | Alone | | With Arith. Comp. | | Composite | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean Square≠ | F | Mean Square≠ | F | Mean Square≠ | F | Mean Square≠ | F |
| Groups (G) | 2 | 189 | 3.67* to 4.15* | 302 | 5.87** to 6.64** | 484 | 9.41*** to 10.64*** | 234.7 | 4.56* to 5.16** |
| Days (D) | 3 | 145 | 1.91 | 191 | 1.12 | 227 | 4.41** to 4.99** | 139.7 | 2.72* to 3.07* |
| Hours (H) | 3 | 34 | --- | 52 | 1.01 to 1.14 | 131 | 2.55 to 2.88* | 25.0 | --- |
| D x H | 9 | 45 | 5.92*** | 21 | --- | 64 | 1.24 to 1.41 | 11.4 | --- |
| D x G | 6 | 62 | 1.21 to 1.36 | 179 | 3.48** to 3.93*** | 70 | 1.36 to 1.54 | 37.0 | --- |
| H x G | 6 | 47 | --- to 1.03 | 72 | 1.40 to 1.58 | 47 | - - to 1.03 | 33.2 | --- |
| D x H x G | 18 | 76 | 1.60* to 1.67* | 24 | --- | 56 | 1.09 to 1.23 | 13.8 | --- |

Where two values of F are given, they are the limits of F computed on the basis of maximum and minimum possible theoretical within cell variances, or MS = (1/2)(1/2)/N and MS = (1/3)(2/3)/N, respectively. The mean value of N = 486 was used. Where a single value of F is given, it is the appropriate ratio based on another error term (e.g., significant interaction of fixed variable with a random one).

*$P < .05$; ** $P < .01$; *** $P < .001$

≠ Mean squares have been multiplied by $10^5$.

Table 14

Summaries of Analyses of Variance of Relative Information Rates While Responding to Code-Lock Problems

| Source of Variation | df | With Target ID | | Alone | | With Arith. Comp. | | Composite | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean Square ≠ | F | Mean Square ≠ | F | Mean Square ≠ | F | Mean Square ≠ | F |
| Groups | 4 | 168529 | n.t. | 19002 | n.t. | 234750 | n.t. | 85255 | n.t. |
| Periods | 15 | 19676 | 1.60 | 12899 | --- | 15229 | --- | 3986 | --- |
| Periods x Groups | 60 | 12265 | n.t. | 13778 | n.t. | 17982 | n.t. | 4058 | n.t. |
| Total (Sum of Squares≠) | 79 | (1705203) | --- | (1096225) | --- | (2246035) | --- | (644335) | --- |

≠ Mean Squares and Sums of Squares have been multiplied by $10^6$.
n.t., no test possible.

Table 15

Summaries of Analyses of Variance
of Equivocation Rates While Responding
to Code-Lock Problems

| | | With Target ID | | Alone | | With Arith. Comp. | | Composite | |
|---|---|---|---|---|---|---|---|---|---|
| Source of Variation | df | Mean Square⁄ | F | Mean Square⁄ | F | Mean Square⁄ | F | Mean Square⁄ | F |
| Groups (G) | 4 | 185018 | n.t. | 25545 | n.t. | 277288 | n.t. | 159250 | n.t. |
| Periods (P) | 15 | 12781 | 2.81* | 32057 | 2.64* | 13204 | 1.88* | 11503 | 4.97*** |
| (Days) | (3) | (22211) | (2.86) | (94030) | (3.88*) | (16564) | (1.44) | (22582) | (5.33***) |
| (Hours) | (3) | (25827) | (9.91**) | (34120) | (6.80**) | (18160) | (3.44) | (26017) | (10.61***) |
| (D x H) | (9) | (5288) | (1.28) | (10712) | (1.02) | (10432) | (1.71) | (2973) | (1.82) |
| Periods x Groups | 60 | 4543 | n.t. | 12135 | n.t. | 7015 | n.t. | 2316 | n.t. |
| (D x G) | (12) | (7753) | (n.t.) | (24220) | (n.t.) | (11487) | (n.t.) | (4234) | (n.t.) |
| (H x G) | (12) | (2597) | (n.t.) | (5019) | (n.t.) | (5280) | (n.t.) | (2453) | (n.t.) |
| (D x H x G) | (36) | (4122) | (n.t.) | (10478) | (n.t.) | (6102) | (n.t.) | (1630) | (n.t.) |
| Total (Sum of Squares⁄ ) | 79 | (1204384) | --- | (1311107) | --- | (1728102) | --- | (948490) | --- |

⁄ Mean Squares and Sums of Squares have been multiplied by $10^6$.
* P less than .05; ** P less than .01; *** P less than .001;
n.t., no test possible.

## FACTOR ANALYSIS OF THE INFORMATION MEASURES

The results obtained with each of the five information measures were used to obtain all the possible intercorrelations among the measures. These intercorrelations were then factor analyzed. The intercorrelations, rotated factor loadings, residuals, and communalities resulting from this analysis are presented in table 16. It is apparent that there are two multiple-group factors, A and B, each of which is heavily loaded with a single variable. Factor A will be identified as the relative information rate while responding, and Factor B will be identified as the equivocation rate while responding. In terms of the optimization of performance, these two factors are relatively easy to interpret: the first factor (A) should be maximized, whereas the second factor (B) should be minimized. Since the two measures, R(respdg) and E(respdg), are so highly and orthogonally loaded on the two factors, these two measures will be used in the future to represent all the informational data collected with the code-lock task.

Table 16

Intercorrelations, Rotated Factor Loadings, Residuals, and Communalities of Information Measures of Performance in Code-Lock Solving ≠

| Information Measure | Information Measure | | | | | Factor Loading | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | A | B |
| 1. Relative Information Rate per Period | (.970) | .998 | .944 | .829 | .557 | .53 | .83 |
| 2. Rate of Information Transmission per Period | .022 | (.983) | .944 | .828 | .524 | .51 | .85 |
| 3. Rate of Information Transmission While Responding | -.008 | -.017 | (.942) | .886 | .476 | .45 | .86 |
| 4. Equivocation Rate While Responding | -.006 | -.025 | .025 | (.962) | .062 | .04 | .98 |
| 5. Relative Information Rate While Responding | .016 | .003 | .013 | .003 | (.980) | .99 | .02 |

≠ Figures in upper-right half of the matrix are intercorrelations; those in the lower-left half are residuals. Communalities are given in parentheses on the diagonal; these are consistent with the reliabilities reported in table 3, p. 7.