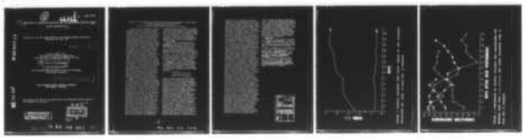
AD-A066 719 NAVAL AEROSPACE MEDICAL RESEARCH LAB DETACHMENT MICHO--ETC F/G 5/9
DEVELOPMENT OF A PERFORMANCE EVALUATION TEST FOR ENVIRONMENTAL --ETC(U)
1979 D L DAMOS, R S KENNEDY, A C BITTNER
UNCLASSIFIED TM-79-01 NL

| OF |
AD
A066719



END
DATE
FILMED
5-79
DDC

(11)

LEVEL II ✓TM-79-01

AD A0 66719

(6) DEVELOPMENT OF A PERFORMANCE EVALUATION TEST FOR ENVIRONMENTAL RESEARCH (PETER): CRITICAL TRACKING TEST

Presented at the 50th Annual Meeting of the Aerospace Medical Association
Washington, DC, May, 1979

(14) TM-79-01

Diane L. Damos, Ph.D.
Department of Industrial Engineering
State University of New York at Buffalo, New York

(10) Diane L. /Damos,
Robert S./ Kennedy, ~~Ph.D.~~
Alvah C./ Bittner Jr., Ph.D.

Naval Aerospace Medical Research Laboratory Detachment
New Orleans, Louisiana

(11) 1979 (12) 5p.

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY
NEW ORLEANS, LOUISIANA

DDC FILE COPY

DDC
RECEIVED
APR 3 1979
B

391 221 79 03 30 023

Development of a Performance Evaluation Test for Environmental Research (PETER): Critical Tracking

Diane Damos, State University of New York at Buffalo
Robert S. Kennedy and Alvah C. Bittner Jr.
Naval Aerospace Medical Research Laboratory Detachment, New Orleans, LA

A need exists for a standardized performance test battery to study the effects of unusual environments which may be encountered by Navy personnel. Such a test battery must be sufficiently sensitive so that the deleterious effects of these exotic environments can be identified. The concern at the Naval Aerospace Medical Research Laboratory (NAMRL) is primarily with inertial environments. Of particular interest from a performance standpoint are the very low ($<$1Hz) frequency motions which occasion seasickness and the higher ($>$1Hz) vibrations which have biodynamic effects. A research program is underway to develop a performance test battery with early emphasis on tests which tap information processing, cognitive and perceptual functions. The general plan for the development of the Performance Evaluation Test for Environmental Research (PETER) is discussed elsewhere (5). The present study is the second in a series which report the results of evaluations of various tests. The Critical Tracking Test selected for study (3), differs somewhat from other tracking tasks in that it requires the operator to stabilize an unstable control element. The degree of instability of the control element is represented by a variable λ, which is the sum of the effective time delays of the display, and operator perceptual processing, neural transport, and neuromuscular delays. The value of λ at which the operator can barely control the system is used as a dependent measure of his performance. The purpose of the study described below was to obtain baseline measures of performance on a Critical Tracking Task to ascertain how much baseline pretesting is required for stability.

Method: The Critical Tracking Task with autopacer used in this experiment was instrumented to replicate that of Jex, McDonnell and Phatak (3). The function of the autopacer was initially to increase the degree of instability quickly and then more slowly as the operator's control limits were approached. The only input to this system was the operator's remnant. The task was displayed on a 14 cm, circular CRT with two sets of brackets painted on its surface. The first set of brackets was separated by 0.8 cm, representing the "good performance" range. Subjects were informed that the best performance resulted when the cursor was kept within these brackets. The large brackets, which were separated by 8.0 cm denoted the range outside of which control was lost easily. The cursor moved only in the vertical dimension and was controlled by compatible forward -backward movements of an isometric control stick. The trial ended when the subject lost control of the cursor and it reached the edge of the display. With no human controller, λ =.84. The isometric control stick was inserted in a table top slightly to the right of the display. The task logic was programmed on an EAI PACE TR 48 Analog Computer. The experimenter read the trial value of λ from a digital display. A repeated measures design was used. Each subject received 15 trials on each of 15 consecutive weekdays. Eighteen Navy enlisted men between the ages of 19 and 24, With 20/20 corrected vision, participated in the experiment. All subjects were volunteers recruited, evaluated and employed in accordance with Secretary of the Navy Instruction 3900.39 and Bureau of Medicine and Surgery Instruction 3900.6.

Results: The overall impression is of a learning curve which may not be level by Day 15, ~~(Figure 1)~~. The means range from 4.38 to 6.75 while the standard deviations vary from .62 to .95. An analysis of variance conducted on the data revealed practice and subjects effects, ($P < .01$). The correlations deserve special mention: The average of the test-retest correlations for Days 1 - 5 is .639; for Days 6 - 10, .767; and for Days 11 - 15, .831. Figure 2 shows correlation coefficients of Days 1, 2, 4, 9 and 13 with the days which followed. A generally declining function is evident which is similar to those previously reported, e.g., by Jones ~~(4)~~.

Discussion: The majority of exotic environment experiments have been concerned with demonstrating that a given task is sensitive to the environment under consideration. The rationale for determining the sensitivity of a given task follows the form of "Student's t-Test" for repeated measures:

$$(1) \qquad t = \frac{\bar{X}_e - \bar{X}_c}{\sqrt{(SD_e^2 + SD_c^2 - 2r_{ec}SD_eSD_c)/N}}$$

where X_e and X_c are the respective mean performances for the experimental and control conditions, SD_e and SD_c are the respective standard deviations, r_{ec} is the correlation between scores in two conditions, and N the number of subjects. Generally, the experimenter attempts to stabilize performance on the task before the subject is exposed to the exotic environment. Thus, $\bar{X}_c$ represents a baseline for determining performance changes induced by the environment and is often obtained after several practice sessions. A statistical difference between the baseline mean and measures obtained during the experimental condition, $\bar{X}_e$ would indicate that the task is sensitive to the environment. More sophisticated statistical treatments, which use several pre-, per- and post- measures, generalize on this approach. In most of these approaches symmetry of the correlations is required i.e., the correlations between all sessions must be equal. This assumption is not ordinarily met and will be discussed below. Although practice to a baseline is the most often used approach, determining the point at which performance has "stabilized" is frequently difficult. This may be because performance invariably continues to improve, and an asymptote, while expected, is frequently difficult

- A -

79 03 20 023

to approach. Also, changes in motivation and fatigue can obscure the approach to asymptote (1). Therefore, it is sometimes impractical to obtain a baseline which is stable enough to detect changes in performance induced by the experimental treatment. This problem was apparent in Figure 1 where asymptotic performances were not attained until well into the third week, if at all. Throughout approximately 180 trials the mean values continued to increase. Although the terminal value is similar to that reported by Jex, et al (3), performance may have reached only a temporary plateau. Hence changes in performance induced by an exotic environment might be obscured by learning. To circumvent the problem associated with an unstable baseline, some have attempted to improve the sensitivity of the t-test by addressing elements (SD_c, SD_e, r_{ec} & $\bar{N}$) of the denominator (Eq 1). For exotic environments, however, there are practical limitations to increasing sample size and so repeated measures designs have been employed in order to control variability, i.e. the denominator of (Eq 1). Unwanted, although usually inherent, sequence effects of different sorts (e.g., factor structure changes (2) often makes this approach untenable. Moreover, SD often changes as $\bar{X}$ increases (4) when learning occurs, and large changes in SD can result in lowered reliability. The last element to be addressed is the "sustained" reliability of the retest. It is felt that this is a frequently overlooked statistic at least from the standpoint of improving the precision of a performance test in an unusual environment. Those who have studied how r changes with repetitions were more interested in the subject of skill acquisition, but the findings are directly relevent to PETER and may prove very useful in interpreting the findings. For example, Days 11 - 15 had an average correlation of .831. This value may indicate that the test-retest reliability of the critical tracking task is sufficiently high to provide a sensitive t-test even though the baseline is not stable and the variances are not particularly small. Concerning the reliability, however, an inspection of Figure 2 shows that while reliabilities are generally high, they decline as a function of remoteness from the trial with which they are compared. That is, retest reliabilities are poorer with repeated trials. However, the overall reliability of performance after day 4 appears to be substantially higher than for previous days suggesting that about 5 days practice (75 trials) should be used to attain stability. In summary, many tasks have not been considered for use in exotic environment experiments because their performance does not stabilize without extensive practice. Although performance on the critical tracking task improved substantially during approximately 180 trials, the test-retest reliabilities, while declining with increasing trial differences, were reasonably high. This indicates that, while the power of a statistical comparison (e.g., t-test) may be sufficient to detect performance differences induced by an exotic environment despite the changing baseline, too many sessions (or trials) may attenuate the power. It further suggests that the nature of what is being measured by the critical task may be changing, in "factor structure". It appears that close attention must be paid to the test-retest reliabilities in addition to the stability of the baseline and the size of the variance in selecting battery tasks. From considerations of test-retest reliability, it may be found that tests that usually would be excluded from an exotic environment experiment would be most sensitive to detection of environmental effects. Moreover, inspection of where the reliability reaches its highest and most stable point may be employed as a criterion for how much pretesting is required.

References:

1. Bradley, J. V., 1969. Practice to an asymptote. *Journal of Motor Behavior*. *1*: 285-295.
2. Fleishman, E. A., 1960. Abilities of different stages of practice in rotary pursuit performance. *Journal of Experimental Psychology*. *60*: 162-171.
3. Jex, H. R., McDonnell, J. D. and Phatak, A. V., 1966. A "critical" tracking task for manual control research. *IEE Transactions on Human Factors in Electronics*. HFE-7: 138-145.
4. Jones, M. B., 1969. *Principles of skill acquisition -Differential processes in acquisition*. New York: Academic Press,
5. Kennedy, R. S., and Bittner, A. C., Jr., 1977. The development of a Navy Performance Evaluation Test for Environmental Research (PETER) *Productivity Enhancement: Personnel Performance Assessment in Navy Systems*, Naval Personnel Research and Development Center, San Diego, CA. Oct. 1977, (AD A056047).

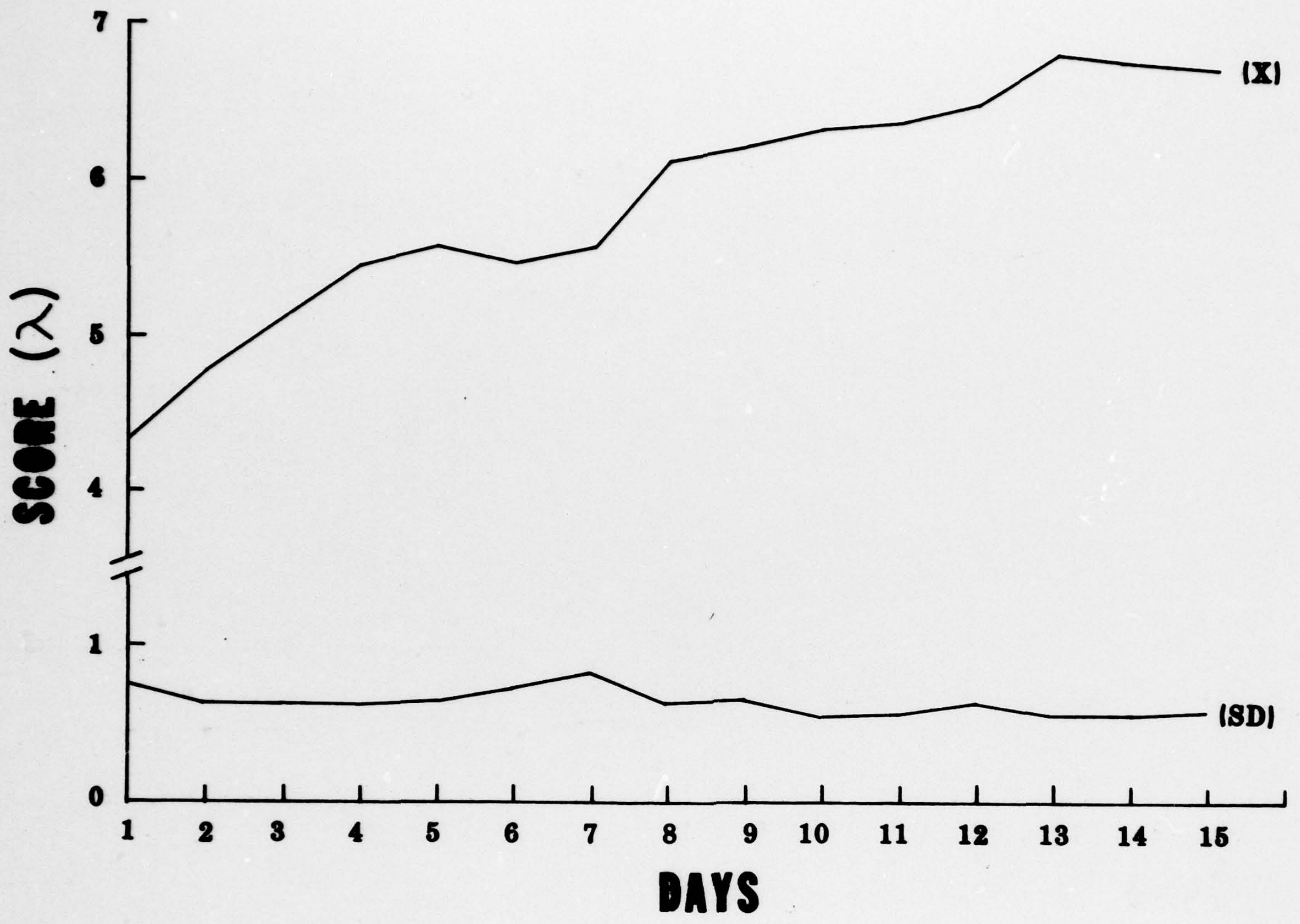


FIGURE 1. CRITICAL TRACKING TEST GROUP MEANS (X) AND STANDARD DEVIATIONS (SD) OVER 15 DAYS FOR 19 SUBJECTS

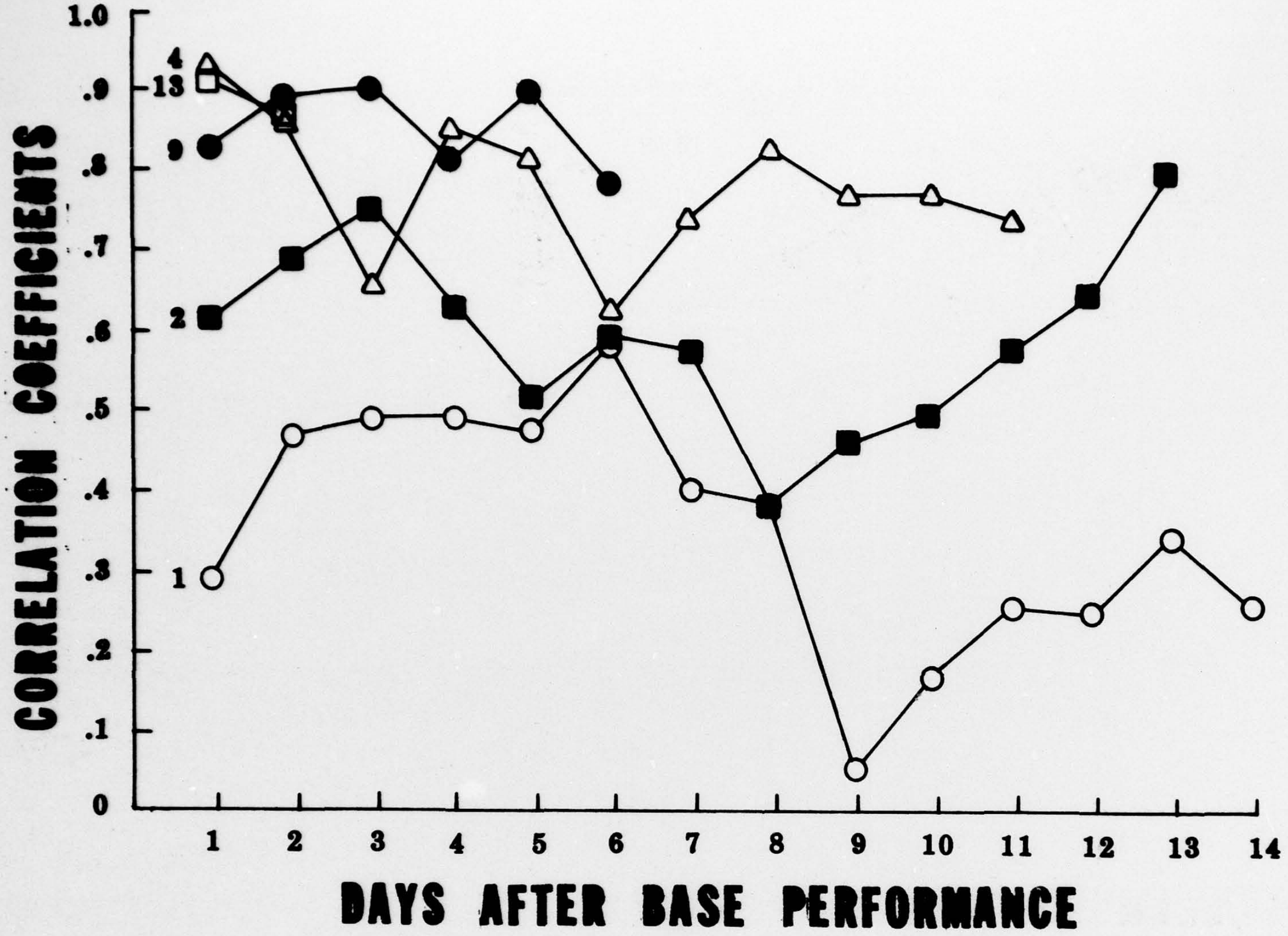


FIGURE 2. COMPARISONS OF CRITICAL TRACKING TEST RELIABILITIES FOR SELECTED BASE DAYS (1, 2, 4, 9, & 13) AND THOSE FOLLOWING OVER 15 DAYS FOR 19 SUBJECTS