TRAINING CHARACTERISTICS OF THE CRITERION TASK SET WORKLOAD ASSESSMENT BATTERY

Robert E. Schlegel

School of Industrial Engineering
University of Oklahoma
Norman, Oklahoma

Clark A. Shingledecker

Ergometrics Technology, Inc.
Dayton, Ohio

ABSTRACT

An evaluation of the Criterion Task Set was performed to determine the training requirements for the various tasks. Twenty subjects were divided into four groups. One group trained on all nine tasks in the battery. The other three groups trained on different three-task subsets. All subjects trained for two hours per day on five consecutive days. Response time, accuracy and subjective workload measures were obtained for each trial.

The required number of trials for stable performance ranged from two to six with a mode of five. Slight improvements were observed on some tasks after eight to ten trials. Performance by the group trained on all nine tasks was equivalent on half of the tasks and worse on the other half. Subjective workload ratings were highly correlated with the actual performance scores.

INTRODUCTION

The USAF Criterion Task Set (CTS) is a human performance test battery composed of nine tasks which measure independent information processing resources. The CTS is based on a synthesis of current human performance models (Wickens, 1981; Sternberg, 1969) which hypothesize that human performance is dependent on a number of information processing resources, stages and specific functions. The three major divisions are perceptual input, central processing and motor output.

The elements of the combined model were operationally defined in terms of the characteristics of tasks which would place predominant demands on them. These definitions were then used to select candidate tasks for the CTS. The nine tasks within CTS Version 1.0 are listed in Table 1. All tasks, except Interval Production, may be conducted at three distinguishable workload levels: low, medium and high. Each of the tasks was subjected to parametric study to establish prescribed testing conditions and loading levels.

A primary application of the CTS is as a test instrument to evaluate the relative sensitivity, reliability and intrusiveness of a variety of available workload measures. Workload metric evaluation studies (Shingledecker et al., 1983) have illustrated the potential variation in diagnosticity that exists among workload measures.

Table 1. CTS Version 1.0 Tasks by Information Processing System Division.

STAGE	TASK	CODE
Input	Probability Monitoring	PM
Central	Memory Search	MS
	Continuous Recall	CR
	Spatial Processing	SP
	Mathematical Processing	MP
	Linguistic Processing	LP
	Grammatical Reasoning	GR
Output	Unstable Tracking	UT
	Interval Production	IP

In addition to the original design intention stated above, the CTS forms an independent Performance Assessment Battery which may be used to assess the effects of various stressors on individual components of the human information processing system.

The battery is presented on a CRT display while the subject responds using a keypad and other controls designed to

allow full response capability to the various tasks. In most tasks, individual stimuli are presented sequentially and response time and accuracy scores are collected. For the input/output tasks, alternative measures are used. A single trial for each level lasts three minutes.

OBJECTIVES

As part of an overall evaluation of the training characteristics of the CTS, a study was initiated with the following objectives:

(1) determine the required number of sessions to achieve asymptotic performance on each CTS task when naive subjects are trained on all tasks or some subset of tasks concurrently,
(2) compare the performance of subjects concurrently trained on all nine tasks with the performance of subjects trained on various three-task subsets,
(3) relate task performance to a subjective workload assessment measure,
(4) examine the inter-subject variability and inter-task performance relationships, and
(5) develop the structure of a testing protocol model to allocate and sequence task training trials within a limited training period.

EXPERIMENTAL METHODOLOGY

Subjects

Four different subject groups were established to compare the performance of subjects trained on all nine tasks (Group A) vs. individual three-task subsets (Groups B, C, and D). In assigning tasks to groups, an attempt was made to balance the category of information processing, type of visual stimuli, task difficulty and other characteristics. Each task was assigned to the overall set and one subset.

Twenty male subjects, age 18 to 25 years, were randomly assigned to the four groups, five subjects per group. Each subject trained on the appropriate CTS tasks for two hours per day on five consecutive days. Due to training time limitations, five trials of each workload level of each task were performed for Group A tasks and fifteen trials were performed for Group B, C and D tasks.

Equipment and Software

The CTS is implemented on a Commodore 64 microcomputer system. Two such systems were used, each consisting of

the following units: Commodore 64 microcomputer, Commodore 1541 disk drive, Commodore 1526 printer, monochrome Panasonic experimenter's monitor, color Commodore 1702 subject's monitor and three subject response devices.

Modifications were made to the CTS software to provide automatic sequencing through the task levels and automatic filename construction for data storage. A coding scheme was devised which included Group, Subject, Task, Level and Trial identifiers. CTS tasks and data were all stored on floppy disk with separate diskettes for each task group and each subject. On the average, two 5-1/4" data diskettes were required per subject with a total of 2500 trials (data files) for the entire study.

Subjective Workload Assessment Technique

To relate task performance to a subjective workload measure, the Subjective Workload Assessment Technique (SWAT) was used. The SWAT Scale (Reid, 1982; Reid, Eggemeier, and Nygren, 1982) is a psychometric instrument subjectively measuring three major dimensions of workload: Time, Effort, and Stress. Given the demands of any specified workload period, subjects rate each dimension on a 1 to 3 Likert-type scale. These ratings were obtained following each task trial.

Procedure

Subjects were trained on their assigned tasks during the same two hour time block on five consecutive days. Subjects in Group A sequenced through the set of tasks in the following order: MS, PM, GR, MP, UT, SP, LP, IP, and CR, yielding 25 trials during the two-hour period. Within each task, subjects sequenced through the levels from low to high in order.

All other subjects performed three trials at each level of each task. The sequence for Group B was MS, PM, GR with a total of 27 trials per day. Group C followed the sequence MP, UT, SP with 27 trials per day and Group D followed the sequence LP, IP, CR with a total of 21 trials due to the single level of the IP task. One trial at each level of each task was completed before repeating the task sequence. As in Group A, subjects sequenced through the levels from low to high in order.
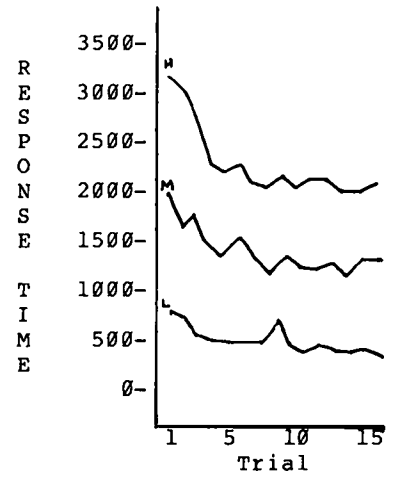
RESULTS

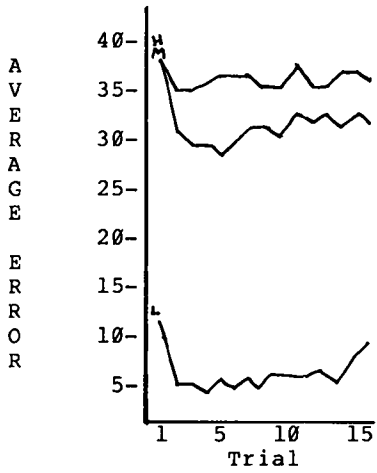Summary statistics were obtained for all trials using the CTS "STATIS-
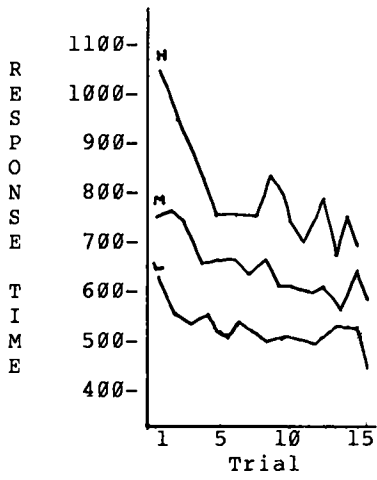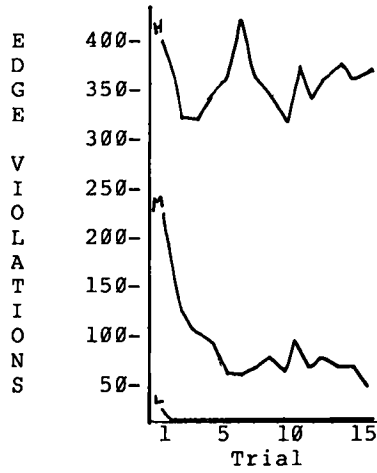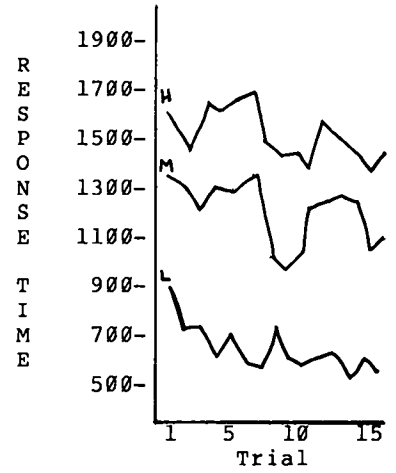
Memory Search
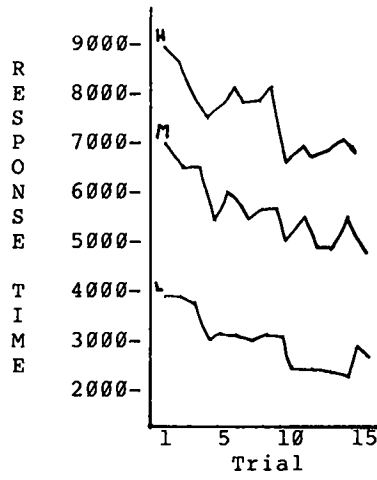
Grammatical Reasoning

Mathematical Processing
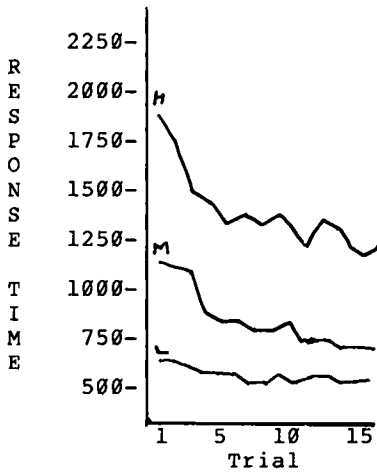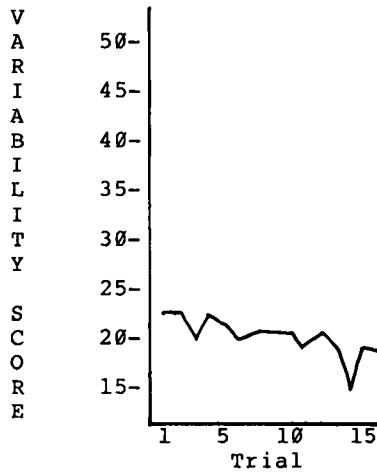

Unstable Tracking
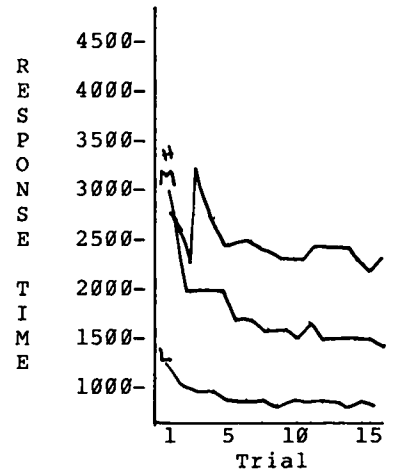

Unstable Tracking

Spatial Processing

Linguistic Processing

Interval Production

Continuous Recall

TICS" option and overall averages were plotted as a function of trial number (Figure 1). For the majority of tasks and levels, a traditional learning curve effect was observed with little difference in the shape of the curves for Group A vs. the other groups.

For approximately half of the tasks (MS, GR, MP, LP), performance for Group A was indistinguishable from that of the other Groups. For other tasks (UT, SP, IP, CR), Group A performance was worse. This may have been due to individual subject differences between the groups but is more likely a result of Group A training on a much larger number of tasks. In addition, a fatigue effect may have existed as the performance differences were worse for those tasks occurring later in the sequence. Probability Monitoring was not evaluated since this task did not provide sufficient stimuli per trial to yield a stable measure.

The SWAT ratings showed consistent ordered differences between workload levels for all tasks. Examination of the ratings also provided a comparison of the relative difficulty across tasks (Table 2). Of the tasks with three distinct workload levels, Mathematical Processing had the lowest rating ("easiest") and Continuous Recall the highest ("most difficult").

Table 2. Subjective Task Difficulty Based on SWAT.

Workload Rank		Task
Low	1	Interval Production
	2	Mathematical Processing
	3	Spatial Processing
	4	Memory Search
	5	Probability Monitoring
	6	Linguistic Processing
	7	Grammatical Reasoning
	8	Unstable Tracking
High	9	Continuous Recall

Additional analyses are currently being performed to isolate differences between groups, subjects, trials and workload levels and to gain additional insight about the factor structure of the CTS battery. The final objective of the research study involves construction of an optimal training protocol model to achieve the greatest degree of training on all tasks within limited time constraints.

REFERENCES

Reid, G.B., Subjective Workload Assessment: A Conjoint Scaling Approach, Aerospace Medical Association Annual Scientific Meeting, Bal Harbor, Florida, 153-154, May 10-13, 1982.

Reid, G.B., Eggemeier, F.T. and Nygren, T.E., "An Individual Differences Approach to SWAT Scale Development", Proceedings of the Human Factors Society 26th Annual Meeting, 639-642, October 1982.

Schlegel, R.E. and Patterson, D.W., "Evaluation of Training Performance for the USAF Criterion Task Set (CTS)," Final Report, 1984 USAF-SCEEE Summer Faculty Research Program, USAF Contract Number: F49620-82-C-0035, Southeastern Center for Electrical Engineering Education, St. Cloud, Florida, 1984.

Shingledecker, C.A., Crabtree, M.S. and Acton, W.H., "Standardized Tests for the Evaluation and Classification of Workload Metrics," Proceedings of the Human Factors Society 26th Annual Meeting, 648-651, October 1982.

Shingledecker, C.A., Acton, W.H. and Crabtree, M.S., "Development and Application of a Criterion Task Set for Workload Metric Evaluation," SAE Technical Paper 831419, Aerospace Congress & Exposition, Long Beach, California, October 1983.

Sternberg, S., "The Discovery of Processing Stages: Extension of Donders' Method," Acta Psychologica, 30, 276-315, 1969.

Wickens, C.D., Processing Resources in Attention, Dual Task Performance, and Workload Assessment, Engineering Psychology Research Laboratory Technical Report EPL-81-3, University of Illinois, 1981.

773