INDIVIDUAL DIFFERENCES IN CRITERION TASK SET PERFORMANCE[1]

Kirby Gilliland Robert Schlegel* Sharon Dannels

Department of Psychology
and
*School of Industrial Engineering
University of Oklahoma
Norman, Oklahoma 73019

ABSTRACT

As part of a larger standardization study of the U.S. Air Force Criterion Task Set (CTS), a number of individual difference variables were assessed to evaluate their relationship to CTS performance. The selection of these individual difference variables was based on their known or hypothesized relationship to performance or perceptual abilities. This paper reports some of the preliminary findings with regard to these individual difference variables. Of the variables measured, Stimulus Screening, Thrill and Adventure Seeking, Neuroticism, Type-A Behavior, and General Intelligence showed the most promising relationships to CTS performance variables.

INTRODUCTION

There is a growing awareness that human capabilities have become the limiting factor in many operational systems. Designers are now faced with situations where operational systems are limited by the information processing, decision making, memory, mental alertness, or physical capabilities of the operator.

In response to the growing awareness of these problems the U.S. Air Force has dedicated considerable effort to the exploration of workload assessment metrics. One major result of these efforts has been the development of the Criterion Task Set (CTS), a set of standardized loading tasks that can be used both to evaluate potential workload measures and to provide graded workload levels of a number of basic performance tasks for a wide range of human factors research (Shingledecker, Acton, & Crabtree, 1983).

The CTS Version 1.0 is composed of nine tasks. Eight of the tasks have three graded workload levels. These tasks include: Memory Search, Continuous Recall, Linguistic Processing, Probability Monitoring, Grammatical Reasoning, Mathematical Processing, Unstable Tracking, and Spatial Processing. The ninth task is a single-level Interval Production (tapping) task.

An important characteristic of the CTS is that it is one of the first task batteries to be developed within the framework of modern information processing theory. Thus, the CTS tasks were principally designed and selected to represent a synthesis of the primary stage (Sternberg, 1969) and multiple resource (Wickens, 1981) models associated with information processing.

In order to support widespread research interest in the CTS, a large-scale standardization and validation study was undertaken (Gilliland & Schlegel, 1985). The resulting data base from this study includes numerous performance and subjective workload measures for each of the CTS tasks collected under standard laboratory conditions, as well as, during exposure to common environmental stressors (time press, noise, sleep loss). A description of the experimental methodology and the CTS performance results of this study are reported elsewhere in these *Proceedings of the Human Factors Society 30th Annual Meeting* (See Schlegel, Gilliland, & Schlegel, 1986).

As a part of this large-scale study, subjects were assessed on a number of individual difference dimensions that have known relationships to performance efficiency and/or are hypothesized to be biologically or perceptually based. There are several individual difference variables that are known to relate to the manner in which a person processes information or to the processes directly related to performance capability (e.g., arousal state). Currently, there are scales that assess the arousal dimension (Eysenck & Eysenck, 1968), as well as related issues such as the degree of sensation seeking (Zuckerman, Kolin, Price, & Zoob, 1964). There are also scales which assess perceptual processing ability (Mehrabian, 1977; Sarason, 1972). It seemed prudent in such a foundational investigation of human performance that psychometric measures such as those described above should be included.

The purpose of this paper is to report on a preliminary analysis of the relationships between the individual difference variables and

---

[1]. This research was sponsored by the Workload and Ergonomics Branch of the Armstrong Aerospace Medical Research Laboratory, United States Air Force, under Contract F33615-82-D-0627 through the Southeastern Center for Electical Engineering Education (SCEEE-ARB/85-62). The United States Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation hereon. The authors wish to thank Gary B. Reid, Dr. F. Thomas Eggemeier, and Dr. Clark A. Shingledecker for their interest and support.

the CTS performance variables derived from this large-scale standardization and validation effort.

METHOD

As noted previously, a detailed description of the methodology used in the large-scale standardization and validation study is reported elsewhere in these *Proceedings*. Briefly, the testing protocol consisted of regularly scheduled two-hour testing sessions conducted once per day, for ten days, over a two-week cycle. Multiple workstations allowed for the simultaneous testing of four subjects during each two-hour session. Testing sessions were scheduled beginning at 8:00 a.m., 10:00 a.m., 1:00 p.m., and 3:00 p.m.

Subjects

Twenty-five men (ages 19 to 32 years; mean = 23.6 years) and twenty-five women (ages 18 to 43 years; mean = 23.0 years) served as volunteer subjects in this project. All subjects were recruited through posted announcements and were paid for their participation in the study. All subjects reported 20/20 actual or corrected vision, no history of hearing impairment, and no current use of medication.

Apparatus

The CTS Version 1.0 tasks were presented on Commodore 64 microprocessor systems with dual floppy disks and a color CRT monitor. Additional software was developed during this project to automate the presentation of CTS tasks, reduce data, and automatically label and store raw data and summary statistics files.

At each workstation the subject was provided three response controllers designed for the CTS battery at the Workload and Ergonomics Branch at Wright-Patterson Air Force Base. These consisted of a tapping button controller box for the Interval Production task, a turn-pot controller box for the Unstable Tracking task, and a four-button keypad for the remaining central processing and input/perceptual tasks.

Psychometric Tests

A battery of psychometric tests measuring individual difference dimensions that have known relationships to performance efficiency and/or are known to be biologically/perceptually based was administered to the subjects. This battery included measures of Generalized Arousal (extraversion), Neuroticism, Stimulus Screening, Sensation Seeking (and its subscale, Thrill and Adventure Seeking), Clinical (trait) Anxiety, Impulsiveness, Test Anxiety, General Intelligence, and Type A Behavior.

Generalized Arousal. The Eysenck Personality Inventory was used to assess generalized anxiety (Eysenck & Eysenck, 1968). This dimension is believed to be directly related to brainstem reticular formation activity which is subsequently reflected in different levels of cortical arousal. Introverts are hypothesized to be higher in arousal than extraverts. This arousal difference often leads to one group or the other being at a performance advantage depending on the environmental or task circumstances. Reviews of both performance and psychophysiological literature generally support this theory. This inventory also yields the Neuroticism scale score.

Stimulus Screening. Also related to the orienting reflex, as well as arousability, is the dimension of stimulus screening (Mehrabian, 1977). Stimulus screening refers to a biologically-based, perception-related dimension that reflects one's ability to screen relevant from irrelevant stimuli during information processing.

Sensation Seeking. Developed from early sensory deprivation and optimal level of arousal research, the sensation seeking scale (Zuckerman, 1979; Form V) assesses the degree to which people actively seek sensory stimuli to increase their stimulation level. This dimension has been related to orienting reflex differences (Zuckerman, 1972), and to regulators of neurotransmitters (Murphy, Belmaker, Buchsbaum, Martin, Ciaranello, & Wyatt, 1977). This scale also provides a subscale assessing the degree to which an individual actively seeks Thrill and Adventure.

Clinical (Trait) Anxiety. Clinical anxiety in a more general sense is simply termed anxiety, as opposed to more specialized forms such as test anxiety. Anxiety is known to disrupt motor performance and cognition. Anxiety is usually viewed as being either of a transient "state" form, often due to situational factors, or a more pervasive protracted "trait" form. Both trait and state anxiety were assessed with the State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970).

Impulsiveness. Impulsiveness has been shown to be related to physiological processes, specifically arousal mechanisms. Impulsiveness was measured with the Barratt Impulsiveness Scale (Barratt, 1965) which also provides subscales of Non-Planning and Motor Impulsiveness.

Test Anxiety. Test anxiety is a form of anxiety associated with demand for performance. One scale of test anxiety (Sarason, 1972) has shown negative correlations with performance efficiency, especially on vigilance and selective attention tasks.

Intelligence. While the theoretical nature of general intelligence remains controversial, this dimension has been shown to be a mediating factor in the performance of many tasks. The Wechsler Adult Intelligence Scale-

Revised (WAIS-R) was administered to all subjects in the study.

Type A Behavior. Type A Behavior refers to a specific coping style which has been linked to coronary prone behavior. This dimension is interesting for two reasons. First, it shows an apparent relationship to physiological processes, e.g., cardiovascular responses. Second, it appears to be related to highly organized, stressful, competitive, and overscheduled approaches to problem solving. This dimension was measured with the Jenkins Activity Survey (Jenkins, Zyanski, & Rosenman, 1979).

Procedure

Subjects were seated in their individual workstations facing the elevated CRT display. The controller boxes were placed on a table in front of the subjects. The workstations were separated by acoustic panels to reduce noise and subject interaction.

During the first two-hour session each subject was oriented to the project, given an introduction to each of the CTS tasks, completed a SWAT Sort (described below), and completed the battery of psychometric tests. In the second through sixth sessions, subjects were given the five training trials on the *entire CTS battery.*

The seventh and ninth sessions provided baseline experimental data collected under standard laboratory conditions, i.e., the same conditions imposed during training. Data from the eighth and tenth sessions were collected during exposure to common environmental stressors. Only performance data from the seventh session (the first baseline session) was included in the analyses reported in this paper.

During each session subjects were presented three-minute trials of each of the 25 CTS tasks (three workload levels of eight tasks, plus the one level Interval Production task). Following each trial was a brief, 1-1.5 minute rest period during which data was stored on the diskette, the next task was prepared for presentation, and subjects completed subjective workload assessments. The total testing time per session was approximately one hour and forty-five minutes.

Throughout the study, subjects were asked to provide subjective assessments of the workload presented by the various CTS tasks by the use of the Subjective Workload Assessment Technique (SWAT). The SWAT Scale (Reid, 1982; Reid, Eggemeier, & Nygren, 1982) is a psychometric instrument for measuring subjective ratings on three major dimensions of workload: Time Pressure, Mental Effort, and Psychological Stress. The unique aspect of SWAT is that it not only provides a means for obtaining an individual subject's workload ratings, but is also a method for establishing cross-subject comparability.

RESULTS

Table 1 presents the preliminary correlational analyses of the individual difference variables with a selected group of CTS task performance variables. Only correlations significant at or below the p =.05 level of significance are listed in Table 1.

It is important to note that these general results are quite preliminary and based on a relatively low number of subjects (N=50). Nonetheless, these data do provide some encouraging information and yield some interesting trends which will be analyzed and reported more fully when collection of the full standardization and validation data set is completed.

The individual difference variables included in Table 1 have been described previously. The CTS Performance measures included in this first summary overview were: Mean Response Time (Mean RT) and Percent Correct (% Corr) for Memory Search (MS), Continuous Recall (CR), Linguistic Processing (LP), Grammatical Reasoning (GR), Mathematical Processing (MP), and Spatial Processing (SP), Interval Production (IP) variability score (var), and Unstable Tracking (UT) mean absolute error (abs error) and edge violations (edge viol).

DISCUSSION

Two trends emerged in the correlational matrix which are important. First, there is a group of individual difference variables that seem to have significant relationships with multiple CTS tasks. Second, five of the CTS tasks seem to have significant relationships with various individual difference measures.

The individual difference variables that seemed to show the strongest relationship to a number of CTS performance variables were: Stimulus Screening, Neuroticism, Thrill and Adventure Seeking, Intelligence, and Type A Behavior.

Stimulus Screening was negatively related to the mean response time of Memory Search, Continuous Recall, Linguistic Processing, Grammatical Reasoning, Mathematical Processing, Unstable Tracking, and Spatial Processing. These results suggest that the higher a person scores on this scale, that is, the more they show the capability to screen out irrelevant stimuli during information processing, the greater the likelihood that they will perform more rapidly on a number of CTS tasks that draw upon a fairly wide range of abilities. However, in only one case was accuracy of responding related to Stimulus Screening. It appears that the higher one scores on Stimulus Screening more likely one is to be correct in the Grammatical Reasoning Task. Screeners were also less likely to make Edge Violations in the Unstable Tracking task. It should also be noted that these relationships did not hold for all workload levels.

Table 1. Correlations Between Selected Individual Difference and CTS Performance Variables*.

INDIVIDUAL DIFFERENCE VARIABLE	WORK LOAD	TASK																	
		MS		IP	CR		LP		GR		MP		UT		SP	
		mean RT	% corr	var	mean RT	% corr	mean RT	% corr	mean RT	% corr	mean RT	% corr	abs err	edge viol	mean RT	% corr
Extraversion	L															
	M		.35													
	H															
Neuroticism	L	.32					.30									
	M									-.29	.37			.35		
	H										.29			.30		
Stimulus Screening	L	-.34			-.33				-.31							
	M						-.28				-.36			-.29		
	H									.31	-.42				-.29	
Sensation Seeking	L						-.28									
	M						-.32									
	H															
Thrill and Adventure Seeking	L		.38				-.41								-.34	
	M		.37				-.33									
	H		.29													.37
Trait Anxiety	L						.29									
	M										.33					
	H										.31					
Impulsiveness	L														.29	
	M															
	H						.29							.32		
Non-Planning Impulsiveness	L															
	M															
	H									-.29				.28		
Motor Impulsiveness	L						.37								.41	
	M															
	H															
Test Anxiety	L															
	M															
	H															
WAIS	L	-.49			-.50				-.54	.69						
	M	-.31	.31			.37	-.48	.40	-.36	.67	-.43				-.35	
	H		.34				-.49	.61		.67	-.42					
WAIS Verbal	L	-.36			-.35				-.41	.61						
	M						-.33	.32		.58	-.31					
	H		.32				-.34	.57		.63	-.32					
WAIS Performance	L	-.62			-.62		-.56		-.65	.68	-.42					
	M	-.46	.36			.51	-.59	.45	-.51	.67	-.51				-.27	
	H	-.32	.31			.36	-.60	.56		.58	-.46					.34
Type-A Behavior	L										-.28				-.31	
	M														-.35	
	H										-.30				-.32	

*Correlations listed in the table were those found to be significant at least at the p=0.05 level.

Neuroticism was also significantly related to several CTS performance variables. Neuroticism, it should be noted, is best viewed as a dimension of emotional stability rather than an index of neurotic behavior in the strict clinical sense. Those individuals who scored higher in Neuroticism (i.e., those more emotionally unstable) were also more likely to respond more slowly in Memory Search, Linguistic Processing, and Math Processing. They were also more likely to commit Edge Violations in Unstable Tracking and respond incorrectly more often in Grammatical Reasoning. Again, these relationships did not hold for all workload levels. Because Stimulus Screening and Neuroticism were found to be correlated $r = -.67$ ($p < .0001$), it should not be surprising to find them having opposite patterns of correlations with CTS performance variables.

Thrill and Adventure Seeking, a subscale of the Sensation Seeking Scale, was also related to several CTS performance variables. It appears that the more one seeks thrill and adventure, the more likely one is to respond

quickly to the Linguistic Processing and Spatial Processing tasks. These individuals are also more likely to be correct in Memory Search and Spatial Processing. Again, these results do not hold at all workload levels for all tasks.

Intelligence was clearly the variable most related to CTS task performance. The more intelligent one is, as measured by the WAIS, the more likely one is to respond rapidly to the Memory Search, Continuous Recall, Linguistic Processing, Mathematical Processing, Spatial Processing, and Grammatical Reasoning tasks. More intelligent subjects were also more likely to be correct on the Memory Search, Continuous Recall, Linguistic Processing, and Grammatical Reasoning tasks, but not on the Mathematical and Spatial Processing tasks. In most cases, these findings generalized to more than one workload level and in some cases to all workload levels. Also, it is interesting to note that motor output tasks (Interval Production and Unstable Tracking) seem to be fairly independent of intelligence.

There were other significant relationships of interest. However these relationships showed less consistency and therefore may be less reliable. For example, Type A behavior seemed to be associated with an unusual combination of speed in Mathematical Processing and Spatial Processing. Also, Edge Violations in Unstable Tracking seemed to be related to Non-planning Impulsiveness (rather than Motor Impulsiveness, as one might predict), and Neuroticism.

It is also interesting to note that some CTS tasks correlated more frequently with the individual difference measures than others. Linguistic Processing and Grammatical Reasoning had the highest number and most complete pattern of correlations between both speed and accuracy measures and intelligence score--undoubtedly due to the emphasis on verbal fluency and facility in these activities. However, Linguistic Processing was related to several other individual difference dimensions while Grammatical Reasoning was significantly related to only a few others. It is also interesting to note that Interval Production appears to be unrelated to any of the individual differences dimensions--despite literature linking some of these variables to differences in rhythmic behavior or task performance (e.g., Extraversion, Impulsiveness).

It is clear that enough relationships of interest exist to warrant more extensive analyses of these data when a larger sample size is obtained. The larger sample size will allow more power for the analyses, as well as provide the opportunity to perform more sophisticated multivariate analyses. In addition, the completed standardization data set will allow analyses of the relationship between individual difference variables and CTS performance variables across environmental stress conditions, i.e., cross-situational paradigm research. Finally, the completed data set will also provide the opportunity to analyze the relationship between individual difference variables and subjective workload assessments.

REFERENCES

Barratt, E. (1965). Factor analysis of some psychometric measures of impulsiveness and anxiety. *Psychological Reports, 16*, 547-554.

Eysenck, H.J. & Eysenck, S.B.G. (1968). *Manual for the Eysenck Personality Inventory*. Educational and Industrial Testing Service, San Diego, California.

Gilliland, K. & Schlegel, R.E. (1985). Evaluation of the Criterion Task Set for drug and stress research. USAF Contract F33615-82-D-0627 through Southeastern Center for Electrical Engineering Education (SCEEE-ARB/85-62).

Jenkins, C., Zyzanski, S., & Rosenman, R. (1979). *Jenkins Activity Survey Manual*. The Psychological Corporation, New York, New York.

Mehrabian, A. (1976). *Manual for the Questionnaire Measure of Stimulus Screening*. Unpublished manuscript. Available from Albert Mehrabian, Los Angeles, California.

Murphy, D.L., Belmaker, R.H., Buchsbaum, M., Martin, N.F., Ciaranello, W., & Wyatt, R.J. (1977). Biogenic amine-related enzymes and personality variations in normals. *Psychological Medicine*, 7, 149-157.

Reid, G.B. (1982). Subjective workload assessment: A conjoint scaling approach. In the *Proceedings of the Aerospace Medical Association Annual Scientific Meeting*, pp. 153-154, Bal Harbor, Florida.

Reid, G.B., Eggemeier, F.T., & Nygren, T.E. (1982). An individual differences approach to SWAT Scale development. In the *Proceedings of the Human Factors Society 26th Annual Meeting*, pp.639-642, Seattle, Washington.

Sarason, I.G. (1972). Experimental Approaches to Test Anxiety: Attention and the Uses of Information. In Spielberger, C.D. (Ed.), *Anxiety: Current trends in theory and research*, Vol. II. New York: Academic Press.

Schlegel, R.E., Gilliland, K., & Schlegel, B. (1986). Development of the Criterion Task Set performance data base. In the *Proceedings of the Human Factors Society 30th Annual Meeting*, Dayton, Ohio.

Shingledecker, C.A., Acton, W.H., & Crabtree, M.S. (1983) *Development and application of a Criterion Task Set for workload metric evaluation*. SAE Technical Paper 831419. Aerospace Congree & Exposition, Long Beach, California.

Spielberger, C., Gorsuch, R., & Lushene, R. (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologist Press, Palo Alto, California.

Sternberg, S. (1969). *Scanning mental processes revealed by reaction time experiments*. *American Scientist*, 57, 421-457.

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

Zuckerman, M. (1979). *Sensation Seeking: Beyond the Optimal Level of Arousal*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Zuckerman, M., Kolin, E.A., Price, L., & Zoob, I. (1964). Development of a Sensation Seeking Scale. *Journal of Consulting Psychology*, 28, 477-482.