FACTOR STRUCTURE OF THE CRITERION TASK SET

Robert E. Schlegel Kirby Gilliland* Betina Schlegel

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

ABSTRACT

A large-scale experimental study was conducted involving the training and testing of 123 human subjects on the Criterion Task Set (Version 1.0). Testing was performed under baseline and stressor conditions. The performance data and Subjective Workload Assessment Technique ratings for the first baseline trial (Trial 6) were analyzed using the SAS VARCLUS procedure to evaluate the structure of the CTS. Seven clusters of response time variables were identified for the nine tasks. In general, the Memory Search, Linguistic Processing and Mathematical Processing tasks were grouped in one cluster with each of the other clusters representing a single task. Five clusters were identified for the SWAT ratings with clusters differentiated along the dimensions of task difficulty and processing stage.

INTRODUCTION

From its initial development in 1983 and subsequent release in 1984, the Criterion Task Set (CTS) has been widely disseminated as a research tool for evaluating workload assessment techniques and as an instrument for human performance assessment. Literature exists concerning the content of the task battery and the development of the individual tasks (Shingledecker, 1984), training characteristics of the battery (Schlegel and Shingledecker, 1985), and preliminary results from the development of a large-scale data base of CTS performance data, Subjective Workload Assessment Technique (SWAT) ratings and individual difference variables (Schlegel et al., 1986; Gilliland et al., 1986).

A major selling point of the CTS is its selection of tasks based on processing stage and multiple resource theories. Table 1 summarizes the characteristics of the nine tasks (25 individual task levels) presented in alphabetic order.

Data collection has recently been completed for the large-scale standardization study of the CTS. Performance and Subjective Workload Assessment Technique (SWAT) data were collected for approximately 123 subjects (96 males, 27 females) performing all nine tasks of the CTS at all levels for nine days. Five days were allocated for training, two days for baseline testing and two days for testing under various stressors. Details of the methodology were reported in Schlegel, Gilliland and Schlegel (1986).

In addition to providing standardization data for the CTS Version 1.0 tasks, the study provided an opportunity to examine the relationships among the dependent measures for the nine tasks and the SWAT ratings for the tasks. Specifically, data from the first baseline trial following training (Trial 6) was analyzed using the Statistical Analysis System VARCLUS procedure to cluster the nine tasks and twenty-five individual task levels. The results of these analyses will help determine if all levels of a particular CTS task are drawing from the same resource pool and how much overlap in resource demands exists between tasks.

THE SAS VARCLUS PROCEDURE

Several techniques may be employed to determine the underlying structure of a set of measures. These include factor analysis, principle components and multidimensional scaling techniques. Another approach is clustering analysis. The SAS VARCLUS procedure divides a set of numeric variables into either disjoint or hierarchical clusters in such a way that each cluster can be interpreted as essentially unidimensional (SAS, 1985). The clusters are chosen to maximize the variation accounted for by either the first principal component or the centroid component of each cluster. Specifically, the procedure attempts to maximize the sum across clusters of the variance of the original variables that is explained by the cluster components.

Downloaded from pro.sagepub.com at UNIV OF OKLAHOMA LIBRARIES on January 20, 2016

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/25467

Table 1. Summary of CTS Task Characteristics.

Task	Code	Stage	Level	Description
Continuous Recall	CR	Central		continuous update memory
			1	1 digit - next item
			2	2 digits - 2 items back
			3	4 digits - 3 items back
Grammatical Reasoning	GR	Central		logic; reasoning
			1	single sentence
			2	two sentences - active/ positive
			3	two sentences - active/ negative or passive/negative
Interval Production	IP	Output	1	tapping at 2 taps/sec.
Linguistic Processing	LP	Central		symbol manipulation
			1	physical identity match
			2	vowel/consonant match
			3	antonym match
Mathematical Processing	MP	Central		math operations (+ or -)
			1	1 operation
			2	2 operations
			3	3 operations
Memory Search	MS	Central		Sternberg memory test
			1	positive set size 1
			2	positive set size 4
			3	positive set size 6
Probability Monitoring	PM	Input		scanning and detection
			1	1 meter, 95% bias
			2	3 meters, 85% bias
			3	4 meters, 75% bias
Spatial Processing	SP	Central		histogram shapes
			1	2 bars, Ø rotation
			2	4 bars, 9Ø or 27Ø rotation
			3	6 bars, 18Ø rotation
Unstable Tracking	UT	Input/Output		manual response
			1	lambda=1
			2	lambda=3
			3	lambda=5

VARCLUS is a type of oblique component analysis related to multiple group factor analysis. It can be used as a variable reduction method to replace a large set of variables by the set of cluster components. A given number of cluster components does not typically explain as much variance as the same number of principal components, but the cluster components are usually easier to interpret than the principal components even after rotation.

RESULTS

Four separate analyses were performed, three involving the performance data and one with the SWAT data.

CTS Performance Data

The first clustering analysis included the response time measures for the discrete stimulus tasks and probability monitoring plus the mean tapping rate (IPMN) and variability score (IPVS) for interval production and the absolute error and edge violations for unstable

Downloaded from pro.sagepub.com at UNIV OF OKLAHOMA LIBRARIES on January 20, 2016

tracking. This provided nine variables for each of the three levels of eight of the tasks plus two variables for Interval Production for a total of 29 performance measures. Each of the task difficulty levels was included separately in order to determine whether the different levels tap the same resource as indicated by the clustering. A summary of the clusters generated from this analysis is given in Table 2.

Table 2. Cluster Analysis for Response Time Measures.

	Cluster					
1	2	3	4	5	6	7
CR1	CR2	GR1	IPMN	PM1	SP1	All
	CR3	GR2	IPVS	PM2	SP2	UT
		GR3		PM3		Var
LP1						
LP2						
LP3						
MP1						
MP2						
MP3						
MS1						
MS2						
MS3						

1 - low level 2 - medium level
3 - high level

Several significant findings are evident from the table. With one major exception, measures from different levels of the same task were placed in the same cluster indicating that the various workload levels or difficulty manipulations of any given task draw from the same resource pool. The exception is the Continuous Recall task at the low level which was placed in the largest cluster with linguistic processing, mathematical processing and memory search. This indicates the much lower difficulty of CR at the low level and its closer association with LP and MP as a symbol manipulation task.

Each task occupied a separate cluster with the exception of LP, MP and MS. This indicates the minimal resource overlap for all tasks except these three. The overlap of these three is probably due to the relative ease of the tasks and the similarities of simple symbol manipulation whether linguistic, mathematical or simple memory update.

The second clustering analysis examined the accuracy measures of proportion correct in place of the response time measures for the discrete stimulus tasks. Performance measures for the non-central processing tasks were not included in the analysis. The clustering exhibited more task overlap (Table 3) than in the analysis of response times. This was probably due to the relatively high level of accuracy in several tasks.

Table 3. Cluster Analysis for Proportion Correct Measures.

	Cluster			
1	2	3	4	5
CR1	CR2	GR1	MS1	SP1
	CR3	GR2	MS2	SP2
MP1		GR3	MS3	SP3
MP2		LP2	LP1	
MP3		LP3		

1 - low level 2 - medium level
3 - hign level

A final clustering analysis with the performance data combined the response time and accuracy measures for a total of 47 variables. Ten clusters were defined as shown in Table 4. It is obvious that the Probability Monitoring measures do not form a separate cluster nor do they clearly associate with other clusters. With minor exceptions, the other measures form logical clusters along the lines of the first two analyses.

SWAT Ratings

The SWAT ratings reflect the perceived difficulty of the task rather than the actual performance. Cluster analysis of the SWAT ratings for the 25 task levels produced the results given in Table 5.

In terms of subjective ratings, the tasks and task levels were clustered along the dimensions of task difficulty and processing stage. It is evident from the table that cluster 1 contains discrete stimulus central processing tasks of low and moderate difficulty while cluster 2 contains the more difficult central processing tasks. Cluster 3 contains the easy levels of the motor output tasks, a spatial task and a math

Downloaded from pro.sagepub.com at UNIV OF OKLAHOMA LIBRARIES on January 20, 2016

Table 4. Cluster Analysis for Performance Measures.

Cluster									
1	2	3	4	5	6	7	8	9	10
CRMN1	CRMN2	CRPC1				IPMN	SPMN1	SPPC1	All
	CRMN3					IPVS	SPMN2	SPPC2	UT
LPMN1	CRPC2			LPPC1			SPMN3	SPPC3	Vars
LPMN2	CRPC3				LPPC2				
LPMN3					LPPC3				
MPMN1		MPPC1							
MPMN2		MPPC2							
MPMN3		MPPC3							
MSMN1				MSPC1					
MSMN2				MSPC2					
MSMN3				MSPC3					
GRPC1			GRMN1						
GRPC2			GRMN2						
		GRPC3	GRMN3						
		PMRT3			PMRT2	PMRT1			

1 - low level 2 - medium level 3 - high level

processing task. Cluster 4 contains the difficult levels of the visual/spatial tasks and cluster 5 contains the difficult levels of the motor output tasks.

Table 5. Cluster Analysis of SWAT Ratings.

Cluster				
1	2	3	4	5
CR1	CR2	UT1	PM1	UT2
	CR3	IP1	PM2	UT3
GR1	GR2		PM3	
	GR3	SP1	SP2	
MP2		MP1	SP3	
MP3				
LP1				
LP2				
LP3				
MS1				
MS2				
MS3				

1 - low level 2 - medium level 3 - high level

DISCUSSION

The results of the cluster analysis help to validate the design goals of the Criterion Task Set as a battery of tasks that tap separate information processing resources and stages. The primary area of concern is the overlap among the Linguistic Processing, Mathematical Processing and Memory Search tasks, all of which involve somewhat simple symbol manipulation. Some differences existed between the cluster structure for the performance data and that for the SWAT ratings indicating that subjects perform differently than their estimate of task difficulty as might be expected. Additional analyses using different techniques are being performed to substantiate these results.

ACKNOWLEDGEMENT

This research was sponsored by the Workload and Ergonomics Branch of the Armstrong Aerospace Medical Research Laboratory, United States Air Force, under Contract F33615-85-D-0514 through the Southeastern Center for Electrical Engineering Education (SCEEE-HER/86-9). The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon. The authors wish to thank Gary B. Reid and Dr. F. Thomas Eggemeier for their interest and support.

Downloaded from pro.sagepub.com at UNIV OF OKLAHOMA LIBRARIES on January 20, 2016

REFERENCES

Gilliland, K., Schlegel, R.E. and Dannels, S.A. (1986). Individual Differences in Criterion Task Set Performance. In **Proceedings of the Human Factors Society 30th Annual Meeting.** (pp. 64-68). Dayton, Ohio: Human Factors Society.

SAS User's Guide: Statistics, Version 5 Edition. (1985). (pp. 801-816). Cary, North Carolina: SAS Institute Inc.

Schlegel, R.E. and Shingledecker, C.A. (1985). Training Characteristics of the Criterion Task Set Workload Assessment Battery. In **Proceedings of the Human Factors Society 29th Annual Meeting** (pp. 770-773). Baltimore, Maryland: Human Factors Society.

Schlegel, R.E., Gilliland, K. and Schlegel, B. (1986). Development of the Criterion Task Set Performance Data Base. In **Proceedings of the Human Factors Society 30th Annual Meeting.** (pp. 58-62). Dayton, Ohio: Human Factors Society.

Shingledecker. C.A. (1984). **A Task Battery for Applied Human Performance Assessment Research** (Tech. Report AFAMRL-TR-84-071). Wright-Patterson AFB, Ohio: Medical Research Laboratory.

Downloaded from pro.sagepub.com at UNIV OF OKLAHOMA LIBRARIES on January 20, 2016