AFHRL-TP-88-73

(2)

AIR FORCE HUMAN RESOURCES LABORATORY

ASCII CODAP PROGRAMS FOR SELECTING AND INTERPRETING TASK CLUSTERS

William J. Phalen

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Michael R. Staley

Intelligent Systems Design
NCR Center, Suite 101
15400 S.E. 30th Place
Bellevue, Washington 98007

Jimmy L. Mitchell

McDonnell Douglas Astronautics Company
P.O. Box 516
St. Louis, Missouri 63166

October 1989

Interim Technical Paper for Period October 1987 - December 1988

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

072

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

DANIEL L. LEIGHTON, Colonel, USAF
Chief, Manpower and Personnel Division

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFHRL-TP-88-73	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Manpower and Personnel Division 6b. OFFICE SYMBOL (*If applicable*) AFHRL/MO	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (*City, State, and ZIP Code*) Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235-5601	7b. ADDRESS (*City, State, and ZIP Code*)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory 8b. OFFICE SYMBOL (*If applicable*) HQ AFHRL	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (*City, State, and ZIP Code*) Brooks Air Force Base, Texas 78235-5601	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 62205F; PROJECT NO. 7719; TASK NO 20; WORK UNIT ACCESSION NO. 14

11. TITLE (*Include Security Classification*)
ASCII CODAP Programs for Selecting and Interpreting Task Clusters

12. PERSONAL AUTHOR(S)
Phalen, W.J.; Staley, M.R.; Mitchell, J.L.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (*Year, Month, Day*)	15. PAGE COUNT
Interim	FROM Oct 87 TO Dec 88	October 1989	14

16. SUPPLEMENTARY NOTATION
Paper presented at the 30th Annual Conference of the Military Testing Association, Arlington, Virginia, 27 November - 2 December 1988.

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
05	08	
05	09	

18. SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)
CODAP; occupational analysis; job analysis; task clustering; job typing

19. ABSTRACT (*Continue on reverse if necessary and identify by block number*)

In the process of developing an operational revision of the CODAP system, several major new programs were created to extend the capabilities of the system for identifying and assisting analysts in interpreting potentially significant jobs (groups of similar cases) and task modules (groups of co-performed tasks). The four major programs are CASSET, which identifies the cases that best represent and discriminate among potential job types; CORTAS, which highlights the tasks most characteristic of potential job types; TASSET, which identifies those tasks that best represent and discriminate among task modules (clusters); and CORCAS, which highlights the cases that are most characteristic of potential task modules (clusters). Used with additional supporting programs, these four primary cluster interpretation programs should greatly expedite and improve the typical occupational analysis of Air Force specialties. Additionally, experimental programs are now available to examine the joint results of both task and case clustering results simultaneously. Such techniques hold considerable promise for making better use of analysts' time by focusing the analysis on key case and task groupings on the basis of predefined analysis criteria.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Nancy J. Allin, Chief, STINFO Office	22b. TELEPHONE (*Include Area Code*) (512) 536-3877 22c. OFFICE SYMBOL AFHRL/PRT

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

AFHRL Technical Paper 88-73

October 1989

ASCII CODAP PROGRAMS FOR SELECTING AND INTERPRETING TASK CLUSTERS

William J. Phalen

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Michael R. Staley

Intelligent Systems Design
NCR Center, Suite 101
15400 S.E. 30th Place
Bellevue, Washington 98007

Jimmy L. Mitchell

McDonnell Douglas Astronautics Company
P.O. Box 516
St. Louis, Missouri 63166

Reviewed and submitted for publication by

Lawrence O. Short, Lt Col, USAF
Chief, MPT Technology Branch

Paper presented at the 30th Annual Conference of the Military Testing Association, Arlington, Virginia, 27 November - 2 December 1988.

SUMMARY

In the process of developing an operational revision of the CODAP system, several major new programs were created to extend the capabilities of the system for identifying and assisting analysts in interpreting potentially significant jobs (groups of similar cases) and task modules (groups of co-performed tasks). The four major programs are CASSET, which identifies the cases that best represent and discriminate among potential job types; CORTAS, which highlights the tasks most characteristic of potential job types; TASSET, which identifies those tasks that best represent and discriminate among task modules (clusters); and CORCAS, which highlights the cases that are most characteristic of potential task modules (clusters). Used with additional supporting programs, these four primary cluster interpretation programs should greatly expedite and improve the typical occupational analysis of Air Force specialties. Additionally, experimental programs are now available to examine the joint results of both task and case clustering results simultaneously. Such techniques hold considerable promise for making better use of analysts' time by focusing the analysis on key case and task groupings on the basis of predefined analysis criteria.

PREFACE

This work was completed under Work Unit 77192014, Research in Manpower and Personnel Technologies, Advanced and Exploratory Development of Occupational Measurement Technology. This paper was presented at the 30th Annual Conference of the Military Testing Association and published in the proceedings of that event.

TABLE OF CONTENTS

ASCII CODAP PROGRAMS FOR SELECTING AND INTERPRETING TASK CLUSTERS

I. INTRODUCTION

Even though revision of the Comprehensive Occupational Data Analysis Programs (CODAP) has been completed and ASCII CODAP implemented for operational use, advanced development continues--particularly in the areas of new techniques and automated support systems to assist job analysts in developing job and task clusters and to expedite making analysis decisions. Within the past 2 years, considerable progress has been made in developing and validating new technologies for selecting, refining, and interpreting job and task clusters.

This paper will focus entirely on new technologies for defining and interpreting task clusters, because this technology is less well known and understood, even though it has become an increasingly important component of major Air Force Manpower-Personnel-Training (MPT) research and development (R&D) programs such as the Training Decisions System (TDS), the Advanced On-the-job Training System (AOTS), and the Job Performance Measurement (JPM) System, to name but a few. Some basic questions asked by these research efforts have been: What tasks are co-performed? What tasks should be trained together and do co-performance measures accurately and uniformly identify these tasks? How can we meaningfully define a job in terms of a relatively small number of descriptive components, rather than the 500 to 1,500 task statements routinely contained in an occupational survey? Every one of these questions involves the clustering of tasks on an appropriate criterion. A major contribution of the new ASCII CODAP system has been to offer new ways of clustering tasks and interpreting task clusters.

II. PHASE I: ADAPTATION OF EXISTING SOFTWARE

Several new approaches to clustering tasks--most notably, semantic clustering procedures--are under consideration and experimentation, but the currently preferred approach has been to cluster tasks on a measure called "co-performance." Co-performance can and does mean many things in the ASCII CODAP system. In general terms, it refers to measuring the commonality of pairs of task profiles across all the cases in a survey sample. The measure can take on various forms for a pair of tasks; e.g., tasks "A" and "B," which have values across "i" cases as shown in the four examples below:

$$\text{OVERLAP}_{AB} = \frac{1}{2}\left[\frac{\sum_i \text{MIN}\,(A_i, B_i)}{\sum_i A_i} + \frac{\sum_i \text{MIN}\,(A_i, B_i)}{\sum_i B_i}\right] 100 \qquad (1)$$

$$\text{OVERLAP}_{AB} = \left[\frac{\sum_i \text{MIN}\,(A_i, B_i)}{\sum_i A_i + \sum_i B_i - \sum_i \text{MIN}\,(A_i, B_i)}\right] 100 \qquad (2)$$

$$\text{OVERLAP}_{AB} = \left[\frac{2\sum_i \text{MIN}\,(A_i, B_i)}{\sum_i A_i + \sum_i B_i}\right] 100 \qquad (3)$$

$$\text{OVERLAP}_{AB} = \left[\frac{\sum_i \text{MIN}\,(A_i, B_i)}{\sqrt{(\sum_i A_i)\,(\sum_i B_i)}}\right] 100 \qquad (4)$$

The values of "A_i" and "B_i" may be the raw task ratings, or the percent time spent values, or simple, dichotomous "do/don't-do" values compared and summed across all cases (N_i) in the sample. When the measure is percent time spent, the A_i and B_i values will sum to 100% and all four measures shown will reduce to the more simplified expression: $\Sigma_i[MIN(A_i, B_i)]$. Many other co-performance measures are available in the ASCII CODAP system, but the four listed here suggest some of the possibilities.

In actuality, the procedure of choice currently has been to convert the task vectors (rows) to percent time spent values after the case vectors (columns) have been converted to percent time spent values by the input standard (INPSTD) program and, then, to use the $_i[MIN(A_i,B_i)]$ overlap option to compute the co-performance between all possible pairs of tasks. Because the overlap (OVRLAP) program works only with columns, this procedure requires front-end application of a program called "XPOSE" to transpose the case-by-task matrix into a task-by-case matrix prior to computing pairwise overlap values. The grouping procedure used for task clustering is a form of average linkage identical to that used for collapsing the overlap matrix when cases are clustered. The resultant hierarchical clustering solution can then be displayed by the DIAGRM (diagram) program.

Other off-the-shelf software also proved useful for displaying the clustering results. The PRTVAR (print variables) and PRTFAC (print factors) programs were used to print the task titles in TPATH order, a product that mimicked the PRTVAR presentation of job titles in KPATH order for cases. The PRTVAR and PRTFAC reports of tasks could also display any other available task variables, such as percent time spent on each task by the total sample, mean task learning difficulty, and mean recommended training emphasis ratings for each task, etc. PRTVAR was also used to provide a product conveying powerful visual effects; namely, a simultaneous display of a case clustering and a task clustering. As in a standard PRTVAR report, cases would be shown in KPATH order as rows; tasks would be shown in TPATH order as columns (just like background variables). The data items within the cells of this matrix-like report would be the raw task ratings (1 through 9 on a relative time spent scale). The visual effect was that of blocks of ratings representing homogeneous clusters of cases (job types) performing homogeneous clusters of tasks (task co-performance modules). When the two types of clustering are displayed simultaneously, each tends to highlight the less obvious breakpoints in the other clustering solution. A planned improvement to enhance the visual effect of the blocks of ratings is to use symbols with varying levels of print intensity to replace the 1- through 9-point ratings as the elements of the matrix.

III. PHASE II: DEVELOPMENT OF NEW INTERPRETIVE SOFTWARE

To interpret the task clusters derived by analysis of the hierarchical clustering solution, two new programs called "TASSET" (task set) and "CORCAS" (core cases) were developed. These two programs completed a set of four CORSET (sets of core cases or tasks) programs for analyzing and interpreting job and task clusters. Figure 1 shows the interrelationships among the four programs.

	Task Clusters (Task Modules)	Case Clusters (Job Types)
Core Tasks	TASSET	CORTAS
Core Cases	CORCAS	CASSET

Figure 1. The CORSET Programs.

Each diagonal has its own naming convention. The TASSET and CASSET programs on the principal diagonal have to do with defining core tasks for task clusters and core cases for case clusters, respectively; whereas the CORTAS and CORCAS programs on the other diagonal are concerned with defining core tasks for job types and core cases for task modules, respectively. This paper will discuss only the TASSET and CORCAS programs, which are used to interpret task clusters.

TASSET Report

Once task clusters have been identified as being distinctly different from one another, the focus moves to pinpointing the nature of the differences. This is the purpose of the TASSET report. It is such a complex report and contains so many items of information that we will only be able to touch upon some of the more important items in this paper.

Supergroup/Subgroup Matrix. This part of the TASSET report is an asymmetric matrix of percentage values indicating the degree to which each cluster of tasks is co-performed with every other task cluster. For example, if task cluster "A" is performed, what is the probability that task cluster "B" is performed? And if task cluster "B" is performed, what is the probability that task cluster "A" is performed? From this matrix we can ascertain whether "A" subsumes "B," or "B" subsumes "A," or whether they subsume each other and thus ought to be merged into a single group.

Average Co-Performance. We now look at each cluster separately. Within each cluster, TASSET computes the average co-performance of each task with every other task in the cluster. For example, if a cluster consists of tasks A, B, C, and D, the average co-performance of task "A" would be the average of the co-performances of "A" with "B," "A" with "C," and "A" with "D." This computation is done for each task in the cluster. The tasks are then sorted from high to low on these values. The significance of this measure is that it shows which tasks are most representative of the cluster and which are least representative. Such information is valuable in determining the dominant characteristics of the cluster and in giving the cluster a name.

Task Co-Performance Discrimination. TASSET computes a discrimination value for each task in a cluster as a measure of how well each task fits in that cluster compared to how well it might have fit, on the average, if it had been placed in each of the other selected task clusters. By this procedure, we can compare task clusters that consist of distinctly different sets of tasks. For example, we have three task clusters: #1 contains tasks A, B, C; #2 contains tasks D, E, F; and #3 contains tasks G, H, I. To compare these three clusters, we would compute the discrimination of task "A" in cluster #1 by evaluating its average co-performance with the tasks in cluster #1 relative to its average co-performance with the tasks in cluster #2 and cluster #3; i.e., assuming task "A" to have been placed in each of the three clusters. If the average co-performance of task "A" is 80% within its own cluster (#1), 40% within cluster #2, and 20% within cluster #3, its discrimination value as a member of cluster #1 would be computed as:

$$DISC_A = \left[\frac{\sqrt{80-40} + \sqrt{80-20}}{2}\right]^2 = 49.5\% \qquad (5)$$

A discrimination value of 49.5% indicates that, on the average, task "A" has a 49.5% higher co-performance with cluster #1 than with #2 or #3. Provision is made for negative discriminations, as well. As a rule of thumb, the cutoff to identify a discriminating task is 25%. Thus, task "A" serves well to discriminate task cluster #1 from task clusters #2 and #3. When TASSET sorts the tasks within a cluster from high to low based on their discrimination values, a clearer picture emerges of the cluster's unique characteristics. This further helps identify the salient characteristics of the task cluster for describing and naming it.

Potential Core Tasks. TASSET lists a set of potential core tasks for each task cluster. Although these are tasks which grouped into other task clusters, they had average co-performance values for the target cluster that were at least as high as one of the tasks actually placed in the target cluster. TASSET also lists the identification numbers (IDs) of the clusters in which the potential core tasks were placed. A large number of potential core tasks with the same cluster ID would indicate a high degree of commonality between this cluster and the target cluster.

Uniquely Co-Performed Tasks. Sometimes "uniquely co-performed tasks" are reported by TASSET for a task cluster. These are tasks which are performed by so few job incumbents that they do not appear in any task cluster; instead, they are listed at the end of the report unique to a particular cluster if all or most of their co-performance is associated with the tasks in that particular cluster. For example, "Compute complex statistics, such as correlation coefficients and standard deviations" was a task performed by only 1.59% of a sample of clerical workers, but it appeared as a uniquely co-performed task for a cluster of tasks having to do with "compilation of information."

Negatively Unique Tasks. Negatively unique tasks are reported by TASSET for a target task cluster when there are tasks which have high co-performance values for most other task clusters (because they are performed by a high percentage of all workers in the occupation) but are rarely co-performed with tasks in the target cluster. For example, clerical tasks related to "dealing with the public" appear as negatively unique tasks for task clusters totally devoted to "transcribing" or "filing."

Various other kinds of task-related information are available in a TASSET report, but the several described here give a good idea of the value of the report for detailed analysis and interpretation of task clusters.

CORCAS Report

In the TASSET report, a task cluster is looked at in terms of its most highly co-performed and most discriminating tasks, so that an analyst can better characterize and name the cluster. Another useful way to characterize a task cluster is in terms of the people who most perform it, and especially those principal performers whose jobs are concentrated in this task cluster to the exclusion of all or most other task clusters. This is precisely what the CORCAS report does; namely, aid the user in characterizing clusters. The CORCAS report may contain any type of background variable information describing a case that will fit in the allocated space, just as on a PRTVAR report; however, "job title" is often the most useful variable. By way of example, suppose an analyst were trying to understand why a task cluster consisted of a set of nondescript clerical tasks. By applying the CORCAS program to the task cluster, the analyst finds that most of the individuals who are most heavily involved with this task cluster listed as their job title "library clerk." This would indicate that this set of tasks (which nowhere mentions the word "library") is peculiar to clerical work in a library setting. Other variables, such as grade level, organization type or level, official job codes, and even KPATH sequence numbers (job type relationships) can be useful interpretive indicators in coming to an understanding of "why" a task cluster occurred in terms of "who" performs it.

JOBMOD Report

The JOBMOD (job type versus task module mapping) program aggregates the case- and task-level indices computed by the four "CORSET" programs and uses these aggregate measures to relate task clusters to job types and vice versa. The description of job types by a handful of discriminant

clusters of tasks, and the association of each task cluster to the types of jobs of which it is an important component, is a basic requirement for defining and integrating the MPT components of an existing or potential Air Force specialty (AFS) or weapon system. If AFSs are to be collapsed or shredded out, or new jobs are to be assigned to an occupational area, or old jobs are to be moved to another occupational area, such highly summarized, yet meaningfully discriminant hard data are essential. There is an especially critical need for this kind of data if numerous AFSs are to be reviewed and compared by panels of subject-matter experts (SMEs) or functional managers who must make broad-based judgments, recommendations, or decisions. JOBMOD products will provide a welcome assist to those who are responsible for meeting another important MPT requirement, that of integrating task data from a variety of databases, such as the Logistics Composite Model (LCOM), the Logistics Support Analysis (LSA), Maintenance Data Collection (MDC), and the Occupational Survey Report (OSR) databases. Task clusters and their associated job types provide fairly stable, functionally discrete categories into which tasks from other data sources can be mapped and ultimately integrated.

IV. PHASE III: AUTOMATION OF CLUSTER SELECTION AND REFINEMENT

MODTYP Program

Just as the JOBTYP (job typing) program automatically selects from a hierarchical clustering of cases the "best" set of job types based on similarity of time spent across tasks, the MODTYP (module typing) program selects from a hierarchical clustering of tasks the "best" set of task module types based on task co-performance across cases. The term "best" means that the evaluation algorithm initially optimizes on four criteria simultaneously (i.e., within-group homogeneity, between-group discrimination, group size, and drop in "between overlap" in consecutive stages of the hierarchical clustering). After all stages of the clustering have been evaluated on these criteria, a primary, a secondary, and a tertiary set of mutually exclusive task clusters are selected as first-, second-, and third-best representations of the modular structure of the hierarchical clustering solution. The three sets of groups are then input to another evaluation algorithm which computes supergroup and subgroup indices between all pairs of groups in the primary solution; between secondary groups and primary groups within the same KPATH range; and between tertiary and secondary groups, as well as between tertiary and primary groups, within the same KPATH range. Based on the combined results of both evaluations, the primary, secondary, and tertiary sets of groups are revised; i.e., groups at one level may be promoted or demoted to replace a group or groups within the same KPATH range at another level. The final set of primary groups is input to the TASSET and CORCAS programs to provide analytic and interpretive data for each primary cluster of tasks. The MODTYP output also produces a report which shows the initial and final sets of primary, secondary, and tertiary groups and their evaluation indices.

In the several applications of MODTYP to date, MODTYP has picked virtually the same task clusters as the analysts in one-third to one-half of the selections, and has deviated by no more than one or two clustering stages in another one-third to one-half of the selections. In only one-sixth of the selections or less has MODTYP deviated radically from the analyst selections and, in those instances, it was usually arguable as to which was the better selection. When the analyst deviates significantly from the MODTYP selection, it is hypothesized that the upward or downward location of the secondary and tertiary groups on the DIAGRM report will be predictive of the direction of deviation. Insufficient data have been analyzed at this point to warrant acceptance or rejection of this hypothesis.

The special value of automated cluster selection programs like JOBTYP and MODTYP in an MPT research environment is that they provide completely automated, rapid, yet reasonably accurate and highly standardized results that are completely replicable. Accuracy, standardization, and replicability of results across multiple occupational survey studies is essential if MPT

integration is to be a systematic process. Where occupational analyses have been done by inexperienced analysts, or where the principal objective of the analysis was to identify career field structure at the major cluster level rather than at the job type level, the completely automated feature of these programs becomes critical, because the MPT researcher may not be equipped to perform a standard CODAP job type analysis. The rapidity of the procedure will play an important role if a number of job or task module typing exercises must be done in order to proceed to the next step in the MPT integration process.

The JOBMOD program and its role in MPT integration was described earlier in this paper. If JOBTYP and MODTYP are employed to produce the job and task clusters to be input to JOBMOD, JOBMOD then becomes the end product of a totally automated process that may well serve as the appropriate beginning analysis/evaluation point for the MPT integration process.

OVLGRP and SEDGRP Programs

If JOBMOD provides the starting point for the MPT integration process, one might ask whether it could be made a better one. Could the job and task clusters consist of a better balance of within-group homogenity, between-group difference, and number of cases or tasks classified into the selected clusters? The answer is affirmative, and the refinement procedure to be used to accomplish this balancing act is the OVLGRP (overlap grouping) nonhierarchical clustering procedure. OVLGRP can use the centroids of hierarchically formed clusters of jobs or tasks as input seeds about which to optimally cluster all cases or tasks. Or, it can use the input seeds generated by the SEDGRP (seed group) program; i.e., discriminant cases or tasks which represent the entire measurement space encompassed by the sample data. Specific applications of the nonhierarchical clustering procedures are beyond the scope of the present paper. OVLGRP and SEDGRP are mentioned here merely to indicate how job and task clusters can be refined prior to their being input to JOBMOD.

V. CONCLUDING REMARKS

The purpose of this paper was to familiarize actual and potential ASCII CODAP users with a number of ASCII CODAP products specifically designed for the definition, selection, analysis, and interpretation of task clusters. These products allow clusters of tasks to be examined, compared, and interpreted with the same degree of care and meticulousness previously associated only with case clustering and the resultant job clusters. The USAF Occupational Measurement Center (USAFOMC) is already using the new ASCII CODAP technology to combine and integrate the two analytical streams in an overall multidimensional approach to studying the world of work. Future technology R&D efforts will attempt to improve and more fully automate this integration process in support of the MPT integration R&D effort.