

A methodology to facilitate occupational task grouping: Meeting the demands of training analysis and planning

BRUCE M. PERRIN, DAVID S. VAUGHAN, and JIMMY L. MITCHELL
The Boeing Company, St. Louis, Missouri

and

ROBERT M. YADRICK and WINSTON BENNETT, JR.
Air Force Research Laboratory, Brooks Air Force Base, San Antonio, Texas

Methods for grouping occupational tasks are required to support a broad range of personnel actions and organizational planning activities. Having subject matter experts sort tasks into groups is the only methodology generally recognized for these purposes. For many applications, however, and training in particular, analyses that cover large areas of an organization may be required. For such uses, manual sorting is costly and may be infeasible. A new method, based on statistical clustering using task co-performance, is described. Results indicate that this method can replicate much of the structure of the experts' groups and so can be used to facilitate task grouping. Implications of this new approach are discussed.

The need to cluster occupational tasks is ubiquitous in personnel and industrial psychology. In job analysis, a grouping of the tasks in a job traditionally has been sought to more clearly describe important activities and responsibilities (McCormick & Ilgen, 1980). Additionally, the role of task clustering in job analysis is not expected to decline; Cascio (1995) has predicted that task cluster descriptions will be of increasing value in understanding the nature of work in future organizations as traditional organizational barriers are removed and jobs become more broadly defined. For training, task clusters again play a central role, in this case, in training needs analysis and training planning (Goldstein, 1993). Groups of tasks with comparable knowledge and skill requirements pose similar training needs that can be more readily assessed and managed.

For the practitioner, two primary methodologies exist for forming these task groupings: Subject matter experts (SMEs) may generate task groups or groups may be statistically defined on the basis of task characteristics. In the latter area, task characteristics are often generated by SMEs. For example, factor analyzing SMEs' importance ratings has been a widely used approach, according to Cranny and Doherty (1988). Unfortunately, as these authors clearly have demonstrated, ratings of the importance

of individual tasks will not generally reflect the relevant sources of covariance between tasks, so the resulting task groupings are inappropriate for job analysis and training analysis and planning purposes.

Alternatively, SMEs may be asked to rate the similarities between tasks. Then these data can be analyzed using various statistical methods, including cluster analysis, multidimensional scaling, or factor analysis to produce appropriate task groupings. This approach requires the generation or estimation of all of the pairwise comparisons between tasks, so for n tasks, $[n * (n - 1)] / 2$ judgments are required. Having SMEs generate the required number of judgments becomes impractical when the task list is lengthy. A complete similarity matrix can also be estimated from a partial set of ratings—for example, by using multidimensional scaling to estimate the missing values (Perrin, Knight, Mitchell, Vaughan, & Yadrick, 1988). This approach, however, requires some basis on which to select tasks well distributed in the multidimensional space, again requiring the intervention of SMEs in some form, and the practical value of this approach has not yet been demonstrated.

Subjective grouping of examples, concepts, or objects is a widely accepted methodology, and it has been used for a broad range of purposes, from the industrial applications described above to research in cognitive modeling and comparisons of experts' and novices' perceptions (e.g., Schoenfeld & Herrmann, 1982). Cranny and Doherty (1988) cited SME generation of task groups, in addition to statistical analysis of SME-provided similarity ratings, as the applicable methods for job analysis. For training uses, Goldstein (1993) has also noted the usefulness of SME-generated task groups, a process he calls the rational clustering exercise.

This research was supported by Contracts F33615-83-C-0028 and F33615-89-C-0001 from the U.S. Air Force Human Resources Laboratory, Training Systems Division, Brooks AFB, San Antonio, TX. This paper is dedicated to the memory of David S. Vaughan, a major contributor to this article and this line of research. J.L.M. is now at the Institute for Job and Occupational Analysis, San Antonio, TX. Correspondence should be addressed to B. M. Perrin, The Boeing Company, Mailcode 0343065, P.O. Box 516, St. Louis, MO 63166-0516 (e-mail: bperrin@mdc.com).

For job analysis, SME-generated task groupings are often a practical solution. Even when jobs are complex and task statements are detailed, a job incumbent can usually provide systematic and complete groupings in a few hours. For training analysis and planning, however, the solution is frequently not this straightforward. Often, the requirements for training data span several jobs or even an entire career field, such as when a training plan is to be developed for a functional unit of an organization. As a result, the applicable set of task statements may number in the hundreds. Additionally, it may not be possible to find an individual who is familiar with all of the relevant tasks. In these cases, a panel of SMEs may be required to form task groupings.

Because of the time requirements and expense of having panels of SMEs generate groupings from the broad range of tasks required for training analysis and planning, the appropriateness of its routine use is questionable, but deserves evaluation. Additionally, and perhaps more importantly, the question of whether there are other bases on which tasks can be grouped remains unresolved. Some years after Cranny and Doherty's (1988) paper critiquing the use of importance ratings, we know of no published reports of attempts to develop or evaluate alternative approaches to task clustering on importance ratings, and Goldstein (1993) has summarized the status of work in the training area by noting that there are still "questions on the appropriate techniques to use in developing clusters" (p. 60). In short, little is known about methodologies for meeting the real-world demands for grouping occupational tasks for the purposes of training analysis and planning.

This paper reviews the requirements and results of having panels of SMEs generate task groups. The consistency of the results between different panels is reported, and the time required and the personnel involved are described. Results are reported for two Air Force occupations representing very different types of work requirements—aircraft navigation systems maintenance and law enforcement/security police. Then, an alternative methodology is described, based on statistically clustering tasks using existing task data. The results obtained from this alternative methodology are compared with those of SME-generated solutions. Finally, an approach that combines these methods is derived and applied, the use of the resulting task groupings in a training analysis and planning system is described, and implications of this methodology are discussed.

METHOD 1 **SME Task Sorting**

Perhaps one of the most straightforward ways to classify information is simply to have people sort examples into categories. The groups that result constitute the categories. One relatively simple way to implement this approach is to ask individuals to sort the information printed

on cards into appropriate piles. This was the first effort undertaken by our panels of SMEs.

To make the job of sorting all of the tasks in an occupation more manageable, we wanted to provide some initial structure in the form of starter groups. These initial piles of cards were not to limit the sorting of tasks in any way, but merely to provide groups of manageable size based on reasonable divisions of the work. Air Force training and operational managers already organize tasks into groups in a document called Specialty Training Standards (STS), used to promote discussions and reach agreements on the level of training to be provided for each group of tasks. We believed that STS groups might make acceptable starter piles. A second set of starter piles, grouped on the basis of task coperformance clustering, was provided in order to determine whether initial structure unduly affected the final results. Coperformance clustering is more fully described in the second study.

SMEs worked as teams and were asked to rearrange the cards into piles that represented groups of tasks that "should be trained together." The directions to the SMEs stressed the importance of using their expertise, and the instructions specifically noted that the resulting groups could include the same task in two or more groups and could be of any size, including single tasks. In both occupations we studied (navigation systems maintenance and law enforcement), enough SMEs were available to form two or more panels. These panels worked independently, proceeding at their own pace and using their own strategies to sort the tasks. When two SME panels were satisfied with their task clusters, they met to reconcile any differences between them. This reconciliation phase allowed the SMEs to compare strategies, decide on the best criteria, and produce a final set of task clusters.

Procedure

The card-sorting method in the navigation systems maintenance occupation was applied at Keesler AFB, Mississippi. Ten SMEs who were technical trainers at the base served as subjects. The average level of the SMEs' experience in the career field was almost 11 years and varied from about 7 to more than 15 years. The task list for this occupation contained 778 items, and all tasks were descriptions of complete units of work, such as "bench check component X on specification Y," rather than parts of a task, such as steps in a procedure. Initial card sorts required about 1½ days for one panel and approximately 2 days for the other panel. Reconciliation of the results from the two panels was completed on the 3rd day.

The card-sorting method for the law enforcement career field was applied at Lackland AFB, Texas. Fourteen SMEs participated in the exercise. Eight were currently trainers at the base, and the remaining 6 were assigned to operational units. Experience in the career field varied from just under 5 years to more than 24 years, with an average of slightly more than 12 years. Because both training and operational units were represented, individ-

uals were assigned to panels differing in both background (field or training) and type of starter pile (STS or coperformance). The four resulting panels are identified by the type of starter pile and background as follows: STS/school, STS/field, coperformance/school, and coperformance/field.

The law enforcement task list contained 666 items. The two panels with coperformance starter piles finished sorting midway through the 2nd day. They started reconciling their results at that time and completed their work by the end of the 3rd day. The two panels with STS starter piles did not finish their initial sorts until the morning of the 3rd day, but also completed their reconciliation by the end of the day.

Results

Of the six task sorts in law enforcement (four panel results and two reconciliation sorts), descriptive statistics on five of the sorts were very similar. Across these five sets of results, the number of task groups was approximately the same, ranging from 65 to 75 groups, and the average group size ranged from about 9 to 11 tasks. Similarly, few task groups were formed with only a single task in these five sorts, and the largest task group in these sorts varied from 30 to 49 tasks. The same task was seldom placed in two or more groups by these SMEs (a maximum of 46 tasks were duplicated), and most of the tasks were classified (a maximum of 12 tasks were unclassified across the five sets of results). The sixth sort, the one from the STS/school panel, however, showed fewer (only 33) and larger (averaging over 20 tasks per group) groups. Their results involved more duplicate (129) and unclassified (117) tasks. Additionally, their largest group contained 109 tasks.

The results for the card sorts from the navigation systems maintenance panels were, on the surface, also quite dissimilar. Groups formed from the STS starter piles averaged only about one fourth as many tasks per group as the groups formed from the coperformance starter piles (5.6 tasks per group vs. 23.6 tasks), although the largest task group produced by both SME panels was not substantially different (78 vs. 89 tasks). Additionally, neither panel used many duplicate tasks and most tasks were classified. The results of the reconciliation sort by these SMEs reflected compromise on all measures. The number (75) and average size (10.1) of the groups, the number of single-task groups (4), and the number of duplicate (7) and unclassified (26) tasks following the reconciliation sort were all greater than the same measure for one panel's results and less than it for the other.

For our purposes, however, the consistency with which the tasks were grouped is more germane than characteristics such as number or size of the groups, although the two are related. Several statistics are available to compare groupings, based on pairwise classification of items in the two solutions. One such statistic is the Fowlkes and Mallows (1983). If A indicates the number of pairs of tasks grouped in both solutions, and B and C indicate

frequencies of disagreement in which pairs are grouped by one method, but not the other, the Fowlkes and Mallows (F&M) is computed as follows:

$$F\&M = \frac{A}{\sqrt{(A+B) * (A+C)}}$$

The F&M has an upper bound of 1.00 when the two solutions agree perfectly, and a lower bound of 0.0 when A is zero. Additionally, the F&M is undefined when cells A, B, and C are all equal to zero, but this would occur only when the number of groups equals the number of cases for both solutions (i.e., no grouping had occurred).

A sampling distribution for the F&M was generated by randomly assigning tasks to groups and then computing the agreement statistic to determine the level of agreement that would occur by chance for solutions representative of the number, size, and type of our task groups (i.e., same number of duplicate tasks and tasks not grouped). Comparisons of two solutions with relatively large task groups showed higher levels of chance agreement than comparisons involving solutions with smaller groups. For comparisons of solutions with large groups (averaging over 30 tasks per group), the F&M statistic exceeded 0.094 in only 1 case out of 100. We adopted this number as the critical value of the F&M statistic at the .01 significance level. Selection of this value provides a very conservative test since the F&Ms for comparisons of solutions more typical of our results (about 9 tasks per group, on average) exceeded 0.027 in only 1 of 100 cases.

Guidance from the statistical literature was equivocal as to how to treat tasks not classified and how to treat duplicate task statements. In the first instance, unclassified cases are generally omitted from statistical comparisons, when the available information does not support clustering them in any group. Tasks were left unclassified in the card sorts primarily because the SMEs felt that these tasks were no longer part of the responsibilities of the occupation, not because of an inability to place them with related tasks. Therefore, we decided to omit these tasks from the analysis and in the estimation of the sampling distribution. In the case of duplicate tasks, each instance of the task was tabulated individually. That is, one grouping of the duplicate may have agreed with the second solution (and be counted in A), whereas a second sorting would disagree with the other solution (and be counted in B or C).

The results of comparing the SME card sorts are reported in Table 1. All comparisons are F&M agreement statistics, and all are statistically significant at the .01 level. The statistics indicated by an asterisk (*) are not independent comparisons; that is, they are comparisons involving a reconciliation sort and the solution from one of the two panels reconciling their results.

The F&Ms ranged from 0.199 to 0.304 for the 11 independent comparisons involving groupings by law enforcement personnel. The two task groups formed by the field SMEs agreed more closely than any other set of comparisons, even though these SMEs worked from dif-

Table 1
Comparisons of SME Card Sorts

	STS Reconciliation	Coperformance School	STS School	Coperformance Field	STS Field
Military police					
Coperformance reconciliation	0.293	0.293*	0.199	0.716*	0.300
STS reconciliation		0.290	0.227*	0.294	0.757*
Coperformance school			0.233	0.270	0.298
STS school				0.199	0.219
Coperformance field					0.304
Navigation systems maintenance					
Reconciliation		0.520*	0.787*		
Coperformance school			0.476		

Note—SME, subject matter expert; STS, Specialty Training Standards. *Not independent results.

ferent types of starter piles. Additionally, the reconciliation sorts were more consistent with the field SMEs' groupings than the groupings formed by the SMEs involved in training. In fact, the large differences in the F&M agreement statistics suggest that the reconciliation sorts were based to a large degree on the criteria and the groupings developed by the field personnel.

Only one comparison was independent for the navigation systems maintenance task sorts. The F&M statistic for this comparison was quite high (0.476), greater than for any comparison between independent law enforcement SME card sorts.

Discussion

Although the F&Ms were all statistically significant, one would like to develop an impression of the level of agreement expressed in these F&Ms so that their practical importance could be evaluated. Obtaining this insight is difficult given the number of cases (tasks) and groups produced and the complexity of the relationships among solutions, although some inferences can be drawn. For example, if all of the solutions involved task groups of 9 (near the average for most of our results), an F&M of 0.426 would result if 6 of 9 tasks in one solution were also grouped in the second solution. This result is quite comparable to that obtained by the navigation systems maintenance SMEs, with an F&M of 0.476 for the comparison of their independent sorts. The card sorting results are not this simple, of course, since group sizes varied widely within a given sort.

The F&M, like other agreement statistics of this type, is substantially affected by differences in the size of groups being compared (i.e., the specificity or granularity of the solution). Consider, for example, the two independent solutions from the navigation systems maintenance personnel, where one panel of SMEs formed groups that were about one fourth the size of the other panel's groups. Assume that if required to form more specific (smaller) task groups, the SMEs who formed the larger ones would divide each of them into the four groups found in the second solution. This difference, then, is less one of how to group the tasks as it is one of what level of specificity to use. And, these differences in group size alone reduce the F&M to about 0.43, less than the observed F&M for the comparison between these

sorts. Even for solutions in which the average group sizes were the same, observation of the SMEs during card sorting suggested that differences in task group size in different areas of the occupation were common, due to differences in expertise and emphasis. With variation in specificity from area to area, solutions could average 9 tasks per group, yet show substantial differences in specificity across areas of specialization.

Although no systematic analysis of differences in specificity was performed (due to the overall size of the task lists and the complexity of the group structures), we believe that a substantial part of the observed shrinkage in the F&M statistics can be attributed directly to this cause. If task groups could be equated for specificity, we would expect a level of overlap of 70% or more between individual task groups. Across the occupations, SMEs were able to achieve considerable agreement on the tasks that should be trained at the same time.

METHOD 2

Statistical Clustering

Statistical clustering has a long history in military occupational analysis. For more than 30 years, job analysts in the armed services have used case cluster diagrams to assist in identifying jobs—groups of people performing similar sets of tasks. Over that period of time, standardized data collection instruments (occupational survey task inventories) have been developed and refined; computing algorithms and diagnostic statistics have been devised, tested, and implemented in the Comprehensive Occupational Data Analysis Programs (CODAP); important background characteristics have been identified and incorporated to aid in the analysis of the structure of work; and an extensive body of research and application has accrued (Christal, 1974; Christal & Weissmuller, 1988).

By comparison, the notion of using cluster analysis to derive task groups to support training decision making is relatively new. Research on the use of cluster analysis for these purposes began in only the mid-1980s. Nonetheless, a variety of algorithms has been added to CODAP for the analyst interested in statistically clustering tasks (Phalen, Mitchell, & Hand, 1990).

We used task coperformance as the similarity index in this analysis and the average linkage clustering proce-

ture (Ward, 1963) contained in CODAP. Whereas identifying jobs in an occupation using cluster analysis involves grouping persons who perform the same (or similar) sets of tasks, task clustering using copformance similarity produces sets of tasks that tend to be performed as a group; that is, if an individual performs one task in the group, it is likely that he/she also performs others in the same group. Importantly for our work, tasks that are copformed may be similar with respect to requisite skills and knowledge if copformance results from alternative ways of achieving similar objectives, from work using or performed on the same piece of equipment, and the like.

We clustered tasks using the same occupational survey data that had been collected to describe jobs—ratings by each job incumbent of the tasks performed. However, prior to clustering, the data were transposed from the person-by-task format, which is used to identify jobs, to a task-by-person format, which produces copformance clusters. Because of the reuse of existing survey data, task clustering was available to us with only the costs of computer processing and analysis; no additional data collection costs were incurred. In organizations in which task-based occupational analyses are not routinely performed, such data may not be available. All of the SME-based procedures, however, recommend use of task lists. As a result, the costs of using a task list to survey a representative sample of job incumbents can be compared with the costs of convening an SME panel to evaluate it. These costs are often quite comparable, but where the occupation covers a broad range of skills and SMEs are busy or widely dispersed geographically, the costs of convening SME panels may make this approach impractical.

To identify task groups from a copformance cluster diagram, we elected to use the same methods that a job analyst would typically use to interpret a case cluster diagram. The heart of this approach is to identify clusters that maximize the variability between groups while minimizing the variability within groups (Archer, 1966). In practice, the analyst applies various rules of thumb regarding, for example, the expected levels of within- and between-group homogeneity, and may seek the assistance of SMEs in refining job structures when cases cannot be placed solely on the basis of statistical data. Given the importance analysts play in case clustering, we selected the same approach for task clustering. It was apparent from the outset, however, that some of the informal guidelines that job analysts have developed over the years did not apply to task cluster analysis. For example, the 35–50 rule, a general rule of thumb that occupational analysts use to identify initial jobs, could be made more stringent.¹

Results

The analysts who interpreted each of the task copformance cluster diagrams were experienced in occupational analysis and were also somewhat familiar with the occupation they were evaluating. They found that the most useful information for identifying task groups was the pattern of changes in homogeneity, rather than the ac-

tual levels. Specifically, the points at which homogeneity dropped significantly as tasks were combined tended to mark distinct task groups, a heuristic also used to identify jobs. The analyst who interpreted the law enforcement cluster diagram identified 67 task groups, resulting in an average group size of about 8 tasks. The largest task cluster contained 34 tasks, and no groups were composed of a single task (although several involved a pair of tasks). Additionally, 115 tasks could not be classified solely on the basis of the cluster diagram.

Working with the navigation systems maintenance task cluster diagram, the analyst identified 95 task groups, for an average of just under 8 tasks per group. Only 35 tasks were not classified. The largest task cluster contained 47 tasks and the smallest again consisted of a pair of tasks.

Table 2 reports the F&M agreement statistics between the task groups formed by SME card sorting and the task clusters identified from the copformance cluster diagrams in each occupation. All comparisons were statistically significant at the .01 level, indicating that the agreement between the statistically defined and the SME-generated clusters was statistically greater than expected by chance. Additionally, the magnitude of the F&M statistics for these comparisons was roughly equivalent to that found for comparisons between different SME card sorts, although the statistics for the law enforcement occupation consistently exceeded those for the independent card sorts in the same occupation.

Discussion

The descriptive statistics reported in this and the previous section suggest that there is considerable agreement as to the appropriate level of specificity for task groups. Across both occupations, the reconciliation card sorts and analysts' interpretations of the copformance cluster diagrams yielded groups with about 8–11 tasks. Although some of the independent card sorts resulted in both more broadly and more narrowly defined groups, these differences tended to be minimal when final groups were formed.

The type of starter pile used in the card-sorting exercise appears to have had no appreciable effect on the results.

Table 2
Comparisons of SME Card Sorts With the Task Clusters Identified From the Task Copformance Cluster Diagrams

SME Card Sorts	Analyst's Interpretation of the Cluster Diagram
Military police	
Coperformance–reconciliation	0.445
STS–reconciliation	0.463
Coperformance–school	0.441
Coperformance–field	0.442
STS–school	0.305
STS–field	0.472
Navigation systems maintenance	
Reconciliation	0.315
STS–school	0.273
Coperformance–school	0.153

Note—SME, subject matter expert; STS, Specialty Training Standards.

Card sorts that began with coperformance starter piles tended to be no more similar to the analyst-identified statistical clusters, which were based on coperformance, than the card sorts produced by SMEs using STS starter piles. This result was perhaps to have been expected since the card-sorting directions encouraged SMEs to use their expertise to restructure the initial piles as necessary.

In the law enforcement occupation, in which both training and field personnel participated in the card sorts, the background of the SMEs may have affected the results. Field SMEs' card sorts tended to more closely match the coperformance clusters than did the school SMEs' sorts, although the difference was small. Presumably, field SMEs are more sensitive to performance criteria and so produced groupings similar to those of task coperformance. Additionally, it should be noted that the results from the two field panels showed the highest agreement and that the reconciliation sorts were more similar to the field SMEs' results than to the school SMEs' task groups (as reported in Table 1).

An additional interesting finding in the comparisons between the card sorts and the statistical clusters for law enforcement involves the pattern of results. The F&Ms between the statistical clusters and each of the card sorts were higher than those between any independent card sort in the occupation. The closest agreement between two independent SME sorts was that for the two field panels, which yielded an F&M statistic of 0.304 (Table 1). All comparisons between the card sorts and the coperformance clusters yielded values of the F&M statistic greater than this level. Although this result suggests substantial common agreement between these techniques, the data do not necessarily support this interpretation. Agreement statistics such as the F&M have been shown to be substantially affected by decisions about the exclusion of cases from the analysis (Edelbrock, 1979). With more than 17% (115 of 666) of the tasks omitted from the statistical clusters, the relatively higher level of the F&Ms found may be an artifact of the exclusion criteria used by the analyst.

METHOD 3

Statistical Clustering With SME Refinement

In a number of ways, card sorting by SMEs and task coperformance clustering represent complementary ways of forming task groups. On the one hand, statistical clustering can replicate, to a great extent, the clusters SMEs produce when asked to sort tasks into groups that should be trained together, doing so efficiently and using existing information. The resulting statistical clusters can be expected generally to agree with SME-produced groups as well as solutions from independent panels of SMEs agree with each other.

On the other hand, statistical clustering has certain limitations in its flexibility, regardless of its ability to replicate SME-produced task groups. It is not always possible to assign a given task to a cluster solely on the basis

of a cluster diagram. In identifying jobs from a case cluster diagram, for example, it is common to have from 5% to 10% of the cases unclassified. This problem was even more pronounced for task clustering in the law enforcement occupation, where 115 of 666 tasks were omitted from any group. One possible solution would be to leave these tasks as single task groups; however, this solution does not compare favorably with the card-sorting results. Another solution would be to relax the restrictions on the coperformance clusters so that more of the tasks would be grouped. It is not clear, however, that the hierarchical procedure would produce acceptable groupings using relaxed inclusion criteria. It has been suggested that nonhierarchical methods for refining task and job clusters around analyst-identified "seed" groups might be a more realistic approach (Phalen, Staley, & Mitchell, 1988), but the work needed to demonstrate such an approach has not yet been reported.

SME card sorting represents a very flexible method for grouping tasks to portray training requirements, using established methods and providing descriptive frameworks for the resulting groupings. The method has some limitations as well. Our results lead us to two conclusions. The first is that multiple panels of SMEs should be involved. This conclusion follows from the somewhat inconsistent results produced by one of the law enforcement SME panels. Given the current state of the requirements for card sorting, a single SME panel may produce task groups that cannot be replicated and probably would not adequately fulfill the requirements for training analysis and planning. At a minimum, a second panel of SMEs should independently sort the tasks so that the appropriate comparisons can be made. If the results are not consistent, further SME input would be warranted. Second, our results suggest that having an SME panel that includes operational personnel is also highly desirable. Where data were available, the field SME panels were more consistent with each other, were more consistent with the reconciliation card sorts, and were more consistent with the coperformance clusters than were the SME panels composed solely of training personnel.

Given the requirement for multiple SME panels composed of both training and field personnel and the fact that card sorting required a minimum of 2 days (3 if a reconciliation sort was produced), arranging for card sorting in any particular occupation can be difficult. These difficulties are likely to be compounded for small, highly specialized career fields in which personnel are dispersed geographically, or when workload is high.

Because the strengths of one method tend to compensate for the weaknesses of the other, we formed a composite procedure for grouping tasks—coperformance clustering interpreted by an analyst, followed by SME refinement. This procedure capitalizes on the strengths of both clustering and card sorting, while avoiding many of the pitfalls. Coperformance clusters are identified to capture the core of agreement in task groupings with moderate costs. SMEs then refine these groups, classify any

tasks that the analyst could not place, and supply descriptive labels for the groups, giving the procedure additional flexibility. This composite method was applied to both the navigation systems maintenance and the law enforcement occupations with the following results:

1. All refinement of task groups, including placement of all tasks not clustered, was completed in 1 day. Additionally, the number and magnitude of changes the SMEs made to refine the analyst-identified task groups were rather small. The comparison between the task groups initially identified by the analysts and the SME-refined groups yielded an F&M agreement statistic of 0.919 for law enforcement and 0.877 for navigation systems maintenance.
2. The descriptive statistics on the resulting task groupings were comparable to previous efforts in the size of the task groups (between 9 and 11 tasks); use of duplicate tasks (0 and 6 duplicate tasks); and number of single-task groups (4 and 5 single-task groups).
3. The F&M agreement statistics for the comparisons between the card sorts and the task groups formed by this composite method were all statistically significant, tended to exceed the statistics for comparisons between independent sorts, and tended to be greatest for the reconciliation card sorts. Overall, this method appears to have tapped a common core of agreement among SMEs as to which tasks should be trained at the same time.

Thus, the combined procedure has the strengths of both computerized statistical clustering and expert human judgment. It uses statistical clustering to quickly and economically capture general patterns in task groupings about which experts can agree. It then uses expert judgment to refine and succinctly characterize these task groups.

CONCLUSIONS

The ability of copformance clustering to provide a first approximation of SME-generated task groupings was not universally predicted by the authors, and in fact, several other methods were initially considered better candidates. For example, in the navigation systems maintenance occupation, tasks were also clustered on the basis of similarities of aircraft, systems, and components maintained, as well as on common test equipment operated. For law enforcement, task similarity matrices were also based on equipment, weapons, and tactics used; courses taken; and functions performed by the incumbent in his/her job. In no case did the clusterings based on these task characteristics exceed the task copformance groups in terms of their agreement with the SME-generated solutions (Perrin et al., 1988). In short, task copformance clustering provided the best approximation of SME-generated groups of all the measures assessed.

To have maximum benefit to practitioners faced with similar problems, a more complete understanding of why copformance clustering provided a good approximation of SME-generated task groups is needed. Ultimately, that understanding will come only with additional research on this issue. We may, however, speculate that two

characteristics of the data available on these occupations may have contributed to our positive findings. First, each task statement represented a complete unit of work, such as, "Bench check component X on specification Y," rather than steps in a unit of work, such as steps in a procedure. With tasks representing complete units of work, copformance appears to be high when other task statements represent alternative ways of achieving the same function, such as, "On the aircraft, check component X on specification Y"; or of testing a different aspect of the same equipment, such as, "Bench check component X on specification Z"; or of performing the same test on a different but related system, such as, "Bench check component A on specification Y." In each case, the tasks share significant components of systems or procedural knowledge. We speculate that the ability of copformance clustering to replicate SME-produced groups more effectively than clusterings based on other factors may be because copformance can be sensitive to multiple factors—equipment maintained, systems on which the work occurs, methods used, and so on—concurrently or individually, as they are relevant to the work performed.

Second, copformance clustering may provide a good approximation of SME task groups if there are opportunities for variability in the structure of work across the organization. Similar tasks, we have argued, may tend to be copformed because they involve similar functions, equipment, or actions. Dissimilar tasks, however, may be copformed largely because of convenience or local policy. For example, one individual may perform two dissimilar tasks, such as, "Bench check component X on specification Y" and "Complete inventory form A." If individuals in other areas of the organization, due to differing demands on their time, different policies, or different aptitudes, perform the bench check task but not the inventory task, copformance of these tasks will be relatively low. Additionally, even if every individual who performs the bench check task also performs the inventory task, copformance may still be low if other individuals perform the inventory task but not the bench check task. In short, dissimilar tasks must be paired in jobs frequently and exclusively for copformance to be high. With variability in job structure across the organization in either factor, copformance will be reduced.

It seems unlikely that dissimilar tasks would be paired frequently and exclusively by chance, and so copformance for dissimilar tasks may be low unless directed by organizational policy. In such cases, it is likely that the SME review of the statistical clusters would readily identify task groups that were copformed by policy, rather than efficiency. Additionally, however, this information may be relevant to training planning since it may reveal areas where training and perhaps performance inefficiencies exist by design.

The work reported in this paper supports the development of an organizational simulation that allows managers to conduct training utility analyses over a broad range of changes to manpower, personnel, and training programs.

This simulation, called the Training Impact Decision System (TIDES), permits tradeoffs in organizational outcomes such as overall costs and resource usage. Specifically, the TIDES holds organizational productivity constant so that alternative manpower, personnel, and training programs, including different types and levels of formal training and on-the-job training (OJT), can be compared. Within TIDES, the demand for training under varying levels of staffing, different job structures and patterns, and different mixes of formal training is derived from an entity-based simulation that represents the flow of personnel between jobs and training throughout their careers (Mitchell, Yadrick, & Bennett, 1993). Proficiency gains from the use of formal training methods, such as classroom lecture or supervised task performance in a laboratory, are represented as a learning curve for a given set of tasks within the occupation. Recognizing that a given training method generally has a point of diminishing returns, statistical models fit to SMEs' estimates of proficiency gains under varying allocations of these methods consistently produce the expected, negatively accelerated, learning curves (Bennett & Perrin, 1989; Perrin et al., 1988).

Data from the simulation are used, and the amount of formal training for each individual across all courses is translated into proficiency using the learning curves; any shortfall of these courses to achieve the established level of productivity is estimated. The shortfall, in turn, is translated into OJT requirements, using a learning curve for this type of training. Finally, resource requirements for providing both formal training and OJT are compared with the capacities of the organization to provide these types of training, and overall labor and nonlabor costs are estimated (Rueter & Feldsott, 1989; Rueter, Feldsott, & Vaughan, 1989). More complete overviews of the TIDES system are available to the interested reader (Mitchell et al., 1992; Vaughan et al., 1989).

The TIDES simulation and learning curve models were built from Air Force occupational survey (OS) tasks, the source of the task lists used in this work. Since OS tasks will share skills and knowledge with other tasks in the same occupation, an estimate of the number of the hours required to train all of the tasks in an occupation individually will overstate training requirements. To more accurately estimate training requirements, the TIDES model must recognize when a trainee or job incumbent is partially proficient on a task on the basis of having received training on similar, related tasks. The task groupings produced by the combined copformance clustering-SME refinement methodology provide this required TIDES capability.

In addition to the role in training planning and analysis, some additional benefits of using task groups derived from this methodology are beginning to be realized. Because these analyses are based on a representative sample of the organization's workforce, copformance clusters

provide a broader functional and geographic perspective than is often possible with a group of SMEs. This problem is more severe, of course, where the occupation is diverse and the work is performed at numerous, remote locations. Accordingly, research performed for the U.S. Coast Guard revealed that empirically defined task copformance clusters were preferable to SME defined "duty areas" for obtaining a broad perspective on a career field appropriate to basic job skill training and selection issues (Weissmuller & Driskill, 1991). And, in work with the Internal Revenue Service, task copformance clusters were found to provide linkage between detailed, behavioral job descriptions and broad knowledge, skill, and ability categories that could provide the basis for a wide range of personnel actions (Weissmuller, Driskill, & Moon, 1991).

But perhaps most of the utility of task copformance clusters is yet to be discovered. Organization-wide descriptions of work and skill are expected to play an ever-increasing role in understanding and managing the flexible work structures of the future (Cascio, 1995). The broad perspective that task copformance clusters can provide on how work is structured may play a central role in these efforts. Additionally, these task groupings may provide important diagnostic information, an additional tool for the personnel or industrial psychologist. In particular, groups of tasks that are coperformed but that are dissimilar in content may be symptomatic of problems in the structure of work or training. In cases in which SMEs readily regroup these tasks with others requiring similar skills and knowledge, identification of the root of the high level of copformance seems warranted. For practitioners performing task-based job analysis, task copformance clustering is an additional capability that comes at very little cost. Overall, our success with this methodology suggests that task copformance clusters refined by SMEs warrant evaluation for other applications in personnel and industrial psychology.

REFERENCES

- ARCHER, W. B. (1966). *Computation of group job descriptions from occupational survey data* (PRL-TR-66-12, AD-653543). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.
- BENNETT, W. R., & PERRIN, B. M. (1989, August). *The task characteristics subsystem: Allocating task modules to training settings*. Paper presented at the annual meeting of the American Psychological Association, New Orleans.
- CASCIO, W. F. (1995). Whither industrial and organizational psychology in a changing world of work? *American Psychologist*, *50*, 928-939.
- CHRISTAL, R. E. (1974). *The United States Air Force occupational research project* (AFHRL-TR-73-75, AD-774 574). Lackland AFB, TX: Occupational Research Division, Air Force Human Resources Laboratory.
- CHRISTAL, R. E., & WEISSMULLER, J. J. (1988). Job-task inventory analysis. In S. Gael (Ed.), *Job analysis handbook for business, industry, and government* (pp. 1036-1050). New York: Wiley.
- CRANNY, C. J., & DOHERTY, M. E. (1988). Importance ratings in job analysis: Notes on the misinterpretation of factor analysis. *Journal of Applied Psychology*, *73*, 320-322.

- EDELBROCK, C. (1979). Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody. *Multivariate Behavioral Research*, *14*, 367-384.
- FOWLKES, E. B., & MALLOWS, C. L. (1983). A method for comparing two hierarchical clusterings (with comments and rejoinder). *Journal of the American Statistical Association*, *78*, 553-584.
- GOLDSTEIN, I. L. (1993). *Training in organizations* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- MCCORMICK, E. J., & ILGEN, D. R. (1980). *Industrial psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- MITCHELL, J. L., VAUGHAN, D. S., KNIGHT, J. R., RUETER, F. H., FAST, J., HAYNES, W. R., & BENNETT, W. R. (1992). *Training decisions technology analysis* (AL-TP-1992-0026). Brooks AFB, TX: Technical Training Research Division, Armstrong Laboratory, Human Resources Directorate.
- MITCHELL, J. L., YADRICK, R. M., & BENNETT, W. R. (1993). Estimating training requirements from job and training models. *Military Psychology*, *5*, 1-20.
- PERRIN, B. M., KNIGHT, J. R., MITCHELL, J. L., VAUGHAN, D. S., & YADRICK, R. M. (1988). *Training decisions system: Development of the task characteristics subsystem* (AFHRL-TR-88-15, AD-A199 094). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.
- PHALEN, W. J., MITCHELL, J. L., & HAND, D. L. (1990, November). ASCII CODAP: Progress report on applications of advanced occupational analysis software. In *Proceedings of the 32nd Annual Conference of the Military Testing Association* (pp. 341-346). Orange Beach, AL: Naval Education and Training Program Management Support Activity.
- PHALEN, W. J., STALEY, M. R., & MITCHELL, J. L. (1988). ASCII CODAP programs for selecting and interpreting task clusters. In *Proceedings of the 30th Annual Conference of the Military Testing Association* (pp. 341-346). Arlington, VA: Army Research Institute.
- RUETER, F. H., & FELDSOTT, S. I. (1989, August). *The resource/cost subsystem: Estimating training capacities and costs*. Paper presented at the annual meeting of the American Psychological Association, New Orleans.
- RUETER, F. H., FELDSOTT, S. I., & VAUGHAN, D. S. (1989). *Training decisions system: Development of the resource cost subsystem* (AFHRL-TR-88-52). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.
- SCHOENFELD, A. H., & HERRMANN, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *8*, 484-494.
- VAUGHAN, D. S., MITCHELL, J. L., YADRICK, R. M., PERRIN, B. M., KNIGHT, J. R., ESCHENBRENNER, A. J., RUETER, F. H., & FELDSOTT, S. I. (1989). *Research and development of the training decisions system: Final report* (AFHRL-TR-88-50). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.
- WARD, J. H., JR. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236-244.
- WEISSMULLER, J. J., & DRISKILL, W. E. (1991, May). Improving job type definition for skill training by clustering on task module values. In *Proceedings of the Seventh International Occupational Analysts' Workshop* (pp. 261-267). San Antonio, TX: U.S. Air Force Occupational Measurement Center.
- WEISSMULLER, J. J., DRISKILL, W. E., & MOON, R. A. (1991). Economical method for establishing knowledges, skills, and abilities for job families. In *Proceedings of the Seventh International Occupational Analysts' Workshop* (pp. 268-273). San Antonio, TX: U.S. Air Force Occupational Measurement Center.

NOTE

1. The 35-50 rule of thumb indicates that a good starting point for identifying jobs is to select case clusters with a within-group homogeneity index of about 35.0 or less and a between-group homogeneity index of 50.0 or more, although the final jobs identified by the analyst may deviate substantially from this guideline.

(Manuscript received September 18, 1995;
revision accepted for publication February 25, 1997.)