A Comparison of the Accuracy of Four Methods for Clustering Jobs

Ray Zimmerman, Rick Jacobs, and James Farr
The Pennsylvania State University

Four methods of cluster analysis were examined for their accuracy in clustering simulated job analytic data. The methods included hierarchical mode analysis, Ward's method, k-means method from a random start, and k-means based on the results of Ward's method. Thirty data sets, which differed according to number of jobs, number of population clusters, number of job dimensions, degree of cluster separation, and size of population clusters, were generated using a monte carlo technique. The results from each of the four methods were then compared to actual classifications. The performance of hierarchical mode analysis was significantly poorer than that of the other three methods. Correlations were computed to determine the effects of the five data set variables on the accuracy of each method. From an applied perspective, these relationships indicate which method is most appropriate for a given data set. These results are discussed in the context of certain limitations of this investigation. Suggestions are also made regarding future directions for cluster analysis research.

Cluster analysis and related techniques have been applied in a wide variety of research settings. Although the present study deals with the application of such techniques in Industrial/Organizational Psychology, the results have implications for other fields of psychology as well.

An important research problem which has plagued personnel psychologists for decades is the inability to generalize the criterion-related validities of selection tests across similar jobs and similar situations (i.e., locations and conditions). This has been due, primarily, to the notion of situation specificity or the belief that every job-situation combination is unique. As a result, the traditional model of test validation has required a separate validation study for each job-situation combination. Such studies were often characterized by small sample sizes and low validity coefficients. Recent developments in the field of personnel psychology have made the traditional model obsolete.

Schmidt, Hunter, and Urry (1976) have shown that the sample sizes needed to obtain sufficient power in validation studies far exceed those which have typically been found in most studies in the past. In addition, the results of recent research in the area of validity generalization by Schmidt and Hunter (1977), Colbert and Taylor (1978), and Schmidt, Hunter, Pearlman, and Shane (1979) have shown that criterion-related validities can, in fact, be generalized across similar jobs and situations.

Validity generalization is based upon the premise that psychologists are able to identify relatively homogeneous subsets of jobs based upon job analytic data. Therefore, the resolution of substantive and methodological issues in job families research is essential to the advancement of research in validity generalization. In an exhaustive review of the job families literature, Pearlman (1980) ad-

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 6, No. 3, Summer 1982, pp. 353–366

0146-6216/82/030353-14$1.70

dressed these issues and indicated avenues of future research which might be the most fruitful. In particular, he warned against excessive controversy over methodological issues, as different methods are likely to be more or less appropriate under different circumstances and an undue emphasis on methodological issues may divert the attention of psychologists away from the more important substantive issues. While it may be true that methodology should play a secondary role in job families research, there is, nevertheless, a need for systematic research to evaluate the relative merits of the various techniques in a wide range of situations and conditions. Clearly, an accurate grouping of jobs into homogeneous subsets is an essential prerequisite for any type of validity generalization study. Thus, the resolution of various issues regarding the methods which are appropriate for grouping jobs can only serve to enhance future research on substantive issues.

Methods for Obtaining Job Families

The two approaches for grouping jobs which have received the most attention are cluster analysis and analysis of variance. The clustering algorithm developed by Ward (1963; Ward & Hook, 1963) was used by Mobley and Ramsey (1973), Tornow and Pinto (1976), Taylor (1978), and Taylor and Colbert (1978) to obtain clusters of similar jobs which might be appropriate for validity generalization. This technique has also been used in research on personnel assessment and classification (Brush & Owens, 1979; Feild & Schoenfeldt, 1975a; Owens, 1968, 1971; Schoenfeldt, 1974).

A serious weakness of cluster analysis has been the conspicuous absence of methods for evaluating the quality of obtained solutions. For instance, clustering algorithms are designed to find groupings of entities such that the entities within each cluster are more similar to each other than they are to the members of other clusters. However, these techniques do not include provisions for determining whether the similarities between entities within each cluster are statistically significant or whether the obtained cluster solution adequately and practically describes the structures and relationships in the data as a whole (i.e., as does R^2 in regression analysis). When these techniques are used for obtaining job families, the researcher runs the risk of clustering jobs which are actually dissimilar enough to warrant separate validity studies. If this were the case, it would produce an artificial expansion of the range of criterion scores leading to an overestimation of the validity coefficient (McIntyre, 1978).

The lack of significance tests of intracluster similarities also raises the issue of whether the use of cluster analysis is defensible from a legal standpoint. This issue led Arvey and Mossholder (1977) to propose the use of analysis of variance (ANOVA), using multiple raters for each job, as a means of obtaining job families. In the Arvey and Mossholder design, the levels of the first factor are the dimensions of a job analysis questionnaire such as the PAQ. The various jobs under consideration define the levels of the second factor, and the raters who provide ratings for each job on the various dimensions comprise a third factor. The first and second factors are fully crossed, but the third factor is nested within the second. The dependent variable, then, is the rating assigned to a given job on a particular dimension.

Several pertinent criticisms of the Arvey and Mossholder (1977) procedure were raised by McIntyre and Farr (1979), Hanser, Mendel, and Wolins (1979), and Lissitz, Mendoza, Huberty, and Markos (1979). One of the more important objections had to do with using dimensions of the PAQ as levels of an independent variable. As McIntyre and Farr (1979) and Lissitz et al. (1979) pointed out, the dimensions of the PAQ are qualitatively different and do not represent commensurate levels of a single independent variable. Based on this objection, Lissitz et al. (1979) discussed the use of a one-way multivariate analysis of variance (MANOVA) using jobs as levels of the independent variable and job dimensions as dependent variables. Arvey, Maxwell, and Mossholder (1979) also discussed a repeated measures approach to multivariate data but nevertheless maintained that the original

ANOVA design was appropriate when MANOVA could not be used (i.e., if the number of raters was not sufficiently greater than the number of jobs plus the number of dimensions).

From a practical standpoint, however, the most telling flaw of the ANOVA/MANOVA approach is the difficulty in grouping jobs in the event that significant differences between jobs are found to exist. In this case, the researcher must examine $N(N-1)/2$ pairwise differences (where N is the number of jobs) in order to determine which jobs may be grouped together. This becomes an unwieldy problem when even as few as 10 jobs are being considered. In the final analysis the ANOVA/MANOVA approach leaves the researcher with the task of performing, by visual inspection, what clustering algorithms were designed to do in the first place.

Issues in Cluster Analysis Applications

In view of the above discussion, some type of cluster analysis is clearly the preferred alternative for most applications. Nevertheless, there are a number of problems which arise in the application of cluster analysis to job analytic data. As stated above, clustering techniques will group "similar" jobs together but they do not provide a means for determining whether the similarity between jobs in a given cluster is statistically significant. Secondly, although a number of techniques have been suggested for determining the number of clusters which are present in the data, none of these are entirely satisfactory for all applications (Everitt, 1980; McIntyre, 1978).

The accuracy of a cluster analytic solution, or the degree to which a cluster solution matches the structure of the true solution, is also of critical importance. Related to the accuracy of a solution is the necessity of a means of assessing the stability of a solution across different samples. Accuracy and stability are analogous to the concepts of validity and reliability, respectively, in measurement theory, with stability being a necessary, but not a sufficient, condition for accuracy.

McIntyre (1978; McIntyre & Blashfield, 1980) has developed a method for assessing the stability of a cluster solution based on the cross-validation paradigm in regression analysis. The nearest centroid evaluation procedure assesses the agreement between cluster analyses of two independent samples by means of the kappa statistic (Cohen, 1960). Using computer-generated data, McIntyre and Blashfield (1980) found that the agreement kappa (a measure of agreement between solutions obtained from two different samples) was positively related to the accuracy kappa (a measure of agreement between the obtained solution and the true solution) when the data did not depart too much from normality and there was little overlap between clusters. Although further investigation of this procedure is necessary, it appears to be very promising.

A final problem in the application of cluster analysis is that of choosing the most appropriate technique for a given set of data. It is well known that different clustering methods will often yield different results on the same data. In addition, the accuracy of each method is likely to vary as a function of the characteristics of the data.

Review of Selected Clustering Methods

Previous investigations of the use of cluster analysis for grouping jobs have been restricted primarily to the use of the hierarchical method developed by Ward (1963). Ward's method is one of several agglomerative hierarchical methods which are so named because they successively group entities into clusters and clusters into larger clusters until a single cluster is obtained. This results in a hierarchical structure or taxonomy of nonoverlapping clusters. The feature which distinguishes Ward's method from other agglomerative methods is the criterion for fusing entities or clusters, namely, the minimum error sum of squares (ESS).

Several empirical studies have compared the accuracy of various hierarchical clustering methods (Cunningham & Ogilive, 1972; Rohlf, 1970; Sneath, 1966; Sokal & Rohlf, 1962) and found some support for the relative superiority of average linkage cluster analysis. However, in the definitive study to date, Blashfield (1976) compared the accuracy of four hierarchical clustering techniques, including single linkage, complete linkage, average linkage, and Ward's (1973) method in a monte carlo simulation. He found that, in general, Ward's method yielded the highest accuracy of the four techniques and he determined the extent to which various characteristics of the data affected the accuracy of each method.

Hierarchical techniques are most appropriate when there is interest in discovering a taxonomic structure in a set of data. Thus, such techniques may be less appropriate than others when interest is only in grouping jobs together in order to increase sample size for a validation study. In addition, hierarchical techniques are characterized by one important deficiency which is not shared by nonhierarchical techniques. That is, with the addition of each new entity to a cluster, there is a corresponding shift in the cluster centroid. As a result, entities which were included in previous steps may become more distant from the centroid of the parent cluster than that of a different cluster. Therefore, the use of Ward's method could often lead to heterogeneity within job families that could affect the obtained validity coefficients. In view of this deficiency, there may often be a need to reassign some jobs to different clusters after a solution has been obtained by means of Ward's method.

Feild and Schoenfeldt (1975b) have devised a method for "cleaning up" the clusters obtained by Ward's method which involves the use of a relocation procedure in which an entity is reassigned to another cluster if the distance to the centroid of that cluster is less than the distance to the centroid of the parent cluster. The relocation procedure is then followed by discriminant analysis to confirm the results and indicate final relocations.

It is misleading, however, to discuss this technique as a procedure for cleaning up the results obtained by Ward's method. The first part of the procedure is simply a nonhierarchical clustering algorithm similar to MacQueen's (1967) k-means method. The k-means technique is one of several partitioning or optimization techniques (Everitt, 1980). It involves an initial partitioning of the data set into k groups and the computation of the centroid of each group. After the initial partition has been set up, each entity is examined in turn to determine whether it is closer to the centroid of its parent cluster than it is to the centroid of every other cluster. If an entity is found to be closer to another cluster, it is reassigned to that cluster and the centroids of the losing and gaining clusters are recomputed. A number of relocations may occur during a single pass through the data and several passes may be required. The procedure terminates when a complete pass through the data does not result in a single relocation.

Wishart (1978) developed a variant of the k-means method which allows the user to specify a range of values for k and to identify and remove outliers (also a characteristic of the Feild and Schoenfeldt, 1975b, procedure). The initial partition may be a random assignment of N entities to k clusters or a partition specified by the user. Thus, the researcher has the option of using the results obtained from Ward's method to set up the initial partition, in which case the Wishart k-means technique is essentially the same as the Feild and Schoenfeldt relocation procedure, excluding the use of discriminant analysis. McIntyre (1978) discussed the fact that the initial relocation procedure of Feild and Schoenfeldt is biased toward giving seemingly useful discriminant analysis results, so its use may be unnecessary. It should be noted that the results obtained by the k-means method using a random initial partition could be very different from the solution reached by using Ward's method to select an initial partition. The k-means method, then, would appear to have advantage over Ward's method in allowing jobs to be reassigned to clusters and in dealing with outliers or jobs which should be dealt with individually.

The Ward and k-means techniques both seek to optimize ESS and are, therefore, biased toward finding spheroidal clusters regardless of the natural configuration of each cluster which actually exists in the data. A class of clustering methods known as density search techniques (Everitt, 1980) have been developed for finding natural clusters in the data. The assumption is that in the p-dimensional variable space defined by the data, there are regions of greater and lesser density of points (entities). As the name implies, density search techniques attempt to identify higher density regions or modes in the data.

Wishart's (1978) mode analysis is a density search technique which first examines the sphere of radius R surrounding each entity. Entities are then labeled as dense points or nondense points, depending upon whether the number of other points which are contained within the sphere exceeds a prespecified value. The value of R is gradually increased so that more entities become dense. As more entities become dense, they may initiate new clusters, join previously existing ones, or cause two clusters to fuse.

Mode analysis has been criticized for its failure to identify large and small clusters simultaneously (Everitt, 1980). Hierarchical mode analysis (Wishart, 1978) was developed to overcome this deficiency. Thus, hierarchical mode analysis would be preferred to mode analysis for most applications.

Research Approach and Hypotheses

The present study examined the accuracy of Ward's (1963) method, Ward's method followed by k-means clustering, the k-means method using a random initial partitioning, and hierarchical mode analysis. The first objective of the study was to determine whether any one of the four methods would consistently be more accurate than the others in clustering jobs. Secondly, the study was conducted to determine how each of the methods would be affected by various characteristics of job analytic data.

A monte carlo technique was used to generate data sets designed to simulate those which might be obtained from several job analyses. In all, 30 data sets, which varied according to certain variables relevant to job families research, were constructed in order to determine the degree to which the accuracy of each method covaried with these characteristics. These variables included the total number of jobs in the mixture, the number of populations in the mixture, the relative size of population clusters in the mixture, and the number of orthogonal job dimensions. The relationship between the accuracy of each method and the degree of separation between clusters was also examined. Although a number of other relevant variables may affect the accuracy of these techniques, those listed above were judged to be the most salient characteristics of job analytic data.

As stated previously, a significant problem with hierarchical techniques is that they do not allow for reassignment of jobs to different clusters during the clustering process. This problem would become more apparent as the total number of jobs in the data set increases. Based on this observation, it was expected that the accuracy of Ward's technique would be negatively correlated with the total number of jobs in the data sets. Also, as the two variations of the k-means technique have the advantage of reassigning jobs to clusters, it was hypothesized that they would generally yield more accurate results than Ward's method.

Finally, whereas hierarchical mode analysis attempts to discover natural clusters in the variable space, the Ward and k-means techniques are biased toward finding spheroidal clusters. It was expected, therefore, that the results obtained by hierarchical mode analysis would differ from those of the Ward and k-means techniques.

Method

Generation of Data Sets

A mixture model approach (Blashfield, 1976; McIntyre & Blashfield, 1980; Wolfe, 1970) was used in generating the data. Briefly, the mixture model proposes that a sample of entities (in this case, jobs) is a mixture of samples from k multivariate populations, each having its own mean vector and covariance matrix.

The subroutine GGNSM of the International Mathematical and Statistical Library (IMSL) was used to generate samples from multivariate normal populations for each of the 30 mixtures used in the study. In each case, a $p \times p$ identity matrix was used as the variance-covariance matrix in order to simulate scores on orthogonal job dimensions which would have been obtained as a result of a principal components analysis of job analytic data. It could be argued that the use of an identity matrix as the population variance-covariance matrix limits the validity of the simulation model, since orthogonality is rarely encountered in practice. In such cases, however, the researcher may use the Mahalanobis D^2 metric, which takes into account the covariance structure between the variables. Such practice would be equivalent to the use of squared Euclidean distance on orthogonal data. The values for the population mean vectors were randomly selected from a uniform distribution ranging from 4.0 to 6.0.

Independent Variables

The literature on job families was reviewed in order to select a range of values for the variables which were used in constructing the mixtures. The number of jobs (N) for a given mixture was randomly selected from the interval (10,200). The number of populations (P) in the mixture (based on the number of clusters obtained in previous studies) was chosen from the interval (2,13). This variable was varied systematically, so that the size of the clusters in each mixture would not be too small. Next, the number of orthogonal dimensions (D) for which scores were generated for each mixture was randomly selected from the interval (1,27). Finally, cluster size was systematically varied across mixtures, so that some mixtures contained large and small clusters, while others contained clusters approximately equal in size. The sum of the absolute values of deviations from the mean cluster size (for a given mixture) was used as a measure of the differential size of clusters within mixtures.

The degree of overlap among clusters was measured by Rao's R, a linear transformation of Wilk's lambda (Tatsuoka, 1971). The formula for R is given by:

$$R = [(1-\lambda^{1/S})/\lambda^{1/S}]\{[ms-p(k-1)/2+1]/p(k+1)\}, \qquad [1]$$

where

λ is obtained value of Wilk's lambda,
p is the number of variables,
k is the number of groups,
$m = N - 1 - (p + k)/2$,

and

$$s = [(p^2(k-1)^2 - 4)/(p^2 + (k-1)^2 - 5)]^{.5}.$$

Although R was not systematically manipulated, it was, nevertheless, treated as an independent variable for the purposes of this study.

A number of other relevant variables may affect the accuracy of the clustering techniques used in this study, but the ones listed above were judged to be the most salient characteristics of job analytic data. For this reason, the present study was limited to examining the effects of these variables.

Cluster Analysis

The various cluster analyses were performed using Wishart's (1978) CLUSTAN 1c package. The dimension scores were standardized prior to clustering, and the squared Euclidean distance was used as the measure of dissimilarity. The use of distance measures for assessing similarity/dissimilarity between job profiles has been criticized by Hamer and Cunningham (1981). However, the decision to use the squared Euclidean distance metric reflects common practice in the job families literature.

The problem of choosing the number of clusters in the final cluster solution is a difficult one. As yet, no completely satisfactory solution has been reached for all applications (Everitt, 1980). However, as this problem was not within the scope of the present study, it was decided that the number of terminal clusters for each analysis would equal the number of population clusters in the mixture.

Accuracy of Cluster Solutions

The statistic c (Rand, 1971) was used to measure the accuracy of cluster solutions, or the degree of similarity between the obtained solution and the actual solution. In prior investigations of this type (Blashfield, 1976; McIntyre, 1978; McIntyre & Blashfield, 1980), the statistic kappa has been used as a measure of accuracy. The use of this statistic requires that the researcher must somehow match each cluster in the obtained solution to one of the populations in the mixture. This matching procedure presents a problem, however. If it is done on a random basis, then the degree of accuracy may be severely underestimated. If, however, it is decided to match each cluster to the most appropriate population in the mixture, this serves to inflate the size of the obtained kappa. As no such explicit matching is required in the calculation of Rand's c, it was deemed to be more appropriate for this study.

The similarity, c, between two cluster solutions, Y and Y', for the same data is given by:

$$c(Y,Y') = (N(N-1)/2 - \{1/2[\sum_i(\sum_j n_{ij})^2 + \sum_j(\sum_i n_{ij})^2] - \sum_{ij}\sum_{ij} n_{ij}^2\})/[N(N-1)/2], \qquad [2]$$

where n_{ij} is the number of points simultaneously in the i^{th} cluster of Y and the j^{th} cluster of Y'. The statistic c ranges from 0 (inaccurate solution) to 1 (perfect solution).

Analyses

A one-way analysis of variance was computed, followed by a set of Bonferroni contrasts, to examine the differences between the mean accuracy values of the four methods. In addition, zero-order correlations were computed between the five independent variables and the c values for each method.

Results

The means and standard deviations of the c values for each method are shown in Table 1. A one-way analysis of variance yielded an F of 16.03 ($p < .0001$, $df = 3, 116$) with an associated omega

Table 1
Means and Standard Deviations of c(Y,Y´) Values for Four Cluster Analysis Methods

Method	Mean	Standard Deviation
Hierarchical Mode Analysis	.66	.19
Ward´s Method	.85	.13
K-Means (Random Start)	.86	.14
Ward/K-Means	.90	.10

squared of .29. Four Bonferroni contrasts were constructed to examine the differences in accuracy among the respective methods. These comparisons are shown in Table 2. As expected, the performance of hierarchical mode analysis did differ significantly from that of the other three methods. However, the mean c values obtained by Ward's method and the two variations of the k-means technique were not found to differ significantly from one another.

Zero-order correlations were computed in order to determine the effects of the five independent variables on the accuracy of the various methods. These are reported in Table 3. The pattern of correlations suggests that only the first three variables, number of jobs, number of populations, and number of dimensions in the mixtures had a significant impact on the accuracy of the four methods.

The relationships between accuracy and the independent variables of number of job dimensions, number of jobs, and number of populations are graphically displayed in Figures 1, 2, and 3, respectively. These figures are included to depict the observed relationships between the various cluster methods and the independent variables across the range of independent variable values. These figures should be interpreted with caution in light of the nonsignificant differences among three of the four methods and the nonsignificant correlations.

Discussion

Conclusions

The hypothesized negative relationship between the accuracy of Ward's method and the number of jobs in the mixtures was not supported by the data. This may have been due to the upper limit of

Table 2
Bonferroni Comparisons of the Accuracy of Four Methods of Cluster Analysis

Method*	t	p
M vs. (W + K + WK)/3	6.78	<.05
W vs. K	.16	>.05
W vs. WK	1.35	>.05
K vs. WK	1.20	>.05

*M = Hierarchical Mode Analysis, W = Ward´s Method, K = K-Means (Random Start), WK = Ward/K-Means.

Table 3
Zero-order Correlations Between c(Y,Y′) for Each Method and Five Variables Characterizing the Mixtures

Variable	Method: Hierarchical Mode Analysis	Ward′s Method	K-Means (Random Start)	Ward/ K-Means
Number of Jobs	.04	.10	.54**	.07
Number of Populations	.40*	.31	.43*	.12
Number of Orthogonal Dimensions	-.01	.48**	.20	.50**
Rao′s R	.15	-.26	-.09	-.03
Sum of Absolute Deviations from Mean Cluster Size	.01	.07	.26	-.07

Note: Sample size equals 30 for all relationships except those involving Rao′s R, in which case the sample size equals 28.
* significant at p<.05
** significant at p<.01

200 on the number of jobs. In any case, it seems safe to conclude that the deficiency of not allowing for reassignment of entities to clusters does not pose any serious problems with data sets consisting of 200 or fewer observations.

Figure 1
Regression of Cluster Solution Accuracy (*c*) on Number of Job Dimensions for Four Cluster Analysis Methods

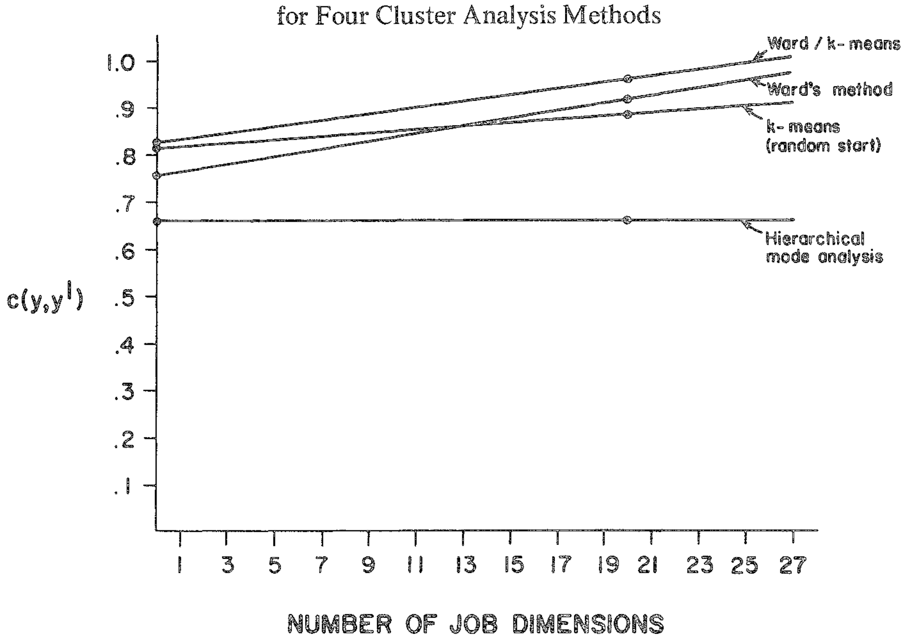


Figure 2
Regression of Cluster Solution Accuracy (c) on Number of Jobs for Four Cluster Analysis Methods

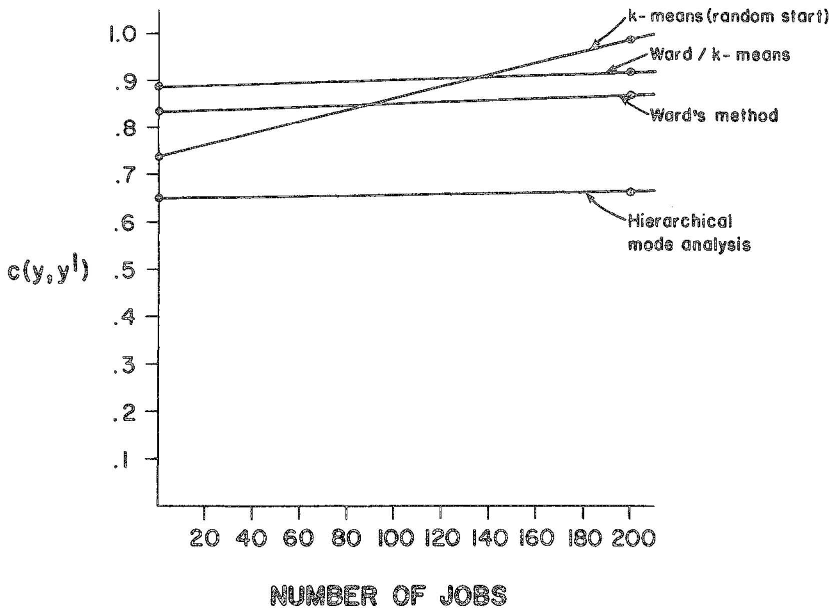
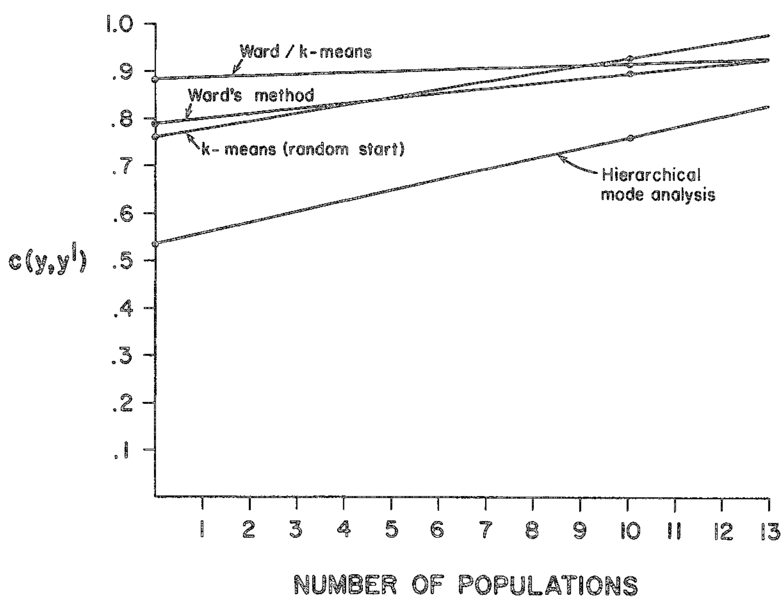


Figure 3
Regression of Cluster Solution Accuracy (c) on Number of Populations for Four Cluster Analysis Methods



The second hypothesis, which was related to the first, was that the two variations of the k-means technique used in this study would yield more accurate results than Ward's method. Although the differences between the means were in the hypothesized direction, they were not found to be significant. One explanation of these results is that, since all three methods use the same clustering criterion (i.e., the minimization of ESS), they might not be expected to differ, to any great extent, in terms of accuracy. Perhaps this is especially so with small to moderate size data sets. An alternative explanation would be that, due to the limited number of mixtures generated, there was insufficient power to allow for a sufficient test of the hypothesis. However, the power for the test between the Ward and Ward/k-means techniques (which was the largest of the three effects) was .84. Thus, the probability was quite high of finding a difference between these techniques if it did, in fact, exist.

The third hypothesis, that the results of hierarchical mode analysis would differ from those of the other three methods, was confirmed. In fact, hierarchical mode analysis was far less accurate than the other techniques. This suggests the possibility that the search for natural clusters may not be appropriate for this type of application. However, other density search techniques that use different procedures for identifying natural clusters might have yielded better results. For instance, the cartet count method of Cattell and Coulter (1966) partitions the variable space into several hypercubes (cartets) and counts the number of points in each hypercube. Wolfe's (1970) method of mixtures, on the other hand, first obtains maximum likelihood estimates of the mean vectors and covariance matrices for each population in the mixture. Entities are then assigned to clusters on a probablistic basis. Since there is little similarity between these methods and hierarchical mode analysis, they might be expected to yield very different solutions.

Definition of a Cluster

The relatively poor performance of the density search technique, which was used in this study, brings up a problem that has received little attention in the psychological measurement literature, namely, the definition of a cluster. This is an important consideration, as different clustering methods define clusters in different ways. The Ward and k-means methods, for example, are biased toward finding spheroidal clusters regardless of the natural configurations of the clusters in a given data set. Methods such as single linkage cluster analysis often yield elongated clusters. In contrast, density search techniques try to find regions of greater and lesser density of entities in the variable space. Another set of methods, referred to as clumping techniques by Everitt (1980), allows for overlapping clusters.

There are, then, several competing definitions of what constitutes a cluster. In view of practical and legal considerations in selection research, this question deserves some consideration. For this type of application (i.e., clustering jobs for purposes of validity generalization), the most reasonable course of action may be to use a method which does not allow for overlapping clusters and which is biased toward finding spheroidal clusters. Using spheroidal clusters of jobs to increase sample size would insure a greater degree of homogeneity among the jobs in a validation study, thus yielding a more accurate estimate of the true validity coefficient. Also, from a legal standpoint, this approach would probably be the most defensible course of action.

The appropriate definition of a cluster should likewise be examined for other areas of psychological research which make use of cluster ananlysis. These considerations should be carefully examined, along with empirical findings, in selecting an appropriate method for a given application.

Recommendations for the Use of Clustering Methods

The results of this study suggest certain recommendations for applying the four techniques used in this study. Obviously, any of the three techniques which minimizes ESS would be preferred over hierarchical mode analysis. Also, as noted previously, Blashfield (1976) found Ward's method to be more accurate, overall, than three other popular hierarchical methods. Considering the results of this study, in light of those obtained by Blashfield, the Ward/k-means technique and k-means technique using a random start would be expected to perform better than single, average, or complete linkage in most instances. Additionally, it might also be expected that these two methods would be similarly affected by some of the independent variables which had an effect on Ward's method in Blashfield's study. In particular, the accuracy of these methods would be expected to be low if the true cluster configuration were ellipsoidal.

Unfortunately, the lack of significant differences among the three clustering techniques precludes any further prescriptions regarding which technique might be preferred under various properties of the data. Although some trends seemed to emerge, future research will have to focus on the subtle differences between the three ESS minimization procedures.

Limitations and Implications for Future Research

There are several limitations in the present study which should be considered in interpreting the results. Most of these have to do with the validity of the simulation model. First, the relatively small number of mixtures used in the study may not have been sufficient for the correlational analyses. Power values for nonsignificant correlations ranged from .01 to .37. In addition, 30 mixtures may not have adequately represented the sample space. Also, the upper limit of 200 on the number of simulated jobs in the mixtures may have served to mask the effect of not allowing relocation of jobs to different clusters when using Ward's method. More importantly, the generation of data samples from multivariate normal populations, rather than data whose marginal distributions were modeled after those of real job analytic data, might be problematic. Finally, although the degree of cluster separation was used as an independent variable, it was not systematically varied. Thus, the range of values on this variable was, no doubt, restricted. If this variable had been systematically varied or if the difficulty level of the mixtures had been varied in some other manner, a significant difference might have been obtained between Ward's method and the two variations of the k-means technique.

Two additional limitations have to do with decisions regarding the clustering phase of the study. As noted earlier, the squared Euclidean distance metric was used rather than a host of other possible similarity/distance measures. The decision to use this measure does not imply that it is the most appropriate one for every application. Also, the decision to make the number of clusters equal to the number of populations in the mixture may have affected the results. Since hierarchical mode analysis attempts to find a set of "natural" clusters, this decision rule may have served to handicap the technique to some extent.

Indeed, there is need for further research on the problem of choosing the number of clusters for a clustering solution and for determining the most appropriate measure of similarity (or dissimilarity) for various applications. Future research of this type should attempt to correct for the limitations of the present study, in order to more effectively model job analytic data. Also, as was mentioned previously, the ability of a particular clustering algorithm to obtain clusters which are relatively stable across different samples is closely related to the accuracy of that method. The investigation of stability would require a more sophisticated research design, involving repeated sampling of the same mixtures, and thus was not incorporated into this study. In addition, it would be important to learn how sensitive these techniques are to unreliability in the data, a prevalent problem with data from job analysis questionnaires.

Another important problem is that of outliers in the data. In research of this type, there may often be one or several jobs which do not fit in any of the clusters very well. It would seem that partitioning techniques would be more able to identify these outliers than hierarchical ones, but this hypothesis should be tested empirically. A related problem is that of adding several jobs to a set of previously existing job clusters. In some cases, such additions could drastically alter the clustering scheme. Thus, it would be important to determine whether the techniques used in this study differ in terms of their ability to obtain a solution which is relatively stable with the addition of new jobs.

Further investigation of these issues would have important implications for various fields of psychology in which cluster analysis is widely used. In particular, personnel psychology would benefit from such knowledge, as it would help to insure the degree of homogeneity within job clusters that is an essential prerequisite for validity generalization research.

References

Arvey, R. D., Maxwell, S. E., & Mossholder, K. M. Even more ideas about methodologies for determining job differences and similarities. *Personnel Psychology,* 1979, *32,* 529-538.

Arvey, R. D., & Mossholder, K. M. A proposed methodology for determining similarities and differences among jobs. *Personnel Psychology,* 1977, *30,* 363-374.

Blashfield, R. K. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin,* 1976, *83,* 377-388.

Brush, D. H., & Owens, W. A. Implementation and evaluation of an assessment classification model for manpower utilization. *Personnel Psychology,* 1979, *32,* 369-383.

Cattell, R. B., & Coulter, M. A. Principles of behavioural taxonomy and the mathematical basis of the taxonome computer program. *British Journal of Mathematical and Statistical Psychology,* 1966, *19,* 237-269.

Cohen, J. A coefficient of agreement of nominal scales. *Educational and Psychological Measurement,* 1960, *20,* 37-46.

Colbert, G. A., & Taylor, L. R. Empirically derived job families as a foundation for the study of validity generalization. Study 3. Generalization of selection test validity. *Personnel Psychology,* 1978, *31,* 355-364.

Cunningham, K. M., & Ogilive, J. C. Evaluation of hierarchical grouping techniques—A preliminary study. *Computer Journal,* 1972, *15,* 209-215.

Everitt, B. S. *Cluster analysis* (2nd ed.). New York: Halstead Press, 1980.

Feild, H. S., & Schoenfeldt, L. F. Development and application of a measure of students' college experiences. *Journal of Applied Psychology,* 1975, *60,* 491-497. (a)

Feild, H. S., & Schoenfeldt, L. F. Ward and Hook revisited: A two-part procedure for overcoming a deficiency in the grouping of persons. *Educational and Psychological Measurement,* 1975, *35,* 171-173. (b)

Hamer, R. M., & Cunningham, J. W. Cluster analyzing profile data confounded with interrater differences: A comparison of profile association measures. *Applied Psychological Measurement,* 1981, *5,* 63-72.

Hanser, L. M., Mendel, R. M., & Wolins, L. Three flies in the ointment: A reply to Arvey and Mossholder. *Personnel Psychology,* 1979, *32,* 511-516.

Lissitz, R. W., Mendoza, J. L., Huberty, C. J., & Markos, H. V. Some further ideas on a methodology for determining job similarities/differences. *Personnel Psychology,* 1979, *32,* 517-528.

MacQueen, J. Some methods for classification and analysis of multivariate observations. In J. Neyman (Ed.), *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1). Berkeley: University of California Press, 1967.

McIntyre, R. M. *The evaluation of a cluster analytic procedure for grouping jobs.* Unpublished doctoral dissertation, Pennsylvania State University, 1978.

McIntyre, R. M., & Blashfield, R. K. A nearest-centroid technique for evaluating the minimum-variance clustering procedure. *Multivariate Behavioral Research,* 1980, *15,* 225-238.

McIntyre, R. M., & Farr, J. L. Comment on Arvey and Mossholder's "A proposed methodology for determining similarities and differences among jobs." *Personnel Psychology,* 1979, *32,* 507-510.

Mobley, W. H., & Ramsey, R. S. Hierarchical clustering on the basis of inter-job similarity as a tool in validity generalization. *Personnel Psychology,* 1973, *26,* 213-225.

Owens, W. A. Toward one discipline of scientific psychology. *American Psychologist,* 1968, *23,* 782-785.

Owens, W. A. A quasi-actuarial basis for individual assessment. *American Psychologist,* 1971, *26,* 992-999.

Pearlman, K. Job families: A review and discussion of their implications for personnel selection. *Psychological Bulletin,* 1980, *87,* 1-28.

Rand, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association,* 1971, *66,* 846-850.

Rohlf, F. J. Adaptive hierarchical clustering schemes. *Systematic Zoology,* 1970, *19,* 58-82.

Schmidt, F. L., & Hunter, J. E. Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology,* 1977, *62,* 529-540.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. Further tests of the Schmidt-Hunter Bayesian validity generalization procedure. *Personnel Psychology,* 1979, *32,* 257-281.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. Statistical power in criterion-related validity studies. *Journal of Applied Psychology,* 1976, *61,* 473-485.

Schoenfeldt, L. F. Utilization of manpower: Development and evaluation of an assessment classification model for matching individuals with jobs. *Journal of Applied Psychology,* 1974, *59,* 583-595.

Sneath, P. H. A. A comparison of different clustering methods as applied to randomly spaced points. *Classification Society Bulletin,* 1966, *1,* 2-18.

Sokal, R. R., & Rohlf, F. J. The comparison of dendograms by objective methods. *Taxon,* 1962, *11,* 33-40.

Tatsuoka, M. M. *Multivariate analysis: Techniques for educational and psychological research.* New York: John Wiley & Sons, 1971.

Taylor, L. R. Empirically derived job families as a foundation for the study of validity generalization. Study 1. The construction of job families based on the component and overall dimensions of the PAQ. *Personnel Psychology,* 1978, *31,* 325-340.

Taylor, L. R., & Colbert, G. A. Empirically derived job families as a foundation for the study of validity generalization. Study 2. The construction of job families based on company-specific PAQ job dimensions. *Personnel Psychology,* 1978, *31,* 341-353.

Tornow, W. W., & Pinto, P. R. The development of a managerial job taxonomy: A system for describing, classifying, and evaluating executive office positions. *Journal of Applied Psychology,* 1976, *61,* 410-418.

Ward, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association,* 1963, *58,* 236-244.

Ward, J. H., & Hook, M. E. Application of an hierarchical grouping procedure to a problem of grouping profiles. *Educational and Psychological Measurement,* 1963, *23,* 69-81.

Wishart, D. *Clustan 1c user manual,* (3rd ed.). Edinburgh, Scotland: Edinburgh University, Program Library Unit, 1978.

Wolfe, J. H. Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research,* 1970, *5,* 329-350.

Acknowledgments

The authors thank the office of the Associate Dean for Research, College of Liberal Arts, for financial support of this project.

Author's Address

Send requests for reprints or further information to Rick Jacobs, Department of Psychology, The Pennsylvania State University, University Park PA 16802.