OMB APPROVED: 1121-0011, EXPIRES 12/31/83

U. S. DEPARTMENT OF JUSTICE
Office of Justice Assistance, Research, and Statistics

CATEGORICAL GRANT PROGRESS REPORT

This recordkeeping requirement falls under the authority of P.L. 96-511, Sec. 3507. The information provided will be used by grant monitors to track grant progress. No further monies or other benefits may be paid out under this program unless this report is completed and filed as required by existing laws and regulations (OMB Circulars A-102 and A-110; Omnibus Crime Control and Safe Streets Act of 1968, as amended; Juvenile Justice and Delinquency Prevention Act of 1974, as amended; and the Justice System Improvement Act of 1979, as amended).

1. GRANTEE: University of Pittsburgh
2. AGENCY GRANT NUMBER: 81-IJ-CX-0059
3. REPORT NO.: 5
4. IMPLEMENTING SUBGRANTEE:
5. REPORTING PERIOD (Dates) FROM: 6/30/82 TO: 9/30/82
6. SHORT TITLE OF PROJECT: Increasing the Statistical Power
7. GRANT AMOUNT: $32,899.00
8. TYPE OF REPORT: ☐ REGULAR ☐ SPECIAL REQUEST ☒ FINAL REPORT
9. NAME AND TITLE OF PROJECT DIRECTOR: Craig S. Edelbrock, Ph.D. Asst. Prof. of Psychiatry
10. SIGNATURE OF PROJECT DIRECTOR: Craig S. Edelbrock
11. DATE OF REPORT: 10/12/82
12. COMMENCE REPORT HERE (Continue on plain paper)

Computer evaluations of inverse factor analysis, Lorr's non-hierarchical clustering technique, and centroid clustering were completed on 20 multivariate normal mixtures generated by Blashfield's 1976. Details regarding both within- and between-method compared comparisons are provided in the following final report. Overall, our research has documented a strong relationship between level of coverage and both accuracy of clustering solutions and the statistical power of cluster based classifications. For each of the three clustering procedures, 100% coverage resulted in less than optimal accuracy in recovering underlying populations from the computer generated mixtures. For all three methods, the accuracy of clustering solutions was substantially increased by leaving 15-25 percent of the subjects unclassified. Our results suggest that accuracy of clustering solutions can be maximized in the range of 55-85% coverage. For all methods we tested, increasing coverage above 85% had deleterious effects on clustering accuracy.

90476

...tical power analyses further indicate that the level of coverage of ...ions is strongly related to the probability of detecting significant correlates membership. For the three methods we tested, 100% coverage produced less ...l probabilities of detecting significant differences among clusters. Instead, power was maximized in the range of 55-85% coverage. Statistical power ...d at levels above 85% coverage and levels below 55% coverage. These results typologies having levels of coverage in the medium range will be optimally ...of external criteria.

...sons among the three methods indicate that clustering procedures are not ...n terms of their accuracy in grouping objects from the computer generated ...or in terms of predictive power. Lorr's non-hierarchical clustering ...rformed exceptionally well in both respects and should be given serious ...n in future clustering efforts. Results on the inverse factoring procedures ...couraging, considering that this method has been widely criticized as a ...cedure. The centroid clustering procedure performed exceptionally well in ...uracy and statistical power. Overall, this method produced the highest levels ...and statistical power for the computer generated data sets we tested.

...F RECEIPT BY GRANTEE CRIMINAL JUSTICE COUNCIL (Official signature)
14. DATE

...V. 4-81)

Efforts to apply these clustering methods to real data regarding juvenile delinquency were not successful. Several clustering methods were applied to this data set, but we were unable to identify homogeneous subgroups of delinquent youth. This is not so much a failure of the clustering methods as an indication that this data set does not warrant cluster analyses. We are currently seeking alternative data sets upon which to further test these clustering methods.

Increasing the Statistical Power of Empirically Derived Taxonomies in Criminal Justice Research

Final Report

Grant # 81-IJ-CX-0059

Submitted to: Office of Research and Evaluation Methods
National Institute of Justice
633 Indiana Avenue
Washington, D.C. 20531

Submitted by: Craig Edelbrock, Ph.D., Principal Investigator
University of Pittsburgh
3811 O'Hara Street
Pittsburgh, PA 15213
(412) 624-0448

TABLE OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

INCREASING THE STATISTICAL POWER OF

EMPIRICALLY DERIVED TAXONOMIES IN CRIMINAL JUSTICE RESEARCH

I. Classification in Criminal Justice

A. Importance of Classification

Classification is of central importance in criminal justice, so much so that the evolution of the field has paralleled the development of more diversified and refined taxonomies of criminal offenses and offenders (Clinard and Quinney, 1967; Ferdinand, 1966; Gibbons, 1975; Warren, 1970). Classifications serve many purposes in criminological research, theory, and practice including the reduction and ordering of the complex phenomena of deviance and the provision of conceptual frameworks for decision making. The role of classification in summarizing complex data is particularly relevant to criminological research and theory construction. Glaser (1974), for example, has argued that the development of reliable and valid classifications is essential to improve the epidemiological mapping of crime, the evaluation of treatment programs, and the explanatory value of theories of criminal behavior. The importance of classffications is further underscored by their use in decision making in all phases of the criminal justice process.

B. Inadequacy of Current Classifications

Despite the importance of classification in criminal justice, there is consensus regarding the inadequacy of current taxonomies. The entire classification enterprise has been assailed for its lack of cumulative and convergent findings and failure to produce taxonomies having practical or theoretical utility (Ferdinand, 1966; Gibbons, 1975; Hood and Sparks, 1970; Opp, 1973). Criminological taxonomies have been critcized for their subjectivity, inadequate reliability, illogical structure, ambiguous nomenclature, impracticality, and lack of predictive validity. The lack of predictive power is perhaps the most debilitating criticism of criminological taxonomies because of the

implications this carries regarding decision making. If classifications are not predictive of the etiology of criminal acts, offender personality, behavior, risk, recidivism or treatment response, their value in decision making regarding the ajudication, sentencing, management, parole, and release of offenders is severely limited.

This presents a disparaging view of the current state of criminological classification, but the situation is not necessarily any better in the other areas of the behavioral and social sciences. The most widely used taxonomy of psychiatric disorders, for example, is that embodied in the Diagnostic and Statistical Manual of the American Psychiatric Association (APA, 1968, 1980). Although this is taken to represent the state-of-the-art taxonomy in psychiatry, it has been widely criticized by psychiatrists, and other mental health professionals on the same grounds as criminological taxonomies (cf. Achenbach and Edelbrock, 1978; Phillips and Draguns, 1971; Zigler and Phillips, 1961). Numerous additional examples of the inadequacies of current classifications can be drawn from the literature of psychology, education, psychiatry, and sociology.

C. Promise of Taxometric Methods

As much as social and behavioral scientists agree regarding the shortcomings of current classifications, there is also agreement among professionals in all fields regarding the potential of numerical taxometric methods to overcome these inadequacies. These methods, variously known as numerical taxonomy, cluster analysis, association analysis, and pattern recognition, have been used in the biological sciences for many years (i.e., Sneath, 1957) but only recently have been added to the methodological armamentarium of social and behavioral scientists. Blashfield (1977), for example, in a comprehensive review of the use of taxometric methods, cited a veritable "explosion" in the use of numerical clustering and classification methods in the social and behavioral sciences since 1970.

The ability of these methods to summarize and order complex multivariate data has made them a valuable tool for the construction of taxonomies. (For reviews of the myriad applications of taxometric methods in various disciplines see Anderberg, 1975; Bailey, 1974; Blashfield, 1976; Sneath and Sokal, 1973). Numerous researchers in criminal justice have recognized the value of taxometric methods for data description, reduction, and management; and experimental design and evaluation research. Brennan (1979) has compiled a bibliography of criminological studies employing multivariate taxometric methods. A wide variety of methods have been used in these studies including hierarchical cluster analysis (Megargee, 1977), Lorr's non-hierarchical clustering technique (Blackburn, 1971), inverse factor analysis (Butler and Adams, 1966; Collins, Burger, and Taylor, 1976) and iterative K-means analysis (Brennan, Huizinga and Elliot, 1978).

Taxometric methods show great promise for the construction of valid and reliable taxonomies of criminal offenses and offenders. However, the use of these methods in criminal justice research has just begun an many difficulties have yet to be ironed out. Several problems stem from the fact that many of these methods have been adopted from the biological sciences. This has far reaching implications, not the least of which is that some taxometric methods are <u>conceptually</u> and <u>methodologically</u> inappropriate for criminal justice applications. Thus, in order to derive more useful and predictive taxonomies in criminology and criminal justice, it may be necessary to develop innovative methods tailored to applications in the social and behavioral sciences.

II. The Issue of Coverage

A. Role of Behavioral Taxonomies

The value of any taxonomy is related to its <u>coverage</u>, which is the proportion of subjects it can classify. In the biological sciences, where many taxometric methods were developed, 100% coverage is an important goal because taxonomies are intended to correspond to definitive classifications of biological species. Such taxonomies represent "real" groupings which have been validated against definitive criteria, such as morphological, physiological, and genetic characteristics, that are reliably assessed and have unquestioned validity as taxonomic criteria. In criminal justice, and other sciences in an earlier "natural history" stage of development, there is a lack of definitive criteria against which to validate empirically derived taxonomies. Thus, rather than constructing definitive taxonomies, taxometric methods are used as a <u>heuristic device</u> for summarizing complex relationships. In most applications, classifying everybody is not necessary, possible, or even desirable. It is recognized, for example, that there is a diversity of causes and modes of expression of criminal behavior, and that the personality and behavioral measures available for deriving taxonomies are not perfectly reliable or valid. Thus, it is not always possible to reliably classify all offenders into categories. The issue of coverage, however, extends beyond the simple fact that some subjects cannot be classified.

B. Bootstrapping - An Example

Owing to the lack of definitive criteria, most attempts to validate taxonomies in the behavioral and social sciences involve "bootstrapping--whereby investigators attempt to "lift themselves by their own bootstraps" by relating taxonomies to other measures known to be imperfect. Megargee, for example, has identified ten types of criminal offenders based on their Minne-

sota Multiphasic Personality Inventory (MMPI) Profiles (Megargee, 1977). The identification of profile types is not an end product in itself.
As Megargee stated:

"The MMPI-based taxonomy is worthless unless it can be established that the ten MMPI-defined groups differ significantly in other respects". (Megargee and Bohn 1977: p. 150).

Thus, in an exemplary fashion Megargee and his coworkers have proceeded to determine how their types differ in demographic characteristics, academic performance, intellectual ability, and social, developmental, and personality characteristics (Megargee and Bohn, 1977). Moreover, they have extended their research to determine how the types differ in long-term prognosis, recidivism, and differential response to treatment.

The value of the Megargee typology, therefore, does not lie in the identification of types but rather in the degree to which the typology relates to other criteria--particularly criteria that are informative regarding possible predisposing causes of criminal acts, management and treatment of offenders, and treatment outcomes. It is at this step of relating typologies to such criteria that the issue of coverage becomes important because of its effect on statistical power.

C. Coverage and Statistical Power.

The ability to detect significant differences among empirically derived groups is a complex function of the number of groups, sample size, separation and homogeneity of groups, and the size of effects under study. In the behavioral and social sciences, some assumptions can be made which simplify this complex set of interrelations. Most empirically derived taxonomies deal with relatively few types (i.e., ten for the Megargee taxonomy) and researchers can generally obtain samples large enough to permit rigorous statistical

analyses within each group. Furthermore, many of the effects examined in the behavioral and social sciences are small.

Whereas there is no absolute rule for defining "small", "medium", or "large", effect sizes, Cohen (1977) has suggested that a small effect size refer to a difference accounting for less than 10% of the variance in a variable; a medium effect size accounts for 20-25% of the variance; and a large effect size accounts for 40-50% of the variance. In criminal justice research, and other areas of the behavioral and social sciences, most effects are small due to less than perfect reliability and validity of the measures, multiple causation, and the lack of experimental control of many impinging sources of variation. This is not to say that small effects are trivial--quite the opposite. Some of the "small" effects in criminal justice research, such as differences in long-term recidivism rates among groups, are of the utmost importance in decision making.

If we assume that samples are sufficiently large relative to the number of groups and that the effects under study are relatively small, statistical power when validating taxonomies is largely a function of the degree to which the groups are distinct from each other, yet homogenous. As a statistician might phrase it, statistical power is optimized when the within-group variance on the taxonomic criteria is small and the between group variance is large. It is important to emphasize that statistical power in this context refers to the ability to detect differences in external criteria, that is, criteria not included in the construction of the taxonomy.

In a taxonomy, small within group variance implies that groups are similar to one another. Large between group variance implies that the types, and hence the groups representing the types, are distinct from one another on the taxonomic criteria. The ideal taxonomy would, therefore, group subjects into homogeneous, yet distinct groups. Such groups would be likely to differ in external criteria.

Unfortunately, subjects do not align themselves into stereotypical "pure types". In the Megargee taxonomy, for example, only 63% of the MMPI Profiles could be classified using computerized classification rules (Megargee and Dorhout, 1977). The remaining 37% were more difficult to classify because they had invalid scores, did not resemble any types, or met inclusion criteria for more than one type. Eventually, however, a total of 96% of the profiles were classified on the basis of clinical inspection and judgement.

D. The Continuum of Classifiability

A high degree of coverage is an important taxonomic goal, but classifying everybody is not always necessary. Moreover, for those sciences in a "boot-strapping" stage of development, attempting to classify everybody may have deleterious effects on the statistical power and reliability of the taxonomy. That is, in most classification efforts in the behavioral and social sciences, there is a "continuum of classifiability." At one end of this continuum are subjects who are easy to classify because they bear close resemblance to empirically derived types. In the Megargee taxonomy, for example, 63% of the sample could be classified on the basis of operationalized classification rules. Moving towards the middle of the continuum, subjects become more difficult to classify because they do not resemble pure types or resemble more than one type. Thus, in the Megargee taxonomy, 33% could not be classified by operationalized classification rules but required clinical inspection and judgment. At the opposite end of this continuum are those who cannot be classified because of invalid scores or lack of resemblances to any of the types (i.e., the 4% unclassified in the Megargee taxonomy).

Thus attempting to classify everybody requires the classification of subjects who are difficult or impossible to assign to groups. As one moves down the continuum classifying subjects who are less and less similar to the "types" the within group variance increases while the between group variance decreases. In other words, the groups become more heterogeneous and begin to overlap. This, of course, dilutes the statistical power of comparisons among groups. Moreover, the continuum of classifiability parallels a "continuum of reliability". Thus, moving down the continuum from pure types to unclassifiable subjects, the reliability of assignment decreases. Subjects resembling pure types can be reliably classified because slight changes on their classification criteria (i.e., behavioral or personality scores) do not substantially alter their similarity to the types. On the other hand, among subjects who are less similar to types, or resemble more than one type, a slight change in scores may result in a different group assignment.

The Thesis of This Research

The thesis of this research was that in criminal justice and other behavioral and social sciences, classification should be viewed as a continuum rather than as a purely discrete phenomenon (as in the biological sciences). Thus, based on the continuum of classifiability, the coverage of classifications can be varied to fit the purposes of the research. In an epidemological study, for example, the goal may be to classify as many subjects as possible, including subjects who are difficult to assign to groups. This procedure results in more heterogeneous groups and decreased reliability of classification but serves a major purpose of epidemological surveys--namely, accounting for the generality of a phenomena. Alternatively, evaluations of a focused treatment may require small homogeneous groups for study. In this situation, perhaps only 10% of the subjects can be classified according to rigorous criteria, but the reliability of assignment to groups is high and the subjects represent relatively "pure types". This low level of coverage has the advantage of increased statistical power for comparisons among groups but carries with it the disadvantage of decreased generalizability of findings. Thus, findings cannot be extrapolated to the population as a whole but only that portion who resemble pure types.

In simple terms, therefore, the issue is whether it is better to "classify some of the people some of the time," or attempt to classify everybody. In order to resolve this issue, it is necessary to determine if the benefits of reduced coverage outweigh the costs. In this "bootstrapping" stage of criminal justice research, however, it may be more valuable to construct taxonomies which classify fewer individuals, if the benefit is increased statistical power and reliability. Unfortunately, conventional taxometric methods available to researchers in criminal justice do not permit the manipulation of the coverage of the resulting taxonomy. Most methods are aimed at simply partitioning (or amalgamating) subjects into a discrete set of groups--despite the fact that the group members differ widely in the degree to which they represent "types". Thus, in order to resolve the issue of coverage, it is necessary to develop, evaluate, and apply new taxometric methods.

In the following section, some innovative methods for manipulating the coverage of empirically derived taxonomies are proposed. These methods represent modifications and extensions of conventional taxometric procedures which have been used by researchers in criminal justice and other areas of the behavioral and social sciences.

III. Methods of Varying Coverage

A. Introduction

Most taxometric methods are not designed to manipulate the coverage of a classification, but rather to identify groupings of individuals based on some statistical criteria. This is not to say that these methods produce taxonomies having 100% coverage. Many empirically derived taxonomies include a subset of subjects who are "unclassified" (i.e., Carlson, 1977; Megargee, 1977). These unclassified groups represent only a tiny proportion of multivariate outliers and do not result from manipulating coverage along a continuum of classifiability. Many taxometric methods, however, are amenable methodological changes that would permit manipulating coverage along this continuum.

In this program of research, innovations are proposed for three methods which have been used successfully to create empirical taxonomies in the social and behavioral sciences. The three methods include (a) a method called centroid analysis (Edelbrock and Achenbach, 1980). (b) Lorr's nonhierarchical clustering technique (Lorr, Bishop, and McNair, 1965; Lorr and Radhakrishnan, 1967), and (c) inverse factor analysis (Monro, 1955; Ryder, 1964: Stephenson, 1936). These methods are not the only candidates for such innovations, but they cover a range of taxometric approaches used in the social and behavioral sciences, have produced useful taxonomies, and have direct applications in criminal justice research. The goal of comparing a range of methods is to determine the degree to which the principle underlying the manipulation of coverage is valid, apart from the idiosyncrasies of one particular method. In the following section, each method is outlined and the innovations proposed to manipulate coverage are described in detail.

B. Centroid Analysis

Centroid analysis is a new taxometric procedure and has several advantages over previous methods, including the abilities to (a) construct hierarchical taxonomies, (b) determine the reliability of profile types, and (c) classify new subjects who were not in the original analysis. Moreover, centroid analysis embodies a mechanism for manipulating the coverage of the classification - although this feature has not yet been fully explored.

Centroid analysis involves three steps:

Step 1 Identification of profile types.

Step 2 Determining the reliability of profile types.

Step 3 Classification of subjects according to the profile types.

1. Identification of profile types. In centroid analysis a conventional hierarchical clustering algorithm is used to identify subgroups of individuals having similar characteristics or patterns of scores. The patterns characterizing such subgroups are termed "profile types". Specifically, the centroid clustering method, also known as the weighted pair group method (Sokal and Michener, 1958), is used. Several issues are involved in the identification of profile types. For one, clustering algorithms will identify homogeneous subgroups of individuals even when applied to random data. That is, to some degree clustering algorithms impose structure on data as well as reveal inherent structure. Thus, some profile types are likely to be methodological artifacts rather than representing reliable profile patterns that characterize subgroups of individuals. One way to deal with this problem is replicate profile types across samples, retaining only those profile types identified in two or more analyses. In almost all applications where this is done, some profile types are identified in one sample that do not replicate in subsequent samples.

2. Measure of similarity. A second issue involves the choice of the measure of similarity among individuals. A variety of similarity measures are available for use in cluster analysis (of Cattell, 1949; Cronbach and Gleser, 1953; Gregson, 1975; Tatsuoka, 1974) and they determine, to a large extent, the nature of the profile types that are identified (i.e., whether the profile

types differ predominantly in elevation, shape, etc.) Some parametric comparisons of clustering methods using different measures of similarity have been performed on computer generated data sets. The results indicate that certain similarity measures (i.e., correlation, intraclass correlations) result in more accurate clustering solutions than other measures (Edelbrock, 1979; Edelbrock and McLaughlin, 1980; Mezzich, 1978). Two measures of similarity will be systematically explored when using centroid analysis, including correlation, and the one-way intraclass correlation (cf. Edelbrock, 1979; Edelbrock and McLaughlin, 1980). These measures cover a broad range of approaches to quantifying profile similarity and are sensitive to various aspects of profile elevation, shape, and scatter.

3. Clustering Algorithm. Assuming that the sample size is large enough to permit replication of profile types and that an appropriate measure of similarity has been chosen, Step 1 of centroid analysis involves separate hierarchical cluster analyses of the data using the centroid method. The centroid algorithm proceeds by first calculating the similarity between each possible pair of profiles in the sample. Next, the two profiles which are most similar to each other are located and combined into a cluster. These two profiles are then replaced by their centroid which is the profile created by averaging the two subject's scores on each scale. On the next step, this centroid is treated just like the profile of a single subject and the similarities between all possible pairs of profiles are recomputed. In each cycle, the two profiles which are most similar to each other are located, combined into a cluster, and replaced by their centroid. Whenever an individual profile or cluster is combined with another cluster, the centroid is computed using a "weighted" procedure. That is, the centroid is obtained by calculating the average

As cycles proceed, larger and larger clusters are formed and combined in a hierarchical manner. The result is a hierarchical clustering of all profiles, in which groups of subjects having similar profile patterns, and the hierarchical relations among these groups, can be identified. At low levels in the hierarchy, profile types are identified which have very specific patterns characterizing small subgroups of subjects. At higher levels, these groups are combined into larger groups of subjects representing more global patterns. Thus, hierarchical taxonomies permit comparisons among groups of various levels of generality. Many small groups having very distinct profile patterns may be compared or a few larger groups representating more global patterns may be analyzed. These multiple levels of analysis are extremely valuable in research. For example, it is difficult, time consuming, and expensive to obtain long-term recidivism data on offenders receiving different treatments. It is possible that significant differences among groups may be detected at one level of the taxonomy but not at other levels. It is a mistake, therefore, to invest research resources in a study wherein the taxonomy permits only one level of analysis.

4. Classification of new subjects. Most taxometric methods classify only those subjects included in the original analysis and do not embody procedures for assigning new cases to groups. This is unfortunate because if empirically derived taxonomies are to have any applications in decision making it will be necessary to classify new subjects. Step 3 of centroid analysis, therefore, involves procedures for classifying new subjects. In order to classify an individual, the similarities between the subject's profile and the reliable centroids identified in the previous two steps are calculated. The subject is then classified according to the profile type with which his/her profile is most similar. Thus, if correlation is the similarity measure, the

subject's profile is correlated with each profile type and is classified according to the type having the highest correlation.

5. Manipulating coverage. The procedure of assigning subjects to groups permits direct manipulation of the coverage of the classification. A minimum similarity required for classification can be specified, such that profiles whose similarities to any of the profile types are less than the minimum cutoff point are not classified. By changing this minimum cutoff point, the coverage of the classification can be varied. That is, the use of the high cutoff point will result in a small proportion of subjects being classified into relatively homogeneous, non-overlapping groups that represent "pure types". Conversely, the use of a low cutoff point results in the classification of a higher proportion of subjects into larger and more heterogeneous groups which have a higher degree of overlap.

C. Lorr's Technique

1. Introduction. Lorr has developed a non-hierarchical taxometric method that has been used to construct a variety of taxonomies (Berzins, Ross, English, and Haley, 1974; Goldstein and Linden, 1969; Lorr, Bishop, and McNair, 1965; Lorr, Pokorny, and Klett, 1973; Lorr and Radhakrishnan, 1967). In most applications of Lorr's technique, Q-correlations were used to measure similarity among profiles, although a variety of similarity metrics could be used. In this research, two measures of similarity (correlation, and the one-way intraclass correlations) will be systematically compared using Lorr's technique.

2. Clustering Algorithm. To illustrate this method, assume correlation is used as the similarity measure. The first step in identifying clusters is to calculate the Q-correlations between all possible pairs of profiles. Considering only those correlations above a certain cutoff point, the profile

having the highest average correlation with other profiles becomes a "pivot" profile. A pivot profile is essentially the "seed" or foundation for building a cluster. To build the first cluster, all profiles with correlations greater than a cutoff point (C_{in}) with the pivot profile are combined into one cluster. Among the remaining profiles, those with average correlations with cluster members that are greater than a lower cutoff point (C_{ex}) are removed from the sample and are not classified. Those with average correlations of $< C_{ex}$ are candidates for other clusters.

In each cycle, all profiles having correlations with the pivot profile C_{in} are combined into a cluster. Profiles with correlations C_{in} but $> C_{ex}$ are removed from the sample and are not classified. This is because even though they may qualify for membership in another cluster, such profiles would still be relatively similar to the first cluster --which would result in overlapping groups. Profiles with correlations $< C_{ex}$ are considered for other clusters. In each cycle, a pivot profile is identified and cluster membership is determined by C_{in}. Those profiles with correlations between C_{in} and C_{ex} are removed and those with correlations $< C_{ex}$ remain for another cycle. Cycles proceed until all profiles are either classified or deemed inappropriate for classification.

3. Manipulating Coverage. Due to the use of the dual cutoff criteria, this method does not result in taxonomies having 100% coverage. Moreover, the cutoff criteria are a convenient mechanism for manipulating the coverage of the classification. By varying C_{in} and C_{ex} the coverage (as well as the homogeneity and degree of overlap) of the groups can be directly manipulated. Specifically, coverage is decreased by setting a high C_{in} value (implying that cluster members are highly similar to the pivot profile) and a low C_{ex} (implying that cluster members are not very similar to other clusters). To

increase coverage, C_{in} is decreased and C_{ex} is increased. Coverage of 100% can be achieved when C_{in} is low and $C_{in} = C_{ex}$. This would, of course, produce relatively heterogeneous, overlapping groups.

Although this mechanism for manipulating coverage is built into Lorr's technique, most researchers employing this method have chosen a single set of C_{in} and C_{ex} values to construct taxonomies and have not systematically explored the effects of varying C_{in} and C_{ex} on the statistical power of their taxonomies. In most applications, C_{in} has been chosen according to some significance criterion (i.e., p .05 at the degrees of freedom determined by the number of variables), and C_{ex} has been chosen to represent a lower significance criterion (i.e., p .10 or .20). This is a rational approach that produces useful taxonomies, but does not take advantage of the built-in mechanisms for manipulating coverage.

D. Inverse Factor Analysis

A. Introduction. Inverse, or Q-type, factor analysis is one of the oldest taxometric methods and has been widely used to construct taxonomies in psychology (Monro, 1955; Overall and Klett, 1972; Stephenson, 1936) and criminal justice (i.e., Butler and Adams, 1966; Collins, Burger, and Taylor, 1976). Although the inverse factor analysis has been criticized as a taxonomic tool (i.e., Baggaley, 1964; Fleiss, Lawlor, Platman, and Fiede, 1971; Fleiss and Zubin, 1969; Jones, 1968; Lorr, 1966), it remains a popular method for taxometric problems and has produced useful taxonomies. Since factor analysis methods have become somewhat standardized and have been discussed in detail elsewhere (Harman, 1976; Fruchter, 1954; Mulaik, 1972) they will not be described in detail here. Instead, the focus of the following section will be on methods for manipulating the coverage of taxonomies constructed using inverse factor analytic methods.

B. Grouping Procedure. Factor analysis is typically used to summarize the matrix of correlations among variables in terms of the limited number of "factors". Each factor is a vector of weights or factor loadings which indicate the degree to which each item is associated with that factor. Since one factor only accounts for a proportion of the variance in a correlation matrix, most applications of factor analysis yield many factors which account for more and more of the remaining variance. Each factor is determined by a group of items which are highly intercorrelated and thus have high loadings on that factor.

As typically applied, factor analysis identifies grouping of items. Simple modifications of the factoring procedure result in the identification of groupings of individuals. In Q-type factor analysis, profile data describing individuals is inverted and intercorrelated producing a correlation matrix representing similarities among individuals, rather than similarities among items. Thus, the factor analysis identifies groups of individuals having similar patterns of scores, rather than identifying groups of intercorrelated items. The factor loadings indicate the degree to which the subjects are similar to the "type" represented by the factor.

C. Manipulating Coverage. In Q-type factor analysis, the factor loadings serve as a way to manipulate the coverage of classification. That is, the loadings represent the continuum of classifiability whereby subjects with high loadings are very similar to the "type" represented by the factor. As loadings decrease, subjects become less similar to the "type". Thus, coverage of the classification can be manipulated by varying the minimum loading required to be classified. A high cutoff point classifies relatively few subjects into homogeneous groups, whereas a low cutoff point classifies more subjects into more heterogeneous groups. Although this is an obvious way to vary coverage, previous researchers employing factor analysis as a taxometric method have simply chosen a single cutoff point (i.e., $\geq .30$) for determining groups.

One problem with using minimum loadings to vary coverage is that subjects may have relatively high loadings on more than one factor. This is because a single factor does not account for all of the variance in a subject's pattern of scores. One solution to this problem is to adopt dual cutoff criteria in a manner similar to Lorr's technique. Specifically, in order to be classified a subject must have a loading greater than a cutoff point on one factor and the loadings on all other factors must be below a second cutoff point. By varying the magnitude and relative difference between the cutoff points, the coverage of the classification can be effectively varied.

IV. Evaluation and Comparison of Methods

A. Goal of this Research. The goal of this research involved the evaluation and comparison of these previously described taxometric methods. Computer generated data sets were used because they have the advantages of having predetermined groups with known correlates. This makes it possible to compare and evaluate the taxometric methods on their ability to recover the groups or "types" built into the data and determine if the resulting taxonomies are predictive of predetermined differences among the groups. One goal of these analyses is to determine, for each method, if systematically reducing coverage improves the statistical power of the taxonomy. That is, we seek to establish the general relations between coverage and statistical power. However, between-method comparisons are also important. Thus, we seek to identify which methods produce the most accurate and predictive taxonomies having the highest coverage.

In the following section the statistical model, data sets, evaluative criteria, and strategy for this phase of the research will be outlined.

B. The mixture model. The mixture model has been proposed as a statistical model for evaluating taxometric methods (Blashfield, 1976; Sclove, 1977; Wolfe, 1970). According to this model, the task of taxometric analysis is to resolve a mixture of populations into its components when the underlying populations and their parameters are unknown. In statistical terms, suppose X is a mixture of k populations, such that

$$X = (x_1, x_2, x_3, \ldots x_K),$$

where x_i denotes an n_i xp matrix based on n_i entities samples from the ith population measured on p variates. Each population has an associated probability distribution, $f(x_i)$, and is defined by parameters u and , where u is a p-length vector of population means, and   is a p x p population covariance matrix. The probability distribution for the mixture X is

$$f(K) = \sum_{i=1}^{K} (N_1/)\ f\ (x_i),$$

$$f(X) = \sum_{i=1}^{K} (n_1/n)\ f\ (x_i),$$

where

$$n = \sum_{i=1}^{K} n_i.$$

The taxonomic problem is to resolve the mixture X into its component populations (x) such that the parameters (n_i, u, and) and members of each population can be specified.

Drawing an example from research, suppose a sample of criminal offenders is described in terms of their scores on a personality inventory (e.g., Megargee, 1977). The total sample (X) is assumed to be composed of several underlying populations or "types" of individuals (x_1, x_2, $x_3 \ldots x_k$). Each type is defined by a particular personality pattern , which can be

described in terms of an average pattern of scores (u) and a covariance structure ( ). Unfortunately, the underlying populations and their parameters are unknown. Thus, a taxometric method, such as hierarchical cluster analysis, would be used to identify the underlying types in the sample.

The mixture model provides a basis for evaluating and comparing taxometric methods. Using "Monte Carlo" procedures, computer generated data sets can be constructed which simulate a mixture of populations. Unlike real data, the parameters of the underlying populations are predetermined and known. Thus, taxometric methods can be evaluated and compared on their ability to resolve mixtures into their component populations (Blashfield, 1976; Edelbrock, 1979; Edelbrock and McLaughlin, 1980; Gross, 1972; Kuiper and Fisher, 1975; Mezzich, 1978; Mojena, 1977; Rand, 1971). Mixture model comparisons are extremely valuable because a wide variety of taxometric methods are available for use and different methods are likely to produce different results when applied to the sme data. Such comparisons help identify those taxometric methods which are most likely to produce fruitful research results when applied to real data.

C. "Benchmark" data sets.

Comparisons are planned for computer generated data, including 20 multivariate normal mixtures generated by Blashfield (197). Each mixture consists of two or more multivariate populations representing underlying groups or "types" which differ in their profile characteristics and external correlates. These mixtures simulate the type of taxonomic problem encountered in the behavioral and social sciences. The profile data consists of scores on continuous, quasi-normal distributions which correspond to the type of data provided by many behavioral and personality measures. The scores on each profile dimension also embody a certain degree of error, which simulates the less than perfect reliability of real data. This error variance implies that the underlying populations overlap to some extent and that population members represent a range in similarity to the population "type". This corresponds to the continuum of classifiability encountered in research applications.

Finally, the populations are constructed to differ in external criteria and these differences represent small to medium effects. This is analogous to a researcher validating a taxonomy of criminal offenders against external criteria such as background variables, recidivism, etc.

These data sets represent a range of parameters describing mixtures and their underlying populations, as well as a range in difficulty of solution. They were selected in preference to generating new mixtures because they have been well characterized, extensively studied, and are available to other researchers (cf. Blashfield, 1976; Edelbrock, 1979; Edelbrock and McLaughlin, 1980; Mojena, 1977). The advantage of using these previously analyzed "benchmark" data sets is that the results of this study can be directly compared to previous studies of different taxometric methods. The generation of new mixtures would preclude direct comparisons to previous work. That is, differences due to methodological innovations would be confounded with differences among the data sets analyzed.

D. Evaluative criteria. The taxometric methods will be evaluated and compared in terms of their accuracy and statistical power.

Accuracy refers to the degree to which a taxonomy recovers the underlying groups in the mixtures. Thus, an accurate taxometric solution is one that groups together members of the same underlying populations fn the mixture, whereas an inaccurate solution would group together members of different populations. A variety of measures of accuracy have been used, most of which are based on quantifying the degree of agreement between the empirically

derived groups and the underlying populations. In this research, the statistic lambda was used to calculate accuracy.

The statistical power of taxonomies refers to the degree to which differences among taxonomic groups can be detected. This involves detecting differences in external criteria, that is criteria not involved in the construction of the taxonomy. This would correspond to a researcher validating a taxonomy based on behavioral or personality measures against external correlates such as background variables, recidivism, etc. In this research, statistical power is estimated by testing differences among the empirically derived groups on artifically generated data. These differences will be tested using one-way analyses of variance (ANOVAs) with group membership serving as the classification variable. These ANOVAs will reveal the degree to which the groups differ on the external criteria. These differences will be quantified in terms of "effect size". Based on the effect size and sample size, statistical power can be derived using tables provided by Cohen (1975). Statistical power is measured on a scale from zero to 1.00, which represents the probability that the null hypothesis will be rejected if it is in fact false. In other words, statistical power is the ability to detect a significant difference if the difference is real.

E. Evaluation strategy. The strategy of this research was to determine the effect of manipulating coverage on the accuracy and statistical power of the taxonomies produced by each method. For each method, mean accuracy and statistical power values calculated separately for the multivariate normal mixtures at various levels of coverage. Since coverage can be directly manipulated, the levels of coverage to be analyzed can be determined in advance. Analyses were made at levels between 100 and 0% coverage at various intervals. These levels thus cover the broadest possible range of cover-

-age. For the multivariate data sets, mean accuracy and statistical power values were based on 20 mixtures.

V. Application of Methods

The mixture model comparisons described above will contribute to a more comprehensive theory and statistical model of taxometric analysis. Moreover, they will identify those methods that are most likely to produce fruitful research results when applied to real data. However, even if the advantages of these innovative methods can be demonstrated on computer generated data sets, these methods will not automatically be used by researchers in criminal justice. In order to be used, the advantages of these methods must be demonstrated on real data relevant to criminal justice. Thus, the third phase of this research involves the application of these new methods to criminal justice data.

Delinquency data. Data collected in a longitudinal study of delinquency and dropping out of school will be analyzed. Data on 2,617 junior and senior high school students were collected using teacher reports, parent interviews, student questionnaires, and school records. The dependent variables were dropping out, self-reported delinquency, juvenile offense record, and adjudication as delinquent. These data and the design of study are described in detail in Elliot and Voss (1974). The taxonomic problem is to determine if subgroups of youth, differing in patterns of scores on such variables as success at home, success at school, normlessness, punitiveness, commitment to peers, and commitment to parents, differ in subsequent school failure, delinquent behavior, criminal offenses, and adjudication. The strategy will be to systematically evaluate the effect of manipulating the coverage of such taxonomies on the ability to detect significant differences in self-reported delinquency.

RESULTS

VI. Within Method Comparisons

A. Inverted Factor Analysis

1. Procedure. Data were double-centered according to the rationale and procedure given by Overall and Klett (1972; pp. 203-204). Variables were standardized (mean = 0, sd = 1) and scores were then standarized equivalently across objects. Each of the 20 (object X variable data sets was then inverted (i.e., to represent a variable X object matrix) and subjected to principal-components factor analysis using the BMDP4M - program. It is important to note that double-centering the data results in bipolarity of the unrotated factors. However, it does not necessarily result in bipolarity in the rotated factors, which were used here.

Two procedures were used to determine the number of factors. First, for each mixture, the number of factors was set to equal the number of underlying populations. Since the rotated factors were not bipolar, each factor comprised only one group of objects having high loadings in the same direction. Thus, determining the number of factors in this way is tantamount to setting the number of groups (j) equal to the number of underlying populations (k). These 20 analyses are subsequently designated by the notation j = k.

Second, the number of factors was determined by examining eigen values. For these data sets, the commonly used eigen value greater than 1" rule resulted in considerable over-factoring. A few factors having large eigen values were obtained followed by several having eigen values slightly greater than 1.00. This problem was also encountered by Blashfield and Morey (1980). Following their procedure, Cattell's (1966) scree test was used to determine number of factors. In this study, two investigators examined the eigen value plot for each mixture and independently selected the number of factors.

Although we agreed for all 20 mixtures, the number of factors indicated by the scree test did not always equal the number of underlying populations. For eight mixtures, the number of factors equalled one more than the number of underlying populations (i.e., k + 1). These 20 analyses are subsequently designated by the notation j ≠ k (i.e., the number of groups did not necessarily equal the number of populations).

One issue in factor analysis is whether to construct orthogonal (uncorrelated) or oblique (correlated) factors. This is an important consideration when deriving typologies because rotational procedures substantially affect final factor loadings, which are the basis for constructing groups. Most previous applications of inverted factor analysis (e.g., Blashfield & Morey, 1980; Collins et al, 1976; Fleiss et al, 1971; Katz & Cole, 1965) involved the varimax rotation--an orthogonal procedure. In this study, both varimax (orthogonal) and direct quartimin (oblique) rotations were compared. This yields four analyses of 20 mixtures each: j = k and j ≠ k with either varimax or direct quartimin rotation.

A crucial issue that arises in inverted factor analysis involves translating factor loadings into discrete groups of objects or individuals. A common procedure has been to assign individuals to groups on the basis of highest factor loadings (in terms of absolute value). Some investigators have specified a minimum loading required for classification. Fleiss et al (1971), for example, selected a minimum loading of .40. Individuals whose highest loadings were less than .40 were left unclassified. In their Monte Carlo study, Blashfield and Morey (1980) selected a minimum loading of .60, with the additional criterion that an object could not have a loading of .60 or higher on any other factor. These rather stringent criteria reduce coverage substantially, but result in more distinct and homogeneous groups.

In this study, objects were assigned to groups on the basis of their highest loadings. This is a simple procedure for constructing groups, but the coverage of the resulting classification can be manipulated by simply changing the minimum loading required for assignment. A low cutoff point results in the classification of a high proportion of objects into relatively heterogenous groups, whereas a high cutoff point results in the classification of a low proportion of objects into more distinct, non-overlapping groups. This assignment procedure therefore makes it possible to evaluate classifications at several levels of coverage.

2. Calculating Accuracy. The accuracy of the inverted factor solutions was defined as the agreement between the obtained groups and the underlying populations in the mixtures. A wide variety of statistics have been used to measure accuracy in mixture model studies, and there is little regarding the "best" accuracy meaeure. Kappa (Cohen, 1960) and Rand's statistic (Rand, 1971) have been used in many studies (e.g., Blashfield, 1976; Edelbrock, 1979; Edelbrock & McLaughlin, 1980; Kuiper & Fisher, 1975; Milligan & Isaac, 1980; Mojena, 1977; Rand, 1971). Both of these measures have drawbacks. Kappa has the advantage of correcting for chance level of agreement in a cross-classification, but it is appropriate only for square matrices (i.e., j = k). Rand's statistic does not require that j = k, but the scale is not uniform from matrix to matrix. That is, the lower bound of Rand's statistic is not zero but is determined by the marginal distributions of the cross-classification.

One way to overcome the idiosyncracies inherent in individual measures is to use multiple criteria for evaluating accuracy. Six measures, including kappa, Rand's statistic, asymmetric lambda, tau, Kramer's v, and the contingency coefficient were used in this study. We chose to report our main findings in terms of asymmetric lambda for several reasons. This statistic is appropriate for nominal level cross-classifications, has a range of zero to 1.00, and can be used with either square (j = k) or rectangular (j ≠ k) matrices. The "asymmetrical" aspect of this statistic also seems well suited to the task of measuring accuracy. The term "asymmetrical" refers to the fact that lambda indexes the degree to which one classification predicts another, and not vice versa. In mixture model studies, the underlying populations comprise a fixed or dependent classification, predicted by empirically derived groups that are free to vary.

Finally, it is worth noting that our conclusions regarding the relative accuracy of various methods were identical for all six measures we explored. This is not surprising, since such measures are all founded on the same information extracted from the cross-classification matrix (cf. Hubert & Levin, 1976). Furthermore, in these analysis, the six measures of accuracy correlated > .95 with one another.

3. Statistical Analyses - Accuracy. For each of the 80 inverted factor solutions, objects were classified according to levels of coverage dictated by the following minimum loadings: .0, .4, .5, .6, .7, .8, and .9. These minimum loadings between were selected because: (a) all objects had highest loadings (b) very few objects had highest loadings between .0, and .4 so accuracy and coverage varied little in this interval, and (c) there were too few loadings above .9 to calculate accuracy.

Accuracy and coverage values were analyzed in separate 2 x 2 x 7 analyses of variance representing: number of factors (j = k vs. j ≠ k) rotational methods (varimax vs. direct quartimin), and minimum loading (.0 to .9), respectively.

Main results are portrayed graphically in Figures 1 and 2. These figures show the relations between the minimum loading reverage (right axis). Figure 1 depicts accuracy and coverage results for the j ≠ k solutions. Overall, accuracy and coverage were significantly related to the minimum loading (p. < 001), but in opposite ways. Raising the minimum loading uniformly increased accuracy, but decreased coverage to a greater and greater extent. No significant differences (F < 1.00) were detected between varimax and direct quartimin rotations for either j = k or j ≠ k solutions. Varimax solutions resulted in consistently higher accuracy and coverage, however.

Paradoxically, j ≠ k solutions resulted in significantly higher accuracy and coverage than j = k solutions (p. < 01). This was the case for both rotational methods. Figure 3 portrays accuracy differences between j = k and j ≠ k solutions in a manner that equates them for verage. Accuracy is shown as a function of coverage, rather than as a function of the minimum loadings as in Figures 1 and 2. At all levels of coverage, j ≠ k solutions resulted in significantly higher accuracy and j = k solutions. Examination of the eight mixtures where j ≠ k confirmed that constraining the number of factors to equal the number of underlying groups substantially reduced accuracy. For these mixtures, higher accuracy was achieved when the number of groups was determined empirically by Cattell's scree test.

4. Effect Size and Statistical Power. To evaluate statistical power of the inverted factor analyses, we analyzed one of the computer generated data sets (Blashfield's Data set #17, See Appendix A) in depth. This data set includes seven external criteria. That is, seven variables which were not included in the cluster analyses, but are statistically related to the clustering variables. These external variables thus serve as "validity" criteria against which to evaluate the empirically derived classifications. This would be analagous to

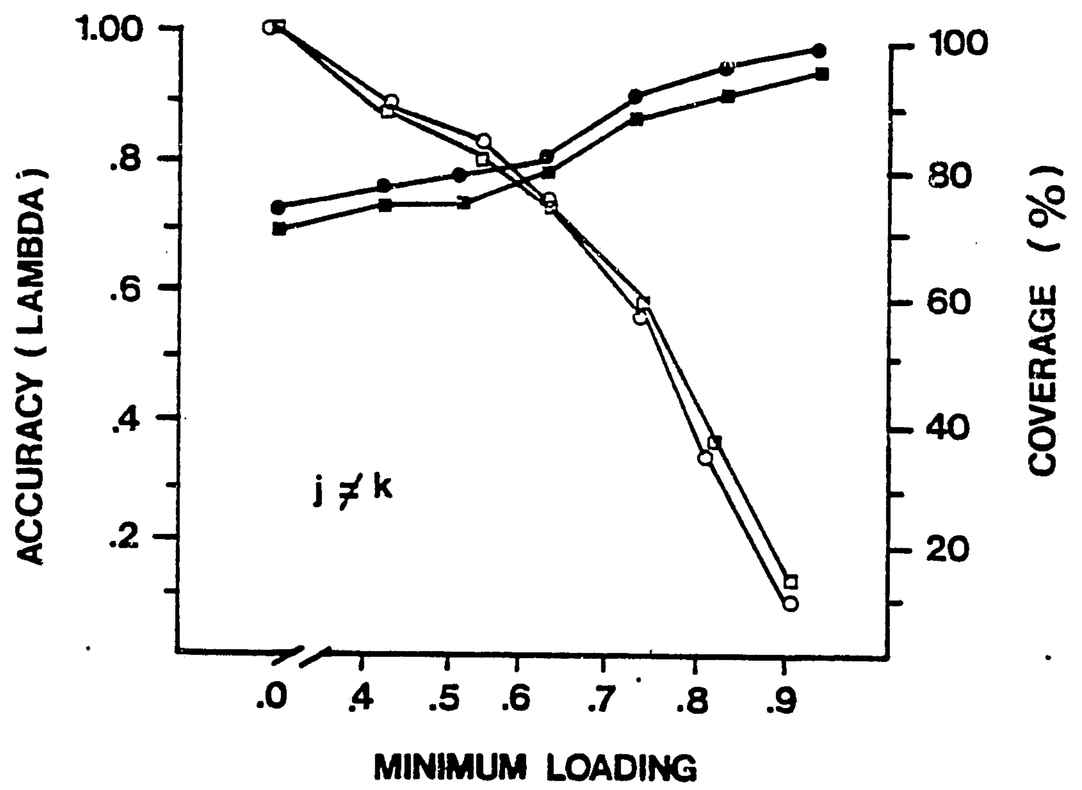

Figure 1. Relations between the minimum loading, accuracy (left axis), and coverage (right axis) for the inverse factoring procedures.
(● Accuracy-Varimax ■ Accuracy-Direct Quartimin
○ Coverage-Varimax □ Coverage-Direct Quartimin)

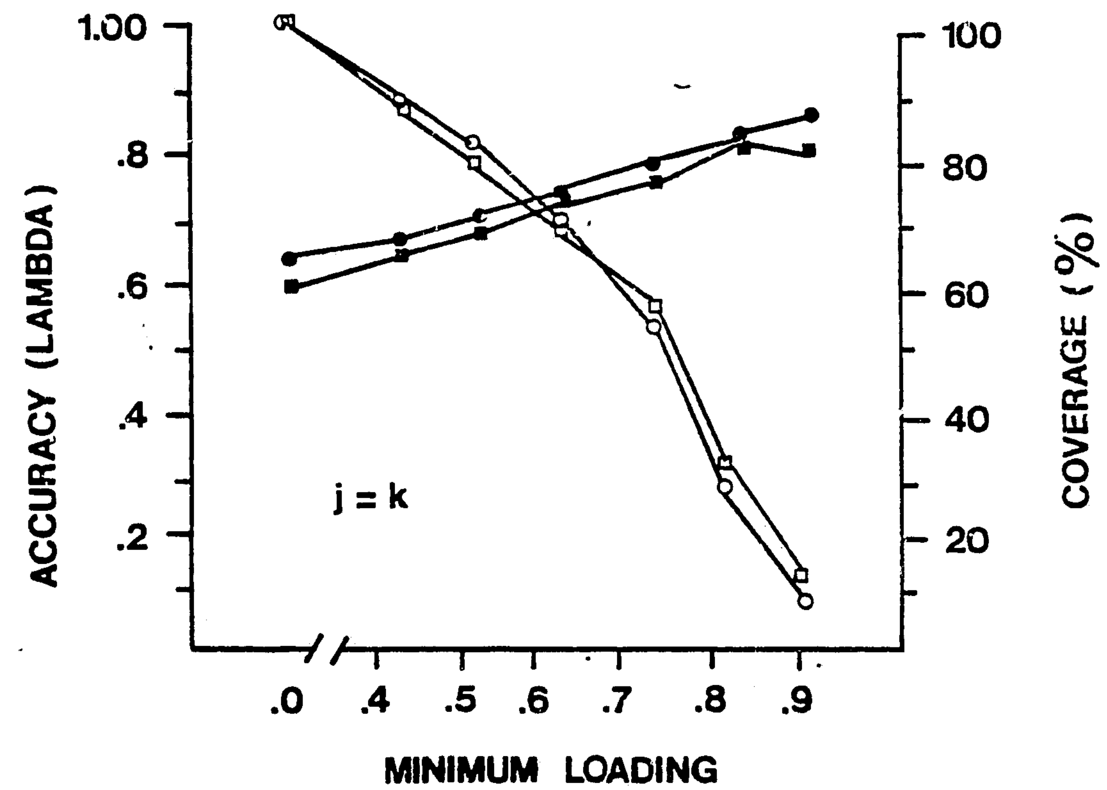

Figure 2. Relations between the minimum loading, accuracy (left axis), and coverage (right axis) for the inverse factoring procedures.
(● Accuracy-Varimax ■ Accuracy-Direct Quartimin
○ Coverage-Varimax □ Coverage-Direct Quartimin)

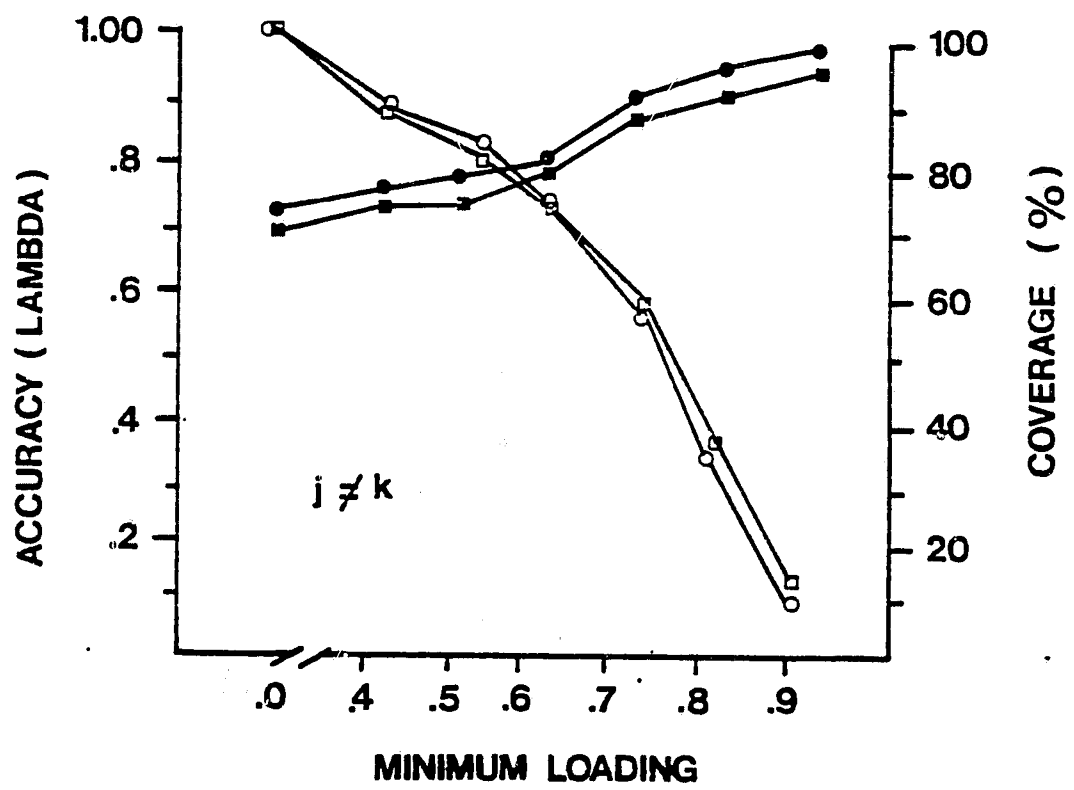

Figure 1. Relations between the minimum loading, accuracy (left axis), and coverage (right axis) for the inverse factoring procedures.
(● Accuracy-Varimax ■ Accuracy-Direct Quartimin
○ Coverage-Varimax □ Coverage-Direct Quartimin)

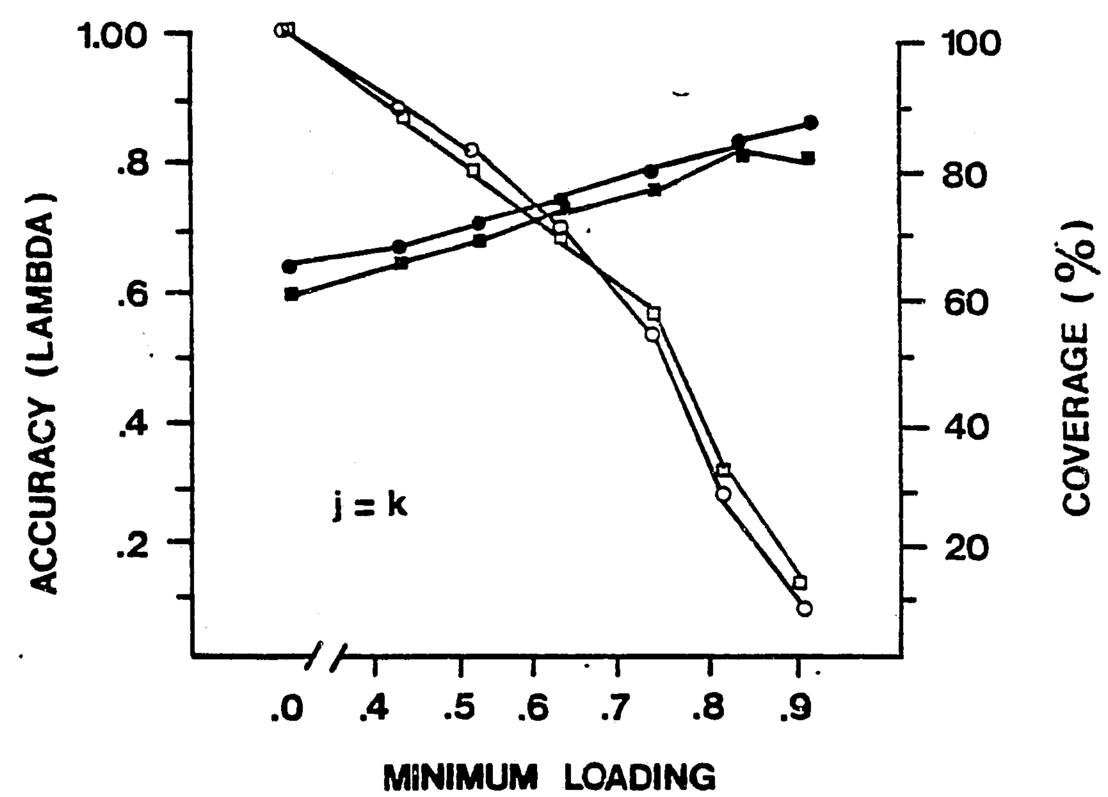

Figure 2. Relations between the minimum loading, accuracy (left axis), and coverage (right axis) for the inverse factoring procedures.
(● Accuracy-Varimax ■ Accuracy-Direct Quartimin
○ Coverage-Varimax □ Coverage-Direct Quartimin)

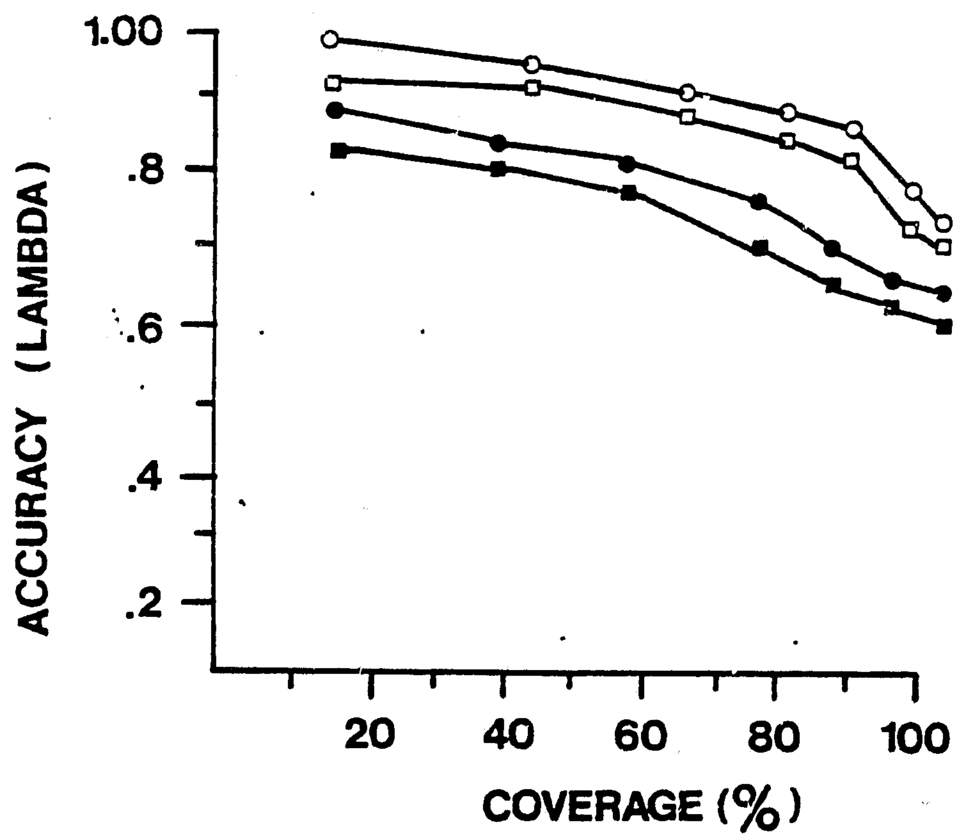

Figure 3. Relations between accuracy and coverage for inverse factoring procedures. (○ Varimax-j≠k □ Varimax-j=k ● Direct Quartimin-j≠k ■ Direct Quartimin-j=k)

Table 1 - Mean Accuracy and Coverage Values for Inverse Factoring Procedures

	Cutoff point (loading) .0	.4	.5	.6	.7	.8	.9
j=k (#clusters=#populations)							
Varimax							
Accuracy	.66	.68	.71	.74	.80	.83	.88
Coverage (%)	100	90	83	71	52	30	9
Direct Quartimin							
Accuracy	.61	.66	.69	.73	.78	.82	.83
Coverage (%)	100	88	81	70	54	33	13
j≠k (#clusters determined via Cattell's scree test)							
Varimax							
Accuracy	.73	.76	.78	.81	.85	.90	.95
Coverage (%)	100	93	87	76	58	34	11
Direct Quartimin							
Accuracy	.71	.74	.77	.80	.84	.88	.90
Coverage (%)	100	92	86	75	59	37	15

to a researcher evaluating an empirically derived taxonomy based on MMPI data against external criterias such as recidivism, subsequent violent behavior, etc. In other words, statistical power is the degree to which the empirically derived clusters differentiate among subjects in terms of external criteria. Statistical power is thus an index of the "predictive power" or "predictive validity" of a classification.

One-way analysis of variance procedures were used to evaluate effect size and statistical power for the clustering solutions obtained via the inverted factor procedure. Since results were almost identical for the varimax and direct quartimin procedures, findings based on only the varimax procedure are reported here. This seemed warranted because the varimax procedure produced slightly higher accuracy values than the direct quartimin procedure in the previous comparisons. Thus, we sought to obtain an estimate of the optimal statistical power of the inverted factoring procedure.

In each of the seven one-way analyses of variance, the empirically derived clusters served as the independent variable and the seven external criteria served as dependent variables. One question we sought to answer was the degree of separation between clusters on the external criteria at various levels of coverage. The statistic $\underline{f}$ (Cohen, 1977) was used as an index of effect size. The statistic $\underline{f}$ serves as an index of the degree to which the clusters explained or accounted for variance in the external criteria. As a rule of thumb, $\underline{f}=.10$ is considered a small effect size; $\underline{f}=.25$ is considered a medium effect size; and $\underline{f}=.40$ is considered a large effect size. Thus, effect size is greater than .40 reflect large differences between clusters on the dependent variables.

Table 2 reports effect sizes for each of the seven external criteria for the varimax grouping procedure. Effect sizes are given at each cutoff point ranging from 0 to .9. This encompasses a range of coverage from 100 to 22%. As shown in Table 2, most effect sizes were in the moderate range. Moreover, there were

Table 2

Effect Sizes for Inverse Factor Analysis (Varimax rotation)

	Cutoff point (loading)						
	.0	.4	.5	.6	.7	.8	.9
External criteria							
1	.12	.09	.07	.08	.14	.35	.70
2	.22	.29	.42	.46	.46	.60	.68
3	.32	.30	.31	.27	.46	.48	.35
4	.41	.47	.48	.52	.54	.64	.92
5	.34	.35	.38	.40	.48	.53	.57
6	.44	.50	.57	.61	.71	.76	.75
7	.19	.27	.28	.29	.44	.48	.35
Mean ES	.29	.33	.36	.38	.46	.55	.62
Coverage (%)	100	92	87	83	70	50	22

differences between the dependent variables. External criterion #1, for example, demonstrated low effect size at all levels of coverage and particularly for high levels of coverage. The most important finding, however, is the relationship between effect size and coverage. Averaging across all seven dependent variables, there is a linear and inverse relationship between coverage and effect size. As coverage decreased, effect sizes increased. In other words, as one classifies fewer and fewer subjects into groups, differences in the external criteria became more pronounced. This can be interpreted in terms of the "continuum of classifiability." At 100% coverage many subjects were grouped into clusters even though they bear little resemblance to other cluster members and do not differ in the same manner on the external criteria. Thus, effect sizes are diluted. At lower levels of coverage, cluster members resemble each other to a greater extent and the clusters themselves are more representative of "pure types." Thus, differences in external criteria become more pronounced.

Considering the linear relationship between coverage and statistical power, one may conclude that the lowest levels of coverage offer the best chance of detecting significant differences in external criteria. This is clearly faulty reasoning. Statistical power, which is the ability to detect significant differences among groups, is a function not only of effect size, but also sample size. As sample size decreases below a certain threshold statistical power also decreases. Given a cutoff point of .9, for example, one would obtain a large effect size (.62) but only 22% of the subjects would be classified. With such a small sample size the probability of detecting significant differences in the external criteria is low, despite the fact that the differences among clusters are relatively "large." Thus, it is important to distinguish between the <u>size</u> of differences between clusters and the <u>significance</u> of these differences.

To evaluate the effects of coverage on statistical power, the probability of

detecting significant differences among groups was calculated according to the procedures provided by Cohen (1977). The $p > .01$ level of significance was used in these statistical power calculations. Results for the inverse factor analysis procedure (varimax rotation) are reported in Table 3. As shown in this table, there was a curvilinear relationship between statistical power and coverage. For these analyses, 100% coverage yielded a .52 probability of detecting a significant difference in external criteria. In other words, one had about a 50/50 chance of finding a significant difference among clusters. However, statistical power increased substantially as coverage decreased. Leaving only 8% of the subject unclassified, for example, boosted statistical power to .60. Leaving 30% of the subjects unclassified boosted statistical power to .78. At lower levels of coverage (e.g., 22%), statistical power again declined to only .51.

These results suggest that classifying all subjects will not result in optimal statistical power. They further suggest that the optimal level of statistical power may be obtained in the range from 70-83% coverage. Reducing coverage from 70 to 50% resulted in negligible gains in statistical power. Moreover, classifying too few subjects had a deleterious effect on statistical power, even though this results in larger effect sizes (See Table 2).

B. <u>Lorr's Technique</u>

1. <u>Procedure.</u> Lorr's non-hierarchical clustering procedure, called "Build-Up," was also evaluated on Blashfield's 20 multivariate normal mixtures. Special computer software was written to perform these cluster analyses and permit us to manipulate the inclusion and exclusion cutoff points as described previously. Lorr's clustering technique was evaluated using 11 combinations of inclusion and exclusion cutoff points, ranging in significance levels from .10 to .001. We expected the lowest levels of coverage to arise from high inclusion cutoff points (.001) and low exclusion cutoff points (.10), as explained previously. The

Table 3

Statistical Power for inverse factor analysis (VARIMAX)

	Cutoff point (loading)						
	.0	.4	.5	.6	.7	.8	.9
External criteria							
1	.05	.03	.03	.03	.05	.36	.65
2	.21	.54	.90	.94	.84	.95	.60
3	.65	.54	.52	.40	.85	.73	.10
4	.91	.95	.94	.99	.97	.97	.99
5	.70	.74	.80	.84	.90	.82	.42
6	.95	.99	.99	.99	.99	.99	.74
7	.19	.40	.42	.47	.84	.73	.10
Mean SP	.52	.60	.66	.67	.78	.79	.51
Coverage (%)	100	92	87	83	70	50	22

highest levels of coverage would arise from inclusions/exclusion cutoff points that were low and close together (.10/.10).

2. Calculating Accuracy. The accuracy of Lorr's clustering solutions were evaluated using asymmetric lambda as described previously for inverse factor analysis. Additional accuracy measures were also employed, but they yielded the same conclusions. Accuracy results were reported in terms of asymmetric lambda because this statistic has conceptual advantages and this would permit direct comparisons with the results obtained for the inverse factoring solutions.

3. Accuracy Results. Mean accuracy results for the 20 multivariate normal mixtures are reported in Table 4. This table reports mean levels of coverage and accuracy for each of the 11 combinations of inclusion/exclusion cutoff points, for both the product-mement correlation and the intraclass correlation. Unexpectedly, there was a curvilinear relationship between accuracy and coverage. In general, accuracy was relatively low at both high and low levels of coverage. For analyses employing the product-moment correlation, accuracy was maximized in the range from 49-72% coverage. For analyses employing the intraclass correlation, the highest accuracies were obtained in the range from 48-68% coverage. The two measures of profile similarity (product-moment correlation and intraclass correlation) did not differ significantly in either accuracy or coverage. These results are similar to those previously reported for inverse factor analysis in that high levels of coverage produced low accuracy.

These accuracy results differ from thos obtained for inverse factor analysis in that low levels of coverage did not produce increasingly higher accuracy values. This can be attributed to the idiosyncratic way in which Lorr's clustering procedure builds clusters. In order to achieve low levels of coverage the inclusion/exclusion cutoff points must be extended beyond reasonable and recommended range.

Table 4

Mean Accuracy and Coverage Values for Lorr's Non-hierarchical Clustering Technique.

C_{in}	.001				.01			.025		.05
C_{ex}	.01	.025	.05	.10	.025	.05	.10	.05	.10	.10
Intraclass correlation										
Accuracy	.74	.74	.74	.71	.96	.95	.96	.94	.93	.88
Coverage (%)	31	29	27	25	59	53	48	68	61	75
Product-moment										
Accuracy	.73	.73	.73	.69	.95	.95	.94	.92	.91	.87
Coverage (%)	34	31	[illegible]	26	62	55	49	72	62	77

Note: Table entries are mean values based on the analysis of 20 mixtures.

That is, the inclusion value must be set exceedingly high while the exclusion value is set exceedingly low. This results in the construction of very peculiar clusters which are not representative of the "pure types" that would be expected at low levels of coverage.

Another unexpected finding was that not combination of inclusion/exclusion criteria resulted in 100% coverage. Even the lowest combination of inclusion/exclusion criteria (.10/.10) failed to classify all subjects. It is possible that 100% coverage could be achieved with even lower inclusion/exclusion criteria, but such cutoff points are not recommended, given the rationale of the grouping procedure.

4. <u>Effect Size and Statistical Power</u>. Effect size and statistical power for Lorr's technique were evaluated on the same multivariate data set used with the inverse factor analysis described previously. Again, clusters defined on the basis of Lorr's grouping procedure served as the independent variable and each of the seven external variables served as the external criteria or dependent variables. One-way analyses of variance were used to evaluate effect size and statistical power according to procedures described previously. Since results for the product-moment correlation and one-way intraclass correlation were almost identical, only results for the product-moment correlation (the measure originally recommended by Lorr) are reported here.

Effect size, as measured by Cohen's $\underline{f}$ statistic, are reported in Table 5. The effect size, coverage, and number of clusters obtained for each of the 11 combinations of inclusion/exclusion criteria are given. Overall, the effect sizes were high, mostly in the medium to large range and showed a linear/inverse relationship to coverage. That is, high levels of coverage resulted in lower effect sizes. As coverage decreased there was a general increase in effect size. These results are similar to those previously reported for inverse factor analysis.

Table 5

Effect Size for Lorr's Non-hierarchical Procedure (correlation)

Cin	.001				.01			.025		.05	.10
Cex	.01	.025	.05	.10	.025	.05	.10	.05	.10	.10	.10
External criteria											
1	.74	.71	.74	.72	.40	.39	.40	.34	.38	.35	.30
2	.88	.88	.89	.90	.64	.62	.63	.56	.59	.60	.40
3	.56	.58	.51	.53	.57	.58	.50	.51	.50	.54	.39
4	.86	.85	.86	.88	.66	.67	.63	.61	.62	.54	.39
5	.57	.49	.32	.22	.54	.51	.29	.43	.39	.41	.39
6	.58	.67	.69	.44	.65	.65	.66	.51	.54	.50	.41
7	.45	.40	[illegible]	.38	.44	.42	.36	.42	.41	.46	.30
Mean ES	.66	.65	.62	.58	.56	.55	.49	.48	.49	.49	.37
Coverage (%)	32	30	26	22	64	56	52	78	71	84	95
#Clusters	6	6	5	4	8	7	6	8	7	7	7

Table 5 also shows an unexpected relationship between the inclusion/exclusion criteria and the number of clusters obtained. With Lorr's technique, the number of clusters is not determined a priori, but is free to vary and is determined by the inclusion/exclusion criteria. Overall, the number of clusters ranged from 4 to 8 and was clearly dependent upon the relative values for inclusion and exclusion. Specifically, combinations of inclusion and exclusion criteria which were close together resulted in more clusters than combinations that were far apart. This can be explained easily, in that the closer the cutoff points, the fewer the subjects excluded from the clustering in each clustering cycle. This leaves more subjects available in subsequent cycles for constructing more clusters. In addition, the number of clusters appear to be related to the absolute value of the inclusion criterion. In general, the lower the level of the inclusion cutoff point the greater the number of clusters.

Statistical power for each of the seven external criteria were also calculated for the 11 runs of Lorr's technique. Statistical power results are reported in Table 6. Overall, statistical power values were high, reflecting a high probability of detecting a significant difference among clusters. Statistical power was also related to coverage. Statistical power was relatively low at extremely high (< 85%) and low (> 55%) coverage. For these data, statistical power was maximized in the range of 64-84% coverage. There were clear disadvantages to reducing coverage below 50%. Classifying only 22% of the sample, for example, resulted in only a .49 probability of detecting a significant difference in external criteria. This can be compared to the .82 probability of detecting a significant difference at 84% coverage.

C. Centroid Clustering.

1. Procedure. The 20 benchmark mixtures were also analyzed using the centroid clustering/nearest centroid assignment procedure previously described by Edelbrock

Table 6

Statistical Power for Lorr's Non-hierarchical Procedure (correlation)

Cin	.001				.01			.025		.05	.10
Cex	.01	.025	.05	.10	.025	.05	.10	.05	.10	.10	.10
External Criteria											
1	.90	.73	.86	.74	.49	.39	.46	.44	.55	.50	.41
2	.99	.95	.95	.95	.95	.90	.94	.93	.95	.99	.76
3	.52	.48	.38	.40	.90	.86	.74	.90	.85	.95	.74
4	.99	.93	.94	.93	.97	.94	.94	.98	.98	.95	.74
5	.60	.31	.10	103	.86	.70	.20	.69	.57	.68	.74
6	.61	.62	.76	.25	.97	.92	.95	.90	.90	.92	.77
7	.32	.16	.12	.14	.60	.48	.32	.69	.62	.79	.41
Mean SP	.70	.60	.59	.49	.82	.74	.65	.79	.77	.82	.65
Coverage (%)	32	30	26	22	64	55	52	78	71	84	95
#Clusters	6	6	5	4	8	7	6	8	7	7	7

and Achenbach (1980). The strategy by which this method constructs clusters was described previously. One practical problem with this method is that it does not yield a discrete number of clusters, but rather a hierarchical arrangement of objects and groups. Determining the appropriate number of clusters in a hierarchical solution is a difficult and thorny problem (cf. Mojena, 1977). In this research, the number of clusters was set to equal the number of underlying populations in the mixture. Outliers were excluded from determining the number of clusters to avoid the problem of a single outlier being considered a "cluster" (See Edelbrock, 1979 for a discussion of this methodological problem).

Two measures of profile similarity were tested using the centroid clustering procedure: the product-moment correlation and the one-way intraclass correlation. Previous research on the centroid clustering procedure (Edelbrock and McLaughlin, 1980) had suggested that the intraclass correlation would result in higher accuracy, but it was necessary to explore this question further here. Running analyses on both of these measures would also maximize comparability with Lorr's technique as reported in previous section.

Each of the 20 multivariate normal mixtures was analyzed separately using the centroid clustering technique employing either the product-moment correlation or the one-way intraclass correlation as the measure of profile similarity. For each clustering solution, the appropriate number of clusters was determined (excluding outliers) and the cluster centroids were calculated. Each object was then classified according to the cluster centroids, using the nearest centroid assignment procedure described previously. The minimum value required for classification could be varied to manipulate coverage. Six minimum cutoff points were used, ranging from .0 to .9.

2. Calculating accuracy. After objects were assigned to clusters using the nearest centroid procedure, asymmetric lambda was calculated as an index of the accuracy of the clustering solution. Again, cluster membership served as the "independent variable" in the cross-classifications, while true population membership served as the "dependent variable." The accuracy values thus reflect the degree to which the true population membership could be predicted by the clustering solution.

3. Accuracy Results. Mean accuracy results for the 20 multivariate normal mixtures are reported in Table 7. This table shows mean accuracy (lambda) for the centroid method using either the product-moment correlation or the intraclass correlation. Mean accuracy and coverage values are given for cutoff points of .0, .1, .3, .5, .7, and .9--corresponding to coverage levels ranging from 100 down to 9.4% As shown in Table 7, accuracy values were moderately high and were related to coverage in the expected way. That is, accuracy increased relative to declines in coverage in a quasi-linear way. Thus, accuracy was lowest at 100% coverage and increased monotonically as coverage decreased. Unlike Lorr's technique, accuracy did not deteriorate at the lowest levels of coverage. For the centroid clustering procedure, accuracy continued to increase even below 15% coverage.

Table 7 also indicates that the intraclass correlation yielded higher accuracy values than the product-moment correlation. For example, at 100% coverage, the intraclass correlation resulted in an accuracy value of .73, whereas the product-moment correlation resulted in an accuracy value of .68. This difference is significant ($p > .05$) by the paired t-test. Although the cutoff points did not yield perfectly comparable levels of coverage for the product-moment correlation and the intraclass correlation, the results in Table 7 indicate that the intraclass correlation resulted in consistently higher accuracy values at lower levels of coverage. This is consistent with previously reported comparisons between these two similarity measures (Edelbrock and McLaughlin, 1980).

Table 7

Mean Accuracy and Coverage Values for the Centroid Clustering Procedure.

	Cutoff Point					
	.0	.1	.3	.5	.7	.9
Intraclass correlation						
Accuracy	.725	.733	.762	.807	.905	1.00
Coverage (%)	100	96.1	84.9	56.3	26.6	1.3
Product moment correlation						
Accuracy	.688	.690	.715	.778	.853	.968
Coverage (%)	100	97.9	92.7	73.0	38.8	14.9

NOTE: Table entries are mean values based on 20 mixtures.

4. Effect Size and Statistical Power. The effect size and statistical power of the centroid clustering procedure was evaluated on Blashfield's data set #17, in a manner described previously for inverse factor analysis and Lorr's non-hierarchical clustering technique. The results for the product-moment correlation and the intraclass correlation were similar. We chose to report the results for the intraclass correlation since this measure produced significantly higher levels of accuracy and appeared to be the similarity measure of choice for this clustering method.

Mean effect sizes for the clustering procedure are reported in Table 8. This table summarizes effect size as measured by Cohen's *f* statistic, for each of the seven external criteria at six levels of coverage dictated by the cutoff points. As shown in Table 8, effect sizes were predominantly in the high range. As expected, effect size had a curvilinear relationship to coverage: effect size was lowest at 100% coverage and increased as coverage declined to about 30%. Below 30% coverage, effect size dropped slightly.

Statistical power results are reported in Table 9. As shown, statistical power was very high for all cutoff points and showed a moderate relationship to coverage. A pronounced curvilinear relationship is also evident. Statistical power was relatively high at 100% coverage, increased moderately as coverage decreased to 60%. Below 50%, statistical power deteriorated rapidly. At extremely low levels of coverage (6%) statistical power was extremely low (e.g., indicating only a .02 probability of detecting a significant difference). Overall, statistical power was maximized at 60% coverage although it was high even at 100% coverage.

D. Between Method Comparisons

A major goal of this research was to document relations between coverage and both accuracy and statistical power for a variety of clustering procedures. An

Table 8

Effect Sizes for Centroid Analysis (intraclass correlation)

	Cutoff point (ICC)					
	.01	.1	.3	.5	.7	.9
External criterion						
1	.49	.50	.55	.87	.79	.99
2	.16	.16	.31	.54	.99	.99
3	.65	.64	.72	.63	.25	.10
4	.67	.65	.80	.86	.99	.82
5	.76	.75	.80	.99	.99	.99
6	.34	.34	.37	.56	.96	.99
7	.52	.52	.53	.50	.43	.21
Mean ES	.51	.51	.58	.71	.77	.72
Coverage (%)	100	98	87	60	29	6

Note: Table entries reflect effect size (f) calculated from analysis of variance according to procedures described by Cohen (1977). Separate one-way analyses of variance were used, with clusters (k=4) serving as the independent variable and the external criteria serving as the dependent variable.

Table 9

Statistical Power for Centroid Analysis (intraclass correlation)

	Cutoff point (ICC)					
	.0	.1	.3	.5	.7	.9
External criteria						
1	.98	.98	.99	.99	.84	.10
2	.08	.07	.48	.89	.99	.10
3	.99	.99	.99	.96	.05	.01
4	.99	.99	.99	.99	.99	.09
5	.99	.99	.99	.99	.99	.06
6	.72	.69	.71	.97	.99	.10
7	.98	.98	.99	.99	.20	.10
Mean SP	.82	.82	.88	.97	.72	.02
Coverage (%)	100	98	87	60	29	6

Note: Table entries reflect probability of detecting significant differences (p<.01) in external criteria between 4 clusters.

equally important goal, however, was to compare clustering methods against each other on the same data sets. Thus, three clustering methods--inverse factor analysis, Lorr's non-hierarchical clustering technique, and centroid clustering--were compared in terms of (a) accuracy in solving 20 multivariate normal mixtures and (b) effect size and statistical power these methods afford on external criteria on the test data set. Our goal was to be able to judge the relative merits of different clustering procedures on the same criteria, even though they construct clusters in vastly different ways. These comparisons provide a basis for selecting the best clustering procedures in future research using real data.

1. Accuracy comparisons. Direct comparisons among the three best clustering procedures (inverse factor analysis-varimax, Lorr's technique-product-moment correlation, and centroid clustering-intraclass correlation) is complicated by the fact that these procedures do not necessarily produce the same number of clusters or identical levels of coverage. Nevertheless, some comparisons between methods are possible, despite the fact that they construct clusters in different ways and employ different mechanisms for manipulating the coverage of the resulting classifications. These comparisons were possible here because the three grouping procedures were evaluated on the same 20 data sets and identical measures of accuracy and statistical power were employed.

Accuracy results for the three grouping procedures are reported in Table 10. This table reports mean accuracy values (lambda) and levels of coverage based on the analysis of the 20 multivariate normal mixtures. Table values reflect some minor interpolation between raw data points to permit comparisons across methods (i.e., coverage values were grouped according to 5% levels). Among the four inverse factoring procedures $j \neq k$ solutions were consistently more accurate than $j = k$ solutions for both the variman and direct quartimin rotations. In addition, varimax solutions were consistently more accorate than the direct quartimin solutions

Table 10

Accuracy Results for Inverse Factor Analysis, Lorr's Technique and Centroid Clustering.

	COVERAGE (%)																		
Method	100		90		80		70		60		50		40		30		20		10
Inverse Factor Analysis																			
Varimax (j=k)	.66	--	.68	--	.71	--	.74	--	--	--	.80	--	--	--	.83	--	--	--	.88
D-Quart (j=k)	.61	--	--	.66	.69	--	.73	--	--	.78	--	--	--	.82	--	--	--	.83	--
Varimax (j≠k)	.73	.76	.78	--	--	.81	--	--	.85	--	--	--	--	.90	--	--	--	--	.95
D-Quart (j≠k)	.71	.74	.76	--	--	.80	--	--	.84	--	--	--	--	.88	--	--	--	--	.90
Lorr's Technique																			
Correlation	--	.75	--	--	--	.87	.92	--	.92	--	.94	--	--	.73	.73	.69	--	--	--
Intraclass r	--	.77	--	--	--	.88	.94	--	.95	.95	.96	--	--	--	.74	.71	--	--	--
Centroid Clustering																			
Correlation	.69	.69	.72	--	--	.78	--	--	--	--	--	--	.85	--	--	--	--	.97	--
Intraclass	.73	.73	--	.76	--	--	--	--	.81	--	--	--	--	--	.91	--	--	--	1.00

NOTE: Table entries represent mean accuracy results (lambda) for 20 multivariate normal mixtures.

at all levels of coverage. Thus, the most accurate inverse factoring procedure was the varimax rotation with the number of clusters determined empirically via Cattell's scree test. For both Lorr's technique and centroid clustering, the one-way intraclass correlation yielded slightly higher accuracy than the product-moment correlation. This is presumably because the intraclass correlation utilizes information based on elevation and scatter of profile scores in addition to shape (cf. Edelbrock and McLaughlin, 1980).

Considering the most accurate of each of the three grouping procedures (inverse factor analysis-varimax, Lorr's technique-intraclass correlation, and centroid clustering-intraclass correlation), differences in accuracy are negligible in the range from 90-100% coverage. For example, the varimax procedure yielded clustering solutions with a mean accuracy of .76 for approximately 95% coverage, whereas Lorr's technique resulted in a mean accuracy of .77 and centroid clustering produced a mean accuracy of .73. These differences are not statistically reliable. At medium levels of coverage (50-85%) there was significant differences among the 3 procedures. Lorr's technique proved to be the most accurate method at medium levels of coverage, with accuracy values ranging from .88-.95. Inverse factor analysis was the next most accurate, with mean accuracy values ranging from .81-.85. Centroid clustering proved to be the lease accurate of the 3 procedures, with mean accuracy values ranging from .76-.81 at medium levels of coverage. At lower levels of coverage (10-45%) the situation was reversed, with inverse factor analysis and centroid clustering (.90-.95 and .91-1.00, respectively) performing better than Lorr's technique (.71-.74).

This finding indicates that comparative evaluation of clustering procedures depends upon level of coverage at which comparisons are made. At high levels of coverage, the three best clustering procedures were comparable in terms of accuracy. However, at medium levels of coverage, Lorr's method proved to be the

most accurate method, followed by inverse factor analysis, and finally the centroid clustering procedure.

2. Statistical Power Comparisons. Mean values for statistical power, averaged across the seven external criteria on Blashfield's data set #17 are reported in Table 11. These mean accuracy values are reported for the most accurate of each of the three grouping methods: (a) inverse factor analysis-varimax, (b) Lorr's technique-intraclass correlation, and (c) centroid clustering-intraclass correlation. As shown in Table 11, centroid clustering produced the highest levels of statistical power, whereas inverse factor analysis produced the lowest. The inverse factoring procedure produced lower levels of statistical power with one exception: at 50% coverage inverse factor analysis produced a higher level of statistical power than Lorr's technique (.79 vs. .65, respectively). However, in the range of coverage in which accuracy is maximized (55-85%), the centroid technique was clearly superior to the other two procedures in terms of statistical power. This suggests that the centroid technique is more likely than other methods to construct clusters that will differentiate among subjects in terms of external criteria. This is particularly true if the procedure is used to construc classifications having coverage in the range of 55-85% coverage. If classifications having high coverage are to be constructed (< 95% coverage), inverse factor analysis and Lorr's technique would appear to be promising methods.

E. Summary and Conclusions

Our research results have documented a strong relationshp between level of coverage and both accuracy of clustering solutions and the statistical power of cluster based classifications. For each of the three clustering procedures tested here, 100% coverage resulted in less than optimal accuracy in recovering predetermined populations from computer generated mixtures. For all three methods, the accuracy of clustering solutions was substantially boosted by leaving a portion

Table 11

Statistical Power Comparison for Three Grouping Methods

Method	COVERAGE																		
	100		90		80		70		60		50		40		30		20		10
A	.52	.60	--	.66	.67	--	.78	--	--	--	.79	--	--	--	--	--	.51	--	--
B	--	.57	--	.82	--	.79	.77	.82	--	.72	.65	--	--	.70	.60	.59	.49	--	--
C	.82	--	--	.88	--	--	--	--	.97	--	--	--	--	--	.72	--	--	--	.02

NOTE: Method A: Inverse Factor Analysis - Varimax rotation, $j \neq k$.

Method B: Lorr's non-hierarchical clustering technique - intraclass correlation.

Method C: Centroid clustering - intraclass correlation.

of subjects unclassified. In fact, leaving only 15-25% of the subjects unclassified resulted in a substantial increase in clustering accuracy for all methods. These results further suggest that the accuracy of clustering solutions can be maximized in the range of 55-85% coverage. For all of the methods we tested, increasing coverage above 85% had deleterious effects on clustering accuracy. In other words, pushing the level of coverage above 85% is likely to result in an increasing probability of misclassification.

Taken together, these results support the concept of a "continuum of classifiability" in the behavioral and social sciences. Based on these findings, we would not recommend that behavioral and social scientists attempt to classify 100% of their subjects into mutually exclusive groups. Alternatively, taxonomists should experiment with the heuristic and predictive value of classifications having less than 100% coverage. Our results further suggest that predictive power of empirically derived taxonomies can be maximized by constructing classifications having coverages of approximately 55-85%.

The statistical power results further indicate that the level of coverage of classifications is strongly related to the probability of detecting significant correlates of cluster membership. For all of the methods we tested, 100% coverage produced less than optimal probabilities of detecting significant differences among clusters. Instead, statistical power was maximized in the range of 55-85% coverage. At the other extreme, statistical power also deteriorated at low levels of coverage (at least for sample sizes in the range of 100-150). These results suggest that tyopologies or classifications having levels of coverage in the medium range will be optimally predictive of external criteria. Although a high level of coverage may be desirable when constructing classifications, it is clear that attempts to classify everybody will result in increasing misclassifications and reduced predictive power.

An additional point to be emphasized, however, is that with the development of methods for manipulating the coverage of empirically derived taxonomies, it is not necessary to derive a single classification of subjects with one level of coverage. Researchers should be encouraged to experiment with various levels of coverage (using procedures such as those developed and presented here). Researchers should also attempt to determine the optimal level of coverage for the classifications they derive. Of course, the level of coverage one seeks is also related to the purpose of the classification, If the goal is to optimize statistical power of a typology, coverage in the medium range may be warranted. Alternatively, if the goal is to characterize "pure types", lower levels of coverage would be useful. In an epidemiological study aimed at accounting for the generality of a phenomena in a population, very high levels of core coverage may be warranted.

The between method comparisons reported here indicate that the clustering procedures we tested are not equivalent in terms of their accuracy in grouping objects drawn from computer generated populations or in terms of predictive power. Lorr's non-hierarchical clustering procedure performed exceptionally well in both respects and should be given serious consideration in future clustering efforts. For Lorr's technique, both the product-moment correlation and the intraclass correlation resulted in high accuracy and high predictive power. Nevertheless, the intraclass correlation performed slightly better. In evaluating inclusion and exclusion criteria for Lorr's technique, inclusion criteria no higher than .01 and exclusion criteria no lower than .05 were optimal. The inclusion/exclusion combination of .01/.025 is recommended as the best compromise in terms of accuracy, coverage, and statistical power. This is, of course, only a general guideline, and the optimal inclusion/exclusion combination will depend upon many factors, including the number of clusters one seeks to derive, the sample size, etc.

Results on the inverse factoring procedures are also encouraging, considering that this method has been widely criticized as a grouping procedure. In our comparisons, varimax procedures performed quite well, both in terms of accuracy and statistical power. This was particularly the case when the number of clusters was determined empirically via Cattell's method. In fact, at high levels of coverage, the varimax procedure produced clustering solutions that were as accurate as those produced by Lorr's technique and the centroid clustering procedure.

The centroid clustering-neura centroid assignment procedure performed exceptionally well in terms of both accuracy and statistical power. Overall, this method produced very high levels of accuracy and statistical power. A major obstacle to this method, however, is the difficulty in determining the appropriate number of clusters. The other two procedures (inverse factor analysis and Lorr's technique) determine the number of clusters empirically. The centroid method, which is a hierarchical clustering procedure does not. Future research is clearly needed to operationalize reliable and objective rules for determining number of clusters in a hierarchical clustering solution (e.g., Mojena, 1977). Unfortunately, comprehensive research on various stopping rules for hierarchical clustering algorithms was beyond the scope of this research. It is possible that the development of objective stopping rules for the centroid clustering procedure would result in even greater accuracy and statistical power. Thus, we would recommend that the centroid clustering procedure be seriously considered, by cluster analysts doing methodological investigations as well as by applied researchers.

References

Achenbach, T.M., and Edelbrock, C.S. The classification of child psychopathology: A review and analysis of empirical efforts. Psychological Bulletin, 1978, 85, 1275-1301.

Anderberg, M.R. Cluster analysis for applications. New York: Academic Press, 1973.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders (2nd. ed.). Washington, DC: Author, 1968.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders (3rd. ed.). Washington, DC: Author, 1980.

Baggaley, A.R. Intermediate correlational methods. New York: Wiley, 1964.

Bailey, K.D. Cluster analysis. In D. Heise (Ed.), Sociological methodology. San Francisco: Jossey-Bass, 1974.

Barr, A.J., Goodnight, J.H., Sall, J.P., and Helwig, J.T. A user's guide to SAS. Raleigh: SAS Institute, 1976.

Bartko, J.J. On various intraclass correlation reliability coefficients. Psychological Bulletin, 1976, 83, 762-765.

Berzins, J., Ross, W.F., English, G.E., and Haley, J.V. Subgroups among opiate addicts: A typological investigation, Journal of Abnormal Psychology, 1974, 83, 65-73.

Blackburn, R. Personality types among abnormal homocides. British Journal of Criminology, 1971, 11, 14-31.

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

Blashfield, R. A consumer report on cluster analysis software. Report prepared for the National Science Foundation, 1977. (Available from the author, Department of Psychiatry, University of Florida, Gainesville, FL 32601).

Blashfield, R., & Morey, L.C. A comparison of four clustering methods using MMPI Monte Carlo data. Applied Psychological Measurement, 1980, 4, 57-64.

Brennen, T. Bibliographic materials on the use of classification in criminological research and its use within criminal justice systems. Report available from the author, Behavioral Research Institute, Boulder, CO, 1979.

Brennan, T., Huizinga, D., and Elliot, D.S. The social psychology of runaways. Lexington, MA: Heath, 1978.

Butler, E.W., and Adams, S.W. Typologies of delinquent girls: Some alternative approaches. Social Forces, 1966, 44, 401-407.

Carlson, K.A. Classes of adult offenders: A multivariate approach. Journal of Abnormal Psychology, 1972, 79, 84-93.

Cattell, R.B. r_p and other coefficients of pattern similarity. Psychometrika, 1949, 14, 279-298.

Cattell, R.B. The screen test for number of factors. Multivariate Behavioral Research, 1966, 1, 245-276.

Clinard, M.B., and Quinney, R. Criminal behavior systems: A typology. New York: Holt, Rinehart, and Winston, 1967.

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

Cohen, J. Statistical power analysis for the behavioral sciences (Rev. ed.). New York: Academic Press, 1977.

Collins, H.A., Burger, G.K., and Taylor, G.A. An empirical typology of heroin abusers. Journal of Clinical Psychology, 1976, 32, 473-476.

Cronbach, L.J., and Gleser, G. Assessing similarity between profiles. Psycyological Bulletin, 1953, 50, 456-473.

Dixon, W.J., and Brown, M.B. (Eds.). BMD biomedical computer programs: P-series. Los Angeles: University of California Press, 1977.

Edelbrock, C. Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody. Multivariate Behavioral Research, 1979, 14, 367-384.

Edlebrock, C., and Achenbach, T.M. Child Behavior Profile patterns of children referred for clinical services. Paper presented at the 86th annual convention of the American Psychological Association, Toronto, August, 1978.

Edelbrock, C., and Achenbach, T.M. A typology of Child Behavior Profile patterns: Distribution and correlates for disturbed children aged 6-16. Journal of Abnormal Child Psychology, in press, 1980.

Edelbrock, C., and McLaughlin, B. Hierarchical cluster analysis using intraclass correlations: A mixture model study. Multivariate Behavioral Research, in press, 1980.

Ferdinand, T.N. Typologies of delinquency: A critical analysis. New York: Random House, 1966.

Fleiss, J.L., Lawlor, W., Platman, S.R., and Fieve, R.R. On the use of inverted factor analysis for generating typologies. Journal of Abnormal Psychology, 1971, 77, 127-132.

Fleiss, J.L., and Zubin, J. On the methods and theory of clustering. Multivariate Behavioral Research, 1969, 4, 235-250.

Fruchter, B. Introduction to factor analysis. New York: van Nostrand, 1954.

Gibbons, D.C. Offender typologies: Two decades later. British Journal of Criminology, 1975, 15, 140-156.

Glaser, D. The classification of offenses and offenders. In D. Glaser (Ed.), The handbook of criminology. New York: Rand McNally, 1974.

Goldstein, S.G., and Linden, J.D. Multivariate classification of alcoholics by means of the MMPI. Jounral of Abnormal Psychology, 1969, 74, 661-669.

Gregson, R. Psychometrics of similarity. New York: Academic Press, 1975.

Gross, A.L. A Monte Carlo study of the accuracy of a hierarchical grouping procedure. Multivariate Behavioral Research, 1972, 7, 379-390.

Harman, H.H. Modern factor analysis (3rd. ed.). Chicago: University of Chicago Press, 1976.

Hindelang, M.J., and Weis, J.G. Personality and self-reported delinquency: An application of cluster analysis. Criminology, 1972, 10, 268-294.

Hood R., and Sparks, R. The classification of crimes and criminals. In R. Hood and R. Sparks (Eds.), Key issues in criminology. New York: McGraw-Hill, 1970.

Hubert, L.J., and Levin, J.R. Evaluating object set partitious: Free sort analysis and some generalizations. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 459-470.

Jones, K.J. Problems of grouping individuals and the method of modality. Behavioral Science, 1968, 13, 496-511.

Katz, M.K., and Cole J.O. A phenomenological approach to the classification of schizophrenic disorders. Diseases of the Nervous System, 1963, 24, 147-154.

Kuiper, F.K., and Fisher, L. A Monte Carlo comparison of six clustering procedures. Biometrics, 1975, 31, 777-783.

Lorr, M. (Ed.). Explorations in typing psychotics. Oxford: Pergammon Press, 1966.

Lorr, M., Bishop, P.F., and McNair, D.M. Interpersonal types among psychotic patients. Journal of Abnormal Psychology, 1965, 70, 468-472.

Lorr, M., Pokorny, A.D., and Klett, C.J. Three depressive types. Journal of Clinical Psychology, 1973, 29, 290-294.

Lorr, M., and Radhakrishnan, B.K. A comparison of two methods of cluster analysis. Educational and Psychological Measurement, 1967, 27, 531-543.

Megargee, E.I. A new classification system for criminal offenders: I. The need for a new system. Criminal Justice and Behavior, 1977, 4, 107-114.

Megargee, E.I., and Bohn, M.J. A new classification system for criminal offenders: IV. Empirically determined correlates of the ten types. Criminal Justice and Behavior, 1977, 4, 149-210.

Megargee, E.I., and Dorhout, B. A new classification system for criminal offenders: III. Revision and refinement of the classificatory rules. Criminal Justice and Behavior, 1977, 4, 125-148.

Meyer, J., and Megargee, E.I. A new classification system for criminal offenders: II. Initial development of the system. Criminal Justice and Behavior, 1977, 4, 115-124.

Mezzich, J.E. Evaluating clustering methods for psychiatric diagnosis. Biological Psychiatry, 1978, 13, 265-281.

Milligan, G.W. and Isaac, P. The validation of four ultrametric clustering algorithms. Pattern Recognition, 1980, 12, 41-50.

Mojena, R. Hierarchical grouping methods and stopping rules: An evaluation. Computer Journal, 1977, 20, 359-363.

Monro, A.B. Psychiatric types: A Q-technique study of 200 patients. Journal of Mental Science, 1955, 101, 330-343.

Mulaik, S. The foundations of factor analysis. New York: McGraw-Hill, 1972.

National Institutes of Health. C-LAB: An on-line clustering laboratory. Available from the Division of Computer Research and Technology, NIH, Bethesda, MD 20205, 1979.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H. SPSS: Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill, 1975.

Opp, K.D. Problems of classification in criminology. Collected Studies in Criminological Research (Vol. 10). Council of Europe, 1973.

Overall, J, and Klett, C.J. Applied multivariate analysis. New York: McGraw-Hill, 1972.

Phillips, L., and Draguns, J.G. Classification of the behavior disorders. Annual Review of Psychology, 1971, 22, 447-482.

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

Ryder, R.G. Profile factor analysis and variable factor analysis. Psychological Reports, 1964, 15, 119-127.

Sclove, S. Population mixture models and clustering algorithms. Communications on Statistical Theory and Methods, 1977, 6, 417-434.

Sneath, P.H.A. The application of computers to taxonomy. Journal of General Microbiology, 1957, 17, 201-226.

Sneath, P.H., and Sokol, R.R. Numerical taxonomy: The principles and practice of numerical classification. San Francisco: W.H. Freeman, 1973.

Sokal, R.R., and Michener, C.D. A statistical method for evaluating systematic relationships. University of Kansas Scientific Bulletin, 1958, 38, 1409-1438.

Sokal, R.R., and Sneath, P.H.A. Principles of numerical taxonomy. San Francisco: W.H. Freeman, 1963.

Stephenson, W. Introduction to inverted factor analysis with some applications to studies of orexsis. Journal of Educational Psychology, 1936, 27, 353-367.

Tatsouka, M.M. Classification procedures: Profile similarity. In M.M. Tatsouka (Ed.), Selected topics in advanced statistics (No. 3). Champaign: Institute for Personality and Ability Testing, 1974.

Warren, M.Q. Classification of offenders as an aid to efficient management and effective treatment. Journal of Criminal Law, Criminology, and Police Science, 1970, 62, 239-258.

Wishart, D. CLUSTAN user's manual (3rd. ed.). Edinburgh: Edinburgh University Press, 1978.

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

Zigler, E., and Phillips, L. Psychiatric diagnosis: A critique. Journal of Abnormal and Social Psychology, 1961, 63, 607-618.

Appendix A

Blashfield's multivariate normal mixture #17

```
9,126
4
36,30,35,25
(I3,6X,9F7.3/)
BLASHFIELD DATA SET 17 MIX17.DAT
 17   1  1 60.340 54.450 31.090 47.366 49.555 44.217 39.482 42.662 57.960 60.455
 17          55.120 57.820 37.394 48.418 48.461 50.768
 17   1  2 56.373 57.155 35.772 43.578 48.686 47.238 43.038 45.584 58.043 57.893
 17          51.767 45.981 40.274 47.552 53.348 51.641
 17   1  3 55.455 55.470 28.816 45.621 48.560 44.981 41.964 44.257 52.645 58.357
 17          54.317 56.540 38.657 44.471 50.696 44.147
 17   1  4 60.415 47.471 49.670 49.143 50.913 38.488 34.377 40.749 51.788 64.180
 17          56.369 53.610 39.252 60.264 44.935 39.473
 17   1  5 65.395 35.209 47.207 60.807 61.632 29.217 26.566 34.332 43.169 75.071
 17          65.618 47.140 34.406 64.050 31.020 41.049
 17   1  6 56.095 49.052 44.774 50.476 50.150 38.715 36.174 41.146 49.489 63.726
 17          55.318 32.522 39.390 68.558 45.223 36.211
 17   1  7 60.959 47.676 46.361 49.870 52.961 40.620 36.887 41.086 53.969 64.137
 17          57.846 37.908 36.058 58.874 44.856 40.174
 17   1  8 54.993 52.030 50.393 45.497 50.460 42.614 38.003 42.913 53.445 61.026
 17          52.559 52.286 42.096 55.463 50.257 39.781
 17   1  9 66.157 42.624 46.468 55.119 55.031 35.852 32.701 37.049 48.876 69.845
 17          62.748 52.650 33.340 54.180 37.921 57.798
 17   1 10 56.388 61.726 53.736 38.379 40.561 49.924 45.362 49.726 59.047 50.209
 17          45.965 36.669 47.061 59.982 60.907 35.612
 17   1 11 51.443 63.232 43.909 37.780 41.761 50.580 45.748 49.055 61.071 49.360
 17          45.479 48.753 46.494 41.587 62.204 34.611
 17   1 12 53.733 57.640 46.000 44.861 46.674 46.529 45.953 45.611 59.629 57.096
 17          51.895 64.556 42.548 62.166 53.961 52.225
 17   1 13 62.886 44.259 47.239 54.497 52.805 37.110 32.680 39.098 48.279 65.607
 17          59.257 48.002 38.638 52.719 40.587 35.715
 17   1 14 58.238 52.835 40.809 47.733 52.661 42.907 40.774 43.073 54.804 63.184
 17          53.898 53.148 38.515 48.852 48.177 56.838
 17   1 15 63.588 41.242 31.807 58.857 55.055 33.482 29.596 35.497 48.322 71.186
 17          65.772 46.430 33.833 57.856 35.426 51.705
 17   1 16 55.389 56.356 40.940 45.411 51.227 46.094 42.959 44.171 56.706 60.268
 17          53.759 55.205 38.286 45.958 50.589 53.040
 17   1 17 65.315 38.869 54.569 58.382 57.301 32.476 28.326 35.220 45.791 71.261
 17          61.368 35.846 35.849 67.855 33.498 35.068
 17   1 18 55.522 63.107 36.876 39.378 43.565 51.766 47.560 48.638 65.053 52.246
 17          47.157 56.945 44.193 46.530 60.249 51.098
 17   1 19 52.835 59.042 46.231 39.410 47.388 49.281 46.645 48.789 62.672 53.237
 17          46.100 50.483 43.941 42.249 58.510 42.140
 17   1 20 64.778 42.455 57.233 55.993 57.232 33.954 30.742 37.772 49.502 70.192
 17          61.155 53.146 36.312 50.530 36.993 41.823
 17   1 21 47.896 66.109 23.971 36.489 40.818 53.771 49.267 48.660 63.619 50.313
 17          45.911 43.876 40.568 40.745 61.896 33.714
 17   1 22 60.022 45.981 35.398 53.291 50.755 37.021 33.093 37.365 48.668 67.475
 17          61.689 42.926 37.286 68.498 40.770 38.429
 17   1 23 61.228 48.509 43.384 52.499 52.914 39.473 37.043 40.791 51.708 65.010
 17          57.759 51.681 39.062 63.345 43.530 52.841
 17   1 24 57.932 56.722 43.519 44.339 48.461 46.295 40.378 45.517 54.612 57.462
 17          50.721 48.882 43.102 46.548 55.179 46.674
 17   1 25 59.811 49.425 38.744 52.344 51.757 38.598 36.089 40.571 50.308 64.696
 17          57.845 35.437 38.887 63.755 43.491 40.770
 17   1 26 57.909 54.275 48.988 43.994 47.016 44.901 42.576 44.872 55.545 58.833
 17          53.687 53.730 42.309 52.124 52.526 39.154
 17   1 27 60.452 57.249 46.197 44.527 46.991 46.110 42.233 45.762 57.363 57.033
 17          53.125 52.875 43.204 54.150 54.482 47.752
 17   1 28 55.737 46.479 26.344 54.598 54.539 37.267 37.930 38.632 49.121 68.406
 17          60.685 49.959 34.544 59.349 39.429 49.891
 17   1 29 56.006 57.319 46.721 46.840 49.319 44.844 39.582 43.868 56.776 60.329
 17          53.858 51.825 42.365 59.754 51.826 40.961
 17   1 30 55.672 56.029 31.731 44.323 47.131 46.566 40.428 44.599 58.021 58.077
 17          50.854 40.532 41.206 51.547 52.779 46.399
 17   1 31 58.591 52.894 42.813 47.823 49.699 41.965 36.791 42.950 54.602 62.240
 17          54.304 47.928 37.557 43.480 47.994 44.019
 17   1 32 62.484 48.064 40.055 50.873 52.699 38.689 34.580 39.283 51.529 66.062
 17          59.526 46.153 35.102 63.316 41.941 38.589
 17   1 33 54.642 56.225 36.983 46.136 48.838 44.644 38.754 43.632 57.464 59.775
 17          51.120 45.159 40.357 50.601 51.467 46.675
 17   1 34 59.509 57.395 48.814 42.269 47.325 46.859 41.440 46.130 57.235 56.333
 17          50.924 39.148 43.291 44.941 54.895 31.475
 17   1 35 59.617 48.600 50.167 49.993 51.324 40.538 36.245 41.566 50.421 65.136
 17          57.047 41.633 38.023 67.513 44.317 39.475
 17   1 36 58.415 53.490 37.069 47.861 50.868 42.103 38.427 42.402 51.440 60.876
 17          53.583 36.587 39.738 56.223 46.788 39.941
 17   2  1 54.969 64.507 51.570 62.837 52.125 39.081 56.376 49.371 56.170 60.463
 17          49.035 53.889 59.164 64.093 34.858 58.753
 17   2  2 45.968 44.437 55.979 55.625 65.251 41.908 59.047 60.509 43.153 48.733
 17          65.496 60.608 54.297 54.038 46.771 44.710
 17   2  3 47.084 32.349 58.300 57.226 49.472 44.011 61.009 61.393 49.953 47.499
 17          62.545 64.141 65.133 58.021 49.584 41.485
 17   2  4 44.918 63.431 50.987 60.817 43.655 38.781 59.517 50.003 48.753 60.593
 17          51.088 56.161 53.754 61.704 35.738 53.188
 17   2  5 54.167 68.979 48.851 65.440 49.572 35.227 53.226 46.732 51.659 67.959
 17          40.272 52.323 83.527 66.221 30.027 62.330
 17   2  6 48.782 55.679 49.228 59.846 52.911 37.463 56.940 54.497 49.346 58.499
 17          59.088 60.596 64.899 58.142 38.295 49.417
 17   2  7 47.213 54.131 57.880 54.864 53.610 47.554 62.152 60.643 45.421 53.325
 17          64.466 58.232 47.981 55.126 44.720 40.776
 17   2  8 45.883 51.981 51.451 54.999 40.622 52.200 66.569 63.162 44.191 48.691
 17          70.564 64.071 51.127 56.561 48.770 41.315
 17   2  9 40.889 35.430 61.124 48.408 48.583 53.471 65.551 50.428 39.219 43.880
 17          77.053 66.885 42.281 46.999 50.256 30.344
 17   2 10 51.943 58.597 48.556 57.780 52.393 47.780 63.595 58.178 50.518 57.744
 17          56.009 53.562 49.284 57.120 40.586 49.511
 17   2 11 48.822 57.465 47.734 56.725 53.396 42.025 60.437 53.085 47.705 61.320
 17          62.119 57.716 58.998 58.198 34.321 47.230
 17   2 12 47.362 51.070 51.228 54.434 48.389 51.774 66.493 41.083 44.579 48.983
 17          59.913 61.545 57.300 54.090 44.098 40.741
 17   2 13 39.580 45.185 64.596 54.590 62.299 46.918 60.561 55.370 41.916 47.179
 17          75.742 62.272 52.984 54.243 40.438 39.983
 17   2 14 55.811 51.853 53.194 58.684 57.663 40.419 53.909 50.280 50.194 56.532
 17          54.789 59.636 64.183 60.871 37.923 52.603
 17   2 15 52.490 54.891 57.448 58.475 58.312 41.299 58.635 52.348 54.619 59.712
 17          50.557 56.996 48.731 58.145 34.270 49.588
 17   2 16 44.499 58.754 47.342 58.777 60.798 37.598 58.140 67.639 46.059 55.090
 17          52.001 59.606 46.270 60.202 36.691 57.347
 17   2 17 51.500 60.541 54.197 60.359 50.814 36.268 57.216 45.895 54.757 64.194
 17          53.467 54.200 67.913 60.473 40.486 47.772
 17   2 18 52.941 67.245 52.011 63.349 61.179 38.427 57.539 53.708 54.329 61.082
 17          43.533 58.730 62.687 65.540 34.713 63.076
 17   2 19 55.383 62.152 59.649 62.695 49.856 38.854 56.608 55.382 51.804 66.923
 17          53.166 54.759 70.050 65.186 35.639 60.442
 17   2 20 47.069 67.347 53.464 60.592 62.843 40.847 58.636 50.727 44.512 56.510
 17          52.461 61.170 70.546 60.356 35.038 51.017
 17   2 21 41.399 46.268 53.104 52.440 52.970 49.507 63.429 57.415 36.809 50.290
 17          70.106 65.226 44.380 52.711 42.254 34.775
 17   2 22 58.603 57.963 44.374 58.871 47.453 42.695 63.783 64.890 52.165 65.965
```

17			61.533	54.009	53.552	59.156	40.176	53.322				
17	2	23	60.217	60.661	52.660	61.305	52.321	44.042	59.390	52.419	46.759	58.380
17			59.913	60.189	74.962	60.745	39.122	54.227				
17	2	24	44.500	52.354	52.841	54.090	52.534	46.588	54.315	57.294	40.398	52.359
17			66.004	63.511	43.910	58.461	37.079	37.519				
17	2	25	44.158	62.558	48.837	60.047	47.352	39.926	58.029	60.951	54.926	62.081
17			58.607	55.076	58.601	62.007	37.981	51.287				
17	2	26	53.218	40.766	68.677	52.672	58.732	55.125	59.828	58.939	35.312	43.519
17			72.640	69.023	46.992	50.893	50.027	40.831				
17	2	27	48.571	69.507	58.102	59.150	47.696	43.611	57.611	57.647	51.260	61.256
17			54.748	59.530	47.685	58.654	36.561	51.331				
17	2	28	45.116	46.593	61.646	54.852	55.918	48.975	59.921	63.114	36.331	42.322
17			72.762	66.620	54.547	54.565	52.676	46.024				
17	2	29	49.651	51.253	56.189	56.700	66.456	46.900	58.081	69.181	42.962	52.039
17			67.537	57.749	63.772	60.631	44.863	39.646				
17	2	30	45.468	58.011	43.187	61.721	53.647	37.246	56.041	58.656	52.327	60.795
17			46.741	58.514	63.428	62.220	28.464	54.716				
17	3	1	62.959	38.709	63.823	33.704	58.699	45.254	46.974	61.617	51.906	55.601
17			47.817	56.835	61.404	47.215	57.626	44.937				
17	3	2	60.600	40.077	60.200	35.600	49.600	34.093	34.089	60.966	37.055	46.658
17			17.761	41.754	69.413	44.056	49.076	43.922				
17	3	3	60.537	38.360	60.074	49.164	56.532	45.603	54.137	57.855	50.011	58.143
17			33.243	61.486	60.527	46.720	47.190	50.679				
17	3	4	55.453	53.356	36.107	47.983	60.030	59.193	54.207	56.863	50.762	56.171
17			46.430	56.481	39.071	33.729	33.863	54.128				
17	3	5	54.336	47.142	50.831	42.530	52.163	51.115	49.242	54.764	49.690	59.181
17			41.054	59.428	51.955	41.987	45.361	51.631				
17	3	6	56.930	54.868	41.078	46.338	61.926	57.458	55.170	58.712	57.358	58.929
17			53.096	74.418	43.823	31.335	37.668	56.839				
17	3	7	56.648	56.427	41.532	49.080	60.502	53.442	53.657	48.203	60.184	59.437
17			51.246	64.782	32.630	34.863	32.711	56.755				
17	3	8	63.254	63.502	54.952	47.370	57.382	48.907	41.451	58.297	36.975	68.247
17			43.912	52.349	50.015	57.084	47.862	50.779				
17	3	9	59.542	57.434	45.446	40.811	52.882	43.899	42.060	55.306	37.810	67.362
17			44.216	48.297	45.766	47.702	52.067	51.197				
17	3	10	55.490	44.631	49.482	47.638	58.034	44.135	38.797	53.724	25.405	57.423
17			31.072	51.412	48.983	35.459	35.604	47.993				
17	3	11	57.171	47.784	55.075	35.556	56.468	53.606	46.464	54.341	50.351	59.580
17			38.856	42.350	49.311	53.521	51.613	49.798				
17	3	12	57.886	59.603	50.886	49.980	62.872	60.857	52.658	58.253	48.287	62.359
17			48.537	61.640	44.093	36.384	45.455	55.013				
17	3	13	59.498	63.319	52.645	43.711	56.476	48.267	45.425	58.358	39.166	63.394
17			42.668	59.943	55.357	50.733	44.796	48.272				
17	3	14	67.107	52.592	50.901	43.180	57.606	50.991	50.488	57.441	46.114	63.994
17			46.255	52.560	44.458	41.958	48.473	49.880				
17	3	15	53.464	46.713	62.079	35.953	55.959	48.272	52.482	61.321	48.781	66.081
17			42.825	60.058	56.840	58.876	56.865	49.882				
17	3	16	55.126	57.118	47.454	46.988	62.072	56.285	61.125	58.641	54.749	71.342
17			50.934	58.039	45.142	38.818	39.323	57.162				
17	3	17	54.450	50.553	50.211	42.044	61.671	52.757	50.082	53.843	59.564	64.315
17			48.905	62.903	43.267	30.360	36.263	52.348				
17	3	18	60.559	54.033	51.763	47.275	56.518	50.456	49.581	61.753	45.672	54.602
17			38.650	65.426	56.546	58.869	56.198	50.449				
17	3	19	56.948	49.552	41.370	49.185	61.233	50.643	50.638	56.298	51.764	62.416
17			55.486	63.391	38.245	38.871	28.186	56.327				
17	3	20	59.639	45.628	61.241	44.093	57.168	46.190	45.703	59.224	46.519	56.838
17			36.287	54.128	58.322	51.281	44.625	50.965				
17	3	21	68.365	28.941	60.047	27.992	47.964	27.758	21.102	65.110	35.325	49.713
17			25.580	43.996	66.621	58.261	48.212	35.900				
17	3	22	71.458	46.964	59.494	41.164	54.889	38.124	35.690	61.332	43.662	51.949
17			35.385	46.734	55.464	47.285	45.678	42.652				
17	3	23	60.549	45.826	54.820	33.448	52.835	52.563	51.814	54.486	41.781	66.009
17			44.068	48.839	43.723	46.652	47.487	49.992				
17	3	24	68.498	41.441	50.001	37.359	50.934	44.243	38.156	57.033	44.452	58.141
17			46.281	60.804	48.611	48.469	43.780	45.330				
17	3	25	60.506	36.148	50.447	49.597	60.461	58.123	56.462	57.347	37.697	66.182
17			49.400	60.396	46.571	28.030	44.130	52.573				
17	3	26	55.380	44.908	60.450	41.727	55.313	47.754	45.496	61.592	38.699	57.479
17			31.011	57.381	63.909	55.735	50.750	45.839				
17	3	27	64.493	35.912	68.216	34.482	52.891	48.015	45.577	65.278	38.562	59.905
17			29.161	50.043	74.241	50.053	51.285	45.846				
17	3	28	58.842	56.015	51.522	47.306	59.102	56.235	66.079	58.888	52.972	73.952
17			64.433	62.335	49.896	44.720	53.867	59.988				
17	3	29	51.779	47.203	44.366	34.275	55.002	55.342	48.442	49.850	58.610	53.581
17			50.992	52.288	40.923	40.462	37.864	50.655				
17	3	30	62.686	55.632	59.192	45.506	60.743	48.696	56.136	60.946	43.542	59.055
17			44.680	53.997	58.561	50.782	59.367	49.637				
17	3	31	60.408	46.721	52.468	40.571	53.276	49.298	38.500	63.312	39.680	55.390
17			29.987	51.958	56.152	63.784	47.609	49.809				
17	3	32	57.534	58.872	49.916	48.774	59.024	50.252	45.318	56.040	41.389	53.647
17			37.800	52.337	53.780	43.562	44.260	45.633				
17	3	33	61.086	37.783	54.256	37.022	49.206	42.225	41.671	57.218	31.431	56.865
17			24.325	41.844	66.490	61.473	41.995	48.015				
17	3	34	53.021	61.130	44.010	54.615	63.309	52.967	51.933	56.672	48.311	63.828
17			49.483	69.846	42.205	25.232	41.475	56.044				
17	3	35	54.669	58.374	38.372	46.333	66.188	59.974	59.417	57.217	59.205	70.220
17			67.336	70.814	31.417	30.828	38.541	62.604				
17	4	1	51.065	50.278	32.098	47.463	50.901	53.466	32.197	39.567	49.960	59.056
17			41.105	37.404	45.429	43.139	37.801	31.193				
17	4	2	58.153	46.997	50.291	55.443	62.701	57.410	49.993	51.075	52.251	66.631
17			52.502	55.176	56.063	36.225	47.838	48.403				
17	4	3	64.854	52.500	43.308	51.862	55.941	57.583	42.483	41.424	52.909	63.477
17			55.043	46.384	45.623	47.320	38.455	52.883				
17	4	4	52.613	44.114	48.034	58.505	60.649	59.432	52.760	49.375	53.629	61.530
17			55.849	50.339	52.077	44.742	53.956	50.793				
17	4	5	61.919	47.660	24.426	43.172	53.981	49.754	35.550	37.366	53.676	54.614
17			38.887	35.998	35.963	43.692	27.705	34.465				
17	4	6	44.779	46.908	49.167	52.552	55.686	55.109	44.152	45.467	57.333	60.837
17			50.515	50.943	53.050	47.533	53.913	47.274				
17	4	7	72.524	46.801	51.993	54.845	60.445	55.279	45.266	39.727	54.401	45.420
17			48.853	45.449	63.110	35.552	44.614	44.531				
17	4	8	34.313	51.800	31.514	50.759	59.161	60.407	51.075	42.702	56.326	62.052
17			49.652	37.117	39.266	43.987	32.712	51.816				
17	4	9	56.946	48.120	35.926	52.469	53.345	53.024	41.147	44.100	49.970	58.499
17			50.865	43.584	46.973	36.361	38.634	42.601				
17	4	10	49.599	43.538	33.638	46.975	47.350	48.396	32.745	40.479	48.856	46.227
17			38.768	40.127	48.185	46.831	33.166	30.255				
17	4	11	61.375	53.429	39.842	46.587	48.096	49.293	21.734	37.845	49.297	47.900
17			33.481	30.253	52.081	47.702	40.924	21.649				
17	4	12	63.641	46.872	42.362	50.535	57.294	53.015	35.863	40.975	50.518	52.600
17			44.242	41.658	50.685	34.122	41.932	41.002				
17	4	13	59.491	46.919	53.394	53.723	50.113	55.111	33.691	43.836	49.217	53.445
17			45.287	35.328	54.971	49.812	49.529	36.071				
17	4	14	35.745	46.300	59.033	55.106	55.852	56.390	43.751	49.635	54.504	68.273
17			47.388	49.366	64.462	42.792	56.292	54.417				
17	4	15	42.238	49.296	41.741	55.130	57.635	56.967	41.664	41.917	52.130	72.119
17			53.969	47.082	45.884	61.383	46.656	51.648				
17	4	16	44.690	47.475	26.732	53.028	54.439	55.414	42.990	37.488	55.870	60.723
17			48.342	39.386	38.652	48.276	39.081	39.005				
17	4	17	57.091	44.899	54.716	57.609	59.086	55.979	49.026	51.057	50.710	69.548

17			54.263	47.402	61.640	36.653	51.684	50.124				
17	4	18	52.927	47.029	37.810	52.708	59.450	58.306	39.703	48.486	52.188	51.982
17			45.745	38.215	51.823	47.075	48.309	40.838				
17	4	19	53.724	44.959	30.034	45.786	50.870	51.213	36.866	35.245	59.035	48.936
17			43.373	34.452	37.974	42.303	33.667	37.308				
17	4	20	63.756	49.125	32.137	47.349	48.139	48.791	29.100	40.552	51.878	42.508
17			38.310	28.540	51.818	52.581	43.411	35.678				
17	4	21	48.772	48.986	36.671	50.266	52.812	53.954	40.557	43.397	53.470	48.418
17			41.957	30.154	49.218	40.171	38.282	38.467				
17	4	22	41.931	49.557	55.240	55.995	59.043	60.352	44.010	48.175	56.254	77.219
17			52.956	51.143	58.549	44.135	51.711	49.832				
17	4	23	55.600	50.612	39.839	49.979	55.359	52.588	36.545	44.839	48.054	46.749
17			40.442	31.432	51.961	41.151	38.224	31.740				
17	4	24	48.330	45.269	33.043	50.604	56.157	58.910	46.062	45.432	55.758	65.278
17			51.948	46.752	40.632	39.429	36.446	54.477				
17	4	25	57.751	44.168	44.825	51.126	56.008	54.574	45.696	43.206	52.200	66.100
17			43.497	41.273	52.064	37.331	40.895	38.333				

Appendix B

FORTRAN Listing for Lorr's non-hierarchical clustering technique

```
**********************************************************
*                                                        *
*     BRIEF DOCUMENTATION FOR LORR'S NON-                *
*     HIERARCHICAL CLUSTERING TECHNIQUE                  *
*                                                        *
**********************************************************
```

PROGRAM CORR.FOR

THIS PROGRAM COMPUTES THE PEARSON CORRELATION COEEFICIENT BETWEEN SUBJECTS. IT GENERATES AN OUTPUT DATA FILE OF CORRELATION COEEFICIENTS.

PROGRAM INTRA.FOR

THIS PROGRAM IS SIMILAR TO CORR.FOR EXCEPT THAT THE ONE-WAY INTRA CLASS CORRELATION COEFFICIENT IS USED.

PROGRAM LORR.FOR

THIS PROGRAM USES LORR'S TECHNIQUE TO FIND A CLUSTERING SOLUTION.

THESE PROGRAMS ALL REQUIRE SUBROUTINES FROM THE IBM PACKAGE INTERNATIONAL AND MATHEMATICAL STATISTICAL LIBRARY.

PROGRAM USAGE

FIRST RUN EITHER CORR.FOR OR INTRA.FOR TO GENERATE A RELATIONSHIP MATRIX TO BE PLACED ON DISK. EACH PROGRAM WILL PROMPT FOR THE RAW DATA SET TO BE CLUSTERED. NEXT RUN LORR.FOR TO GENERATE THE CLUSTERING SOLUTION.

DATA PREPARATION

THE DATA SET TO BE CLUSTERED MUST HAVE SOME INFORMATION AT THE START OF THE FILE.

FOR EXAMPLE:

```
N,M
NCUT
(CUT(I),I=1,NCUT)
FORMAT
TITLE
RAW DATA
```

WHERE N IS THE NUMBER OF FACTORS, M IS THE NUMBER OF CASES, NCUT IS THE NUMBER OF CUT POINTS IF GROUPS ARE KNOWN IN ADVANCE, CUT(I) ARE THE CUT VALUES WHEN THE GROUPS ARE KNOWN IN ADVANCE(THESE MUST ADD UP TO M). FORMAT IS THE FORMAT STATEMENT TO USE WHEN READING THE RAW DATA, AND TITLE IS ANY TITLE UP TO 80 CHARACTERS. WHEN THE GROUPS ARE NOT KNOWN

IN ADVANCE NCUT AND CUT(I) CAN BE SET TO ARBITRARY VALUES WHICH SATISFY THE RESTRICTIONS.

```
**********************************************************
*                                                        *
*                      LORR.FOR                          *
*          THIS PROGRAM USES LORR'S NON-                 *
*      HIERARCHICAL CLUSTERING PROCEDURE TO              *
*            IDENTIFY HOMOGENEOUS SUBGROUPS OF           *
*      SUBJECTS.  EITHER THE PRODUCT-MOMENT              *
*    CORRELATION OR THE INTRACLASS CORRELATION           *
*     MAY BE USED THE MEASURE OF PROFILE SIMILARITY      *
*                                                        *
**********************************************************

      COMMON IDEL(149),PROF(149),R(149,149),ICLUS(149),
     1 NR,NC,CL,CU,NDEL,NPROF
      DIMENSION ICTOT(30),ICMAX(30),IVMAX(30)
     1 ,IRMAX(30),IRTOT(30),IWMAX(30)
     1 ,IOUT(149)
      DIMENSION ICROS(20,0/20)
      INTEGER PROF
      DIMENSION TITLE(20),NCS(6)
      READ(12)TITLE
      READ(12)METH
D     WRITE(6,126)TITLE
D126  FORMAT('  TITLE ',20A5)
 105  FORMAT(20A5)
      READ(12)N,M,IDNUM
      NR=M
      NC=M
D     WRITE(6,125)M
D125  FORMAT(' NO. ROWS ',I10)
      READ(12)NCUT
      READ(12)(NCS(I1),I1=1,NCUT)
      IT=0
      K1=1
      M1=1
D     WRITE(6,122)NCUT,NCS
 122  FORMAT(' NCUT ',I10,' CUTS ',6I5)
      DO 106 J=1,M
      READ(12)(R(J,I),I=1,K1)
      K1=K1+1
 106  CONTINUE
 12   CONTINUE
CDO 11 I=2,NC
CWRITE(6,13)((R(I,J),J=1,I-1))
C13   FORMAT(8F10.6)
C11   CONTINUE
      READ(1,502)CL
      READ(1,502)CU
      READ(1,502)IWRIT
      IF(CL.GT.CU)WRITE(6,611)
 611  FORMAT(' WARNING FIRST PROB MUSTBE LOWER THAN SECOND ON INPUT ')
C     COMPUTE T VALUE FOR SPECFIED PROBABILITY LEVEL
C     USE R**2=T**2/(T**2+(N-2)) TO COMPUTE CL AND CU
      IF(CL.GT.CU)STOP
      SCL=CL
      SCU=CU
```

```
      IF(METH.EQ.2)GO TO 131
      Q1=2.*CL
      Q2=2.*CU
      F=N-2
      CALL MDSTI(Q1,F,X,IER)
      CL=X
      CALL MDSTI(Q2,F,X,IER)
      CU=X
      CL=(CL**2)/(CL**2+F)
      CU=(CU**2)/(CU**2+F)
      CU=CU**.5
      CL=CL**.5
 502  FORMAT(1G)
      IF(METH .EQ.1)GO TO 132
 131  CONTINUE
      SNU1=N-1
      SNU2=N
      X=1-CL
      CALL MDFI(X,SNU1,SNU2,CL,IER)
      CL=(CL-1.)/(CL+1)
      X=1-CU
      CALL MDFI(X,SNU1,SNU2,CU,IER)
      CU=(CU-1.)/(CU+1.)
 132  CONTINUE
      DO 11 I=1,NR
      DO 11 J=1,NC
 11   IF(R(I,J).NE.0.0)R(J,I)=R(I,J)
D     WRITE(9,303)((R(I,J),I=1,NC),J=1,NR)
D303  FORMAT(22F6.2)
 5    CONTINUE
      CALL BUILD(STPE)
      IF(STPE.EQ.1.0)GO TO 100
      K=K+1
      NCLUS=K
      DO 10 I=1,NPROF
 10   ICLUS(PROF(I))=K
C     MARK COORELATION MATRIX FOR PROFILES USED
      DO 20 I=1,NDEL
      DO 20 J=1,NR
      IF(IDEL(I).EQ.0)GO TO 20
      R(IDEL(I),J)=-1.0E+30
      R(J,IDEL(I))=-1.0E+30
 20   CONTINUE
      DO 30 I=1,NPROF
      DO 30 J=1,NR
CWRITE(6,77)NPROF,PROF(I)
 77   FORMAT(' NPROF ',I10,2X,' PROF ',I10)
      R(PROF(I),J)=-1.0E+30
 30   R(J,PROF(I))=-1.0E+30
D     WRITE(6,99)ICLUS
 99   FORMAT(' SOLUTION ',11I5)
      IF(STPE.EQ.1.0)GO TO 100
      IF(NPROF.GE.4)GO TO 5
 100  CONTINUE
      NTOT=0
      IS=0
      DO 500 I=1,NCUT
      NTOT=NTOT+NCS(I)
      NCS(I)=NCS(I)+IS
 500  IS=NCS(I)
```

```
      K=1
      DO 510 I=1,NTOT
      M=ICLUS(I)
      IF(I.GT.NCS(K))K=K+1
 510  ICROS(K,M)=ICROS(K,M)+1
      WRITE(6,610)TITLE
      IF(METH .EQ. 1)WRITE(6,613)
 613  FORMAT(/'          PEARSON CORRELATION'/)
      IF(METH.EQ. 2)WRITE(6,614)
 614  FORMAT(/'          INTRACLASS CORRELATION'/)
 610  FORMAT(5X,20A5/)
      WRITE(6,620)(NCS(I),I=1,NCUT)
 620  FORMAT(6X,' CUTPOINTS ',5I6)
      WRITE(6,621)
 621  FORMAT(//)
      WRITE(6,630)
 630  FORMAT(6X,' CROSS CLASSIFICATION MATRIX'
     1 /6X,' ---------------------------')
      WRITE(6,640)
 640  FORMAT(24X,' CLUSTERS')
      WRITE(6,650)
 650  FORMAT(15X,'U',4X,'1',4X,'2',4X,'3',4X,'4',4X,'5'
     1 ,4X,'6',4X,'7',4X,'8',4X,'9')
      WRITE(6,660)
 660  FORMAT(15X,'-',4X,'-',4X,'-',4X,'-',4X,'-',4X,'-'
     1 ,4X,'-',4X,'-',4X,'-',4X,'-')
      DO 515 I=1,NCUT
 515  WRITE(6,520)I,(ICROS(I,J),J=0,NCLUS)
 520  FORMAT(3X,I3,':',4X,20I5)
      DO 528 I=1,NCUT
 528  IRMAX(I)=-1000
      DO 529 I=1,NCLUS
 529  ICMAX(I)=-1000
      DO 530 I=1,NCUT
      DO 530 J=1,NCLUS
      ICTOT(I)=ICTOT(I)+ICROS(I,J)
      IRTOT(J)=IRTOT(J)+ICROS(I,J)
      IF(ICROS(I,J).LT.ICMAX(J))GO TO 531
      ICMAX(J)=ICROS(I,J)
      IVMAX(J)=I
 531  IF(ICROS(I,J).LT.IRMAX(I))GO TO 530
      IRMAX(I)=ICROS(I,J)
      IWMAX(I)=J
 530  CONTINUE
      N=0
      MAXRJ=-1000
      DO 541 I=1,NCLUS
 541  IF(MAXRJ .LT. IRTOT(I))MAXRJ=IRTOT(I)
      MAXCJ=-1000
      DO 540 I=1,NCUT
      N=N+ICTOT(I)
 540  IF(MAXCJ .LT. ICTOT(I))MAXCJ=ICTOT(I)
      S1=0.0
      DO 550 J=1,NCLUS
      I=IVMAX(J)
 550  S1=S1+ICROS(I,J)
      S2=0.0
      DO 551 I=1,NCUT
      J=IWMAX(I)
 551  S2=S2+ICROS(I,J)
```

```
C       WRITE(6,*)S1,MAXCJ,N
        IF(N.EQ.MAXCJ)GO TO 760
        SLAM=(S1-FLOAT(MAXCJ))/(FLOAT(N)-FLOAT(MAXCJ))
 760    CONTINUE
        IF(N.EQ.MAXRJ)GO TO 761
        SLM2=(S2-FLOAT(MAXRJ))/(FLOAT(N)-FLOAT(MAXRJ))
 761    CONTINUE
        IF(2*N .EQ. MAXCJ+MAXRJ)GO TO 762
        SYMM=(S1+S2-MAXCJ-MAXRJ)/(2*N-MAXCJ-MAXRJ)
 762    CONTINUE
        WRITE(6,560)SLAM
        WRITE(6,561)SLM2
        WRITE(6,562)SYMM
 561    FORMAT(/'     LAMBDA (PREDICT CLUSTERS  ) ',F10.3)
 562    FORMAT(/'     LAMBDA (SYMMETRIC         ) ',F10.3)
        SPCT=FLOAT(N)/FLOAT(NR)
        WRITE(6,544)SPCT
 560    FORMAT(/'     LAMBDA (PREDICT POPULATION) ',F10.3)
 544    FORMAT(/'     PERCENT COVERED             ',F10.3/)
        IF(METH.EQ.1)M=F+2.1
        IF(METH.EQ.2)M=SNU2+.1
        WRITE(6,570)M
 570    FORMAT(' NUMBER OF VARIABLES PER CASE ',I5/)
        WRITE(6,580)CL,CU
 580    FORMAT(' INCLUSION VALUE CIN ',F10.3/
     1  ' EXCLUSION VALUE CEX ',F10.3)
        WRITE(6,590)SCL,SCU
 590    FORMAT(' PROB VALUE CIN  ',4X,F10.3/
     1  ' PROB VALUE CEX  ',4X,F10.3)
        SPCT=FLOAT(N)/FLOAT(NR)
        OPEN(UNIT=2,FILE='CLUST.RES',ACCESS='APPEND')
        WRITE(2,714)TITLE(5),CL,CU,SCL,SCU,M,NCLUS,NCUT,SLAM,SLM2,SYMM
     1  ,SPCT
 714    FORMAT(1X,A2,4F6.3,3I3,4F6.3)
        DO 724 J=0,NCLUS
        M=0
        DO 725 I=1,NTOT
        IF(ICLUS(I).EQ.J)M=M+1
        IF(ICLUS(I).EQ.J)IOUT(M)=I
 725    CONTINUE
        IF(J.EQ.0)WRITE(6,728)
 728    FORMAT(/'          UNCLASSIFIED ENTITIES '/)
        IF(J .GT. 0)WRITE(6,729)J
 729    FORMAT(/'          MEMBERSHIP IN CLUSTER ',I3/)
        WRITE(6,727)(IOUT(K),K=1,M)
 727    FORMAT(10X,25I4)
 724    CONTINUE
        T='*****'
        WRITE(6,598)T,T,T,T,T,T,T,T,T,T,T,T,T,T,T,T
 598    FORMAT(//16A5)
        IF(IWRIT.EQ.1)CALL WRITE(IDNUM,ICLUS)
        END
        SUBROUTINE WRITE(IDNUM,ICLUS)
        DIMENSION FORM(16)
     1  ,X(10),ICLUS(1),NC(5)
        DOUBLE PRECISION DATA1(20),ISET
        ISET=DATA1(IDNUM)
       DATA DATA1/'MIX1.DAT','MIX2.DAT','MIX3.DAT','MIX4.DAT','MIX5.DAT'
     1  ,'MIX6.DAT','MIX7.DAT','MIX8.DAT','MIX9.DAT','MIX10.DAT',
     1'MIX11.DAT','MIX12.DAT','MIX13.DAT','MIX14.DAT','MIX15.DAT',
```

```
     1     'MIX16.DAT','MIX17.DAT','MIX18.DAT','MIX19.DAT',
     1 'MIX20.DAT'/
        OPEN(UNIT=2,FILE=ISET)
        OPEN(UNIT=3,FILE='BLASH.DAT')
        READ(2,100)N,M
 100    FORMAT(2G)
        READ(2,110)NCUT
 110    FORMAT(1G)
        READ(2,120)(NC(I1),I1=1,NCUT)
        READ(2,130)FORM
 120    FORMAT(5G)
 130    FORMAT(16A5)
        READ(2,130)TITLE
        DO 10 I=1,6000
        READ(2,150,END=200)I1,I2,I3,X
 150    FORMAT(3I3,10F7.3)
        IF(I3.EQ.0)GO TO 160
        ICNT=ICNT+1
 160    WRITE(3,170)ICLUS(ICNT),I1,I2,I3,X
 170    FORMAT(4I3,10F7.3)
 10     CONTINUE
 200    RETURN
        END
        SUBROUTINE BUILD(STPE)
        DIMENSION NPR(149),SUM(149),PROF1(149),N(149)
        COMMON IDEL(149),PROF(149),R(149,149),ICLUS(149),
     1  NR,NC,CL,CU,NDEL,NPROF
        INTEGER PROF,PROF1
C       HOUSEKEEPING
        DO 5 I=1,NR
        NPR(I)=0
        PROF(I)=0
 5      PROF1(I)=0
        DO 10 J=1,NC
        DO 10 I=1,NR
 10     IF(R(I,J).GE.CL)NPR(J)=NPR(J)+1
C       FIND MAX OF NPR TO DETERMINE PIVOT
D       WRITE(6,248)NPR
 248    FORMAT(' NPR ',11I3)
        MAX1=-1000
        IMAX=0
        SMAX=-1.0E+30
        DO 20 J=1,NR
        IF(NPR(J).NE.0)IMAX=1
        IF(NPR(J).LT.MAX1)GO TO 20
        MAX1=NPR(J)
        MAX=J
 20     CONTINUE
C       FIND # OF ELEMENTS IN PIVOT
        IF(IMAX.EQ.0)STPE=1.0
        IF(STPE.EQ.1.0)RETURN
        J=0
        DO 30 I=1,NR
        IF(R(I,MAX).LT.CL)GO TO 30
        J=J+1
        NPROF=J
        PROF(J)=I
 30     CONTINUE
D       WRITE(6,78)MAX
D78     FORMAT( ' MAX ',I5)
```

CONTINUED

1 OF 2

```
D       WRITE(6,79)(PROF(I),I=1,NPROF)
D79     FORMAT(' FIRST LIST ',11I5)
C       FIND SECOND MEMBER OF NUCLEUS
        DO 35 I=1,NR
        SUM(I)=0.0
 35     N(I)=0.0
        DO 40 I=1,NR
        DO 40 J=1,NPROF
        IF(R(I,PROF(J)).LT.-2.0)GO TO 40
        N(I)=N(I)+1
        SUM(I)=SUM(I)+R(I,PROF(J))
 40     CONTINUE
        SUM(MAX)=-1.0E+30
C       FIND MAX
        SMAX1=-1.0E+30
        DO 50 J=1,NR
        IF(N(J).EQ.0)GO TO 50
        SUM(J)=SUM(J)/N(J)
        IF(SUM(J).GE.SMAX1)MAX1=J
        IF(SUM(J).GE.SMAX1)SMAX1=SUM(J)
 50     CONTINUE
C       FIRST TWO PROFILES IN PIVOT LIST ARE MAX AND MAX1
C       FIND THIRD MEMBER
D       WRITE(6,93)(SUM(J),J=1,NR)
D 93    FORMAT(' AVE CORR ',11F7.2)
        J=0
        DO 60 I=1,NC
        IF(R(I,MAX1).LT.CL)GO TO 60
        J=J+1
        NPROF1=J
        PROF1(J)=I
 60     C[illegible]
C       C[illegible] OUT PROFILES IN PROF1 ALREADY IN PROF
        D[illegible] I=[illegible],NPROF
        [illegible],0 J=[illegible],NPROF1
 70     [illegible](PROF([illegible]).EQ.PROF1(J))PROF1(J)=0
C       [illegible]ONSOLIDATE PROFILES
        J=0
        DO 80 I=1,NPROF1
        IF(PROF1(I).EQ.0)GO TO 80
        J=J+1
        PROF(NPROF+J)=PROF1(I)
 80     CONTINUE
        NPROF=NPROF+J
D       WRITE(6,91)(PROF(I),I=1,NPROF)
D91     FORMAT(' NEXT LIST ',11I5)
        DO 85 I=1,NR
        N(I)=0
 85     SUM(I)=0.0
C       FIND THIRD MEMBER OF NUCLEUS
        DO 90 I=1,NR
        DO 90 J=1,NPROF
        IF(R(I,PROF(J)).LT.-2.0)GO TO 90
        N(I)=N(I)+1
        SUM(I)=SUM(I)+R(I,PROF(J))
 90     CONTINUE
C       FIND MAX VALUE OF SUM(I)
        SUM(MAX)=-1.0E+30
        SUM(MAX1)=-1.0E+30
        SMAX2=-1.0+E30
```

```
        DO 100 J=1,NR
        IF(N(J).EQ.0)GO TO 100
        SUM(J)=SUM(J)/N(J)
        IF(SUM(J).GE.SMAX2)MAX2=J
        IF(SUM(J).GE.SMAX2)SMAX2=SUM(J)
 100    CONTINUE
D       WRITE(6,93)(SUM(J),J=1,NR)
C       FIRST CLUSTER CONSISTS OF PROFLES MAX,MAX1,MAX2
        PROF(1)=MAX
        PROF(2)=MAX1
        PROF(3)=MAX2
D       WRITE(6,92)PROF(1),PROF(2),PROF(3)
D92     FORMAT(' NUCLEUS ',3I5)
        NPROF=3
        DO 105 I=1,NR
        N(I)=0
 105    SUM(I)=0.0
 108    DO 110 I=1,NR
        DO 110 J=1,NPROF
        IF(R(I,PROF(J)).LT.-2.0)GO TO 110
        IF(I.EQ.PROF(J))GO TO 110
        SUM(I)=SUM(I)+R(I,PROF(J))
        N(I)=N(I)+1
 110    CONTINUE
        DO 115 I=1,NPROF
 115    SUM(PROF(I))=-1.0E+30
C       FIND AVE AND MAX
        DO 120 I=1,NR
        IF(N(I).EQ.0)GO TO 120
        SUM(I)=SUM(I)/N(I)
 120    CONTINUE
D       WRITE(6,94)(SUM(I),I=1,NR)
D94     FORMAT(' AVERAGE CORR ',11F7.2)
C       FIND MAX AVERAGE CORRELATION
        SMAX=-1.0E+30
        DO 130 I=1,NR
        IF(SUM(I).GE.SMAX)MAX=I
 130    IF(SUM(I).GE.SMAX)SMAX=SUM(I)
        IF(SUM(MAX).LT.CL)GO TO 200
        NPROF=NPROF+1
        PROF(NPROF)=MAX
D       WRITE(6,*)(PROF(INN),INN=1,NPROF)
D       WRITE(6,*)SUM(MAX),MAX
        DO 140 I=1,NR
        SUM(I)=0.0
 140    N(I)=0
        GO TO 108
 200    CONTINUE
        NDEL=0
        DO 205 I=1,NR
        N(I)=0
        SUM(I)=0.0
 205    IDEL(I)=0
D       WRITE(6,96)
D96     FORMAT(' DELETE CASES BELOW CU ')
        DO 210 I=1,NR
        DO 210 J=1,NPROF
        IF(R(I,PROF(J)).LT.-2.0)GO TO 210
        IF(I.EQ.PROF(J))GO TO 210
        SUM(I)=SUM(I)+R(I,PROF(J))
```

```
          N(I)=N(I)+1
 210      CONTINUE
          DO 220 I=1,NR
          IF(N(I).EQ.0)GO TO 220
          SUM(I)=SUM(I)/N(I)
          IF(SUM(I).LE.CU)GO TO 220
          IF(SUM(I).GE.CL)GO TO 220
          NDEL=NDEL+1
          IDEL(NDEL)=I
 220      CONTINUE
D         WRITE(6,94)(SUM(I),I=1,NR)
D         WRITE(6,77)(IDEL(I),I=1,NDEL)
 77       FORMAT(' DELETE ',11I5)
          IF(NPROF.LE.3)STPE=1.0
          RETURN
          END
```

```
*************************************************************
*                                                           *
*                       CORR.FOR                            *
*                                                           *
*     CORR.FOR  IS A SUBROUTINE FOR LORR.FOR.  IT           *
*     PRODUCES A CORRELATION MATRIX (PRODUCT-MOMENT         *
*     CORRELATIONS) FOR USE IN LORR'S NON-                  *
*     HIERARCHICAL CLUSTERING TECHNIQUE.                    *
*                                                           *
*************************************************************

          DIMENSION X(23,149),R(11250),S(149),XM(149)
          DIMENSION NC(5),FORM(20)
          DIMENSION TITLE(20)
          DOUBLE PRECISION DATAS
          READ(5,999)DATAS
 999      FORMAT(A10)
          OPEN(UNIT=11,FILE=DATAS)
          READ(11,*)N,M
          IX=23
          READ(11,457)NCUT
          READ(11,460)(NC(I1),I1=1,NCUT)
 457      FORMAT(1G)
 460      FORMAT(5G)
          READ(11,463)FORM
          READ(11,463)TITLE
          DO 111 J=1,M
 111      READ(11,FORM)IDNUM,(X(I,J),I=1,N)
          CALL BECORI(X,N,M,IX,XM,S,R,IER)
D         WRITE(6,78)IER
 78       FORMAT('  IERR  ',I10)
          ITOT=M*(M+1)/2
          K1=1
          IT=0
          M1=1
D         WRITE(6,*)N,M,NCUT,NC
D         WRITE(6,463)TITLE
D         WRITE(6,463)FORM
 463      FORMAT(20A5)
          WRITE(12)TITLE
          METH=1
          WRITE(12)METH
          WRITE(12)N,M,IDNUM
          WRITE(12)NCUT
          WRITE(12)(NC(IO),IO=1,NCUT)
 456      FORMAT(20A5)
          DO 200 J=1,M
          WRITE(12)(R(I),I=M1,K1)
          WRITE(14,1111)(R(I),I=M1,K1)
 1111     FORMAT(F5.2)
D         WRITE(6,467)(R(I),I=M1,K1)
 467      FORMAT(16F5.2)
          IT=IT+1
          M1=M1+IT
          K1=M1+IT
```

```
200     CONTINUE
        END
```

```
*****************************************************
*                                                   *
*                    INTRA.FOR                      *
*                                                   *
*     SUBROUTINE FOR LORR.FOR.  COMPUTES INTRACLASS *
*     CORRELATION MATRIX FOR USE IN LORR'S NON-     *
*     HIERARCHICAL CLUSTERING TECHNIQUE.            *
*                                                   *
*                                                   *
*****************************************************

        DIMENSION X(23,149),R(11250)
        DIMENSION Y(60),NTR(30),TM(30),WTV(30),S(3),NDF(3)
        DIMENSION NC(5),FORM(20)
        DIMENSION TITLE(20)
        DOUBLE PRECISION DATAS
        READ(5,999)DATAS
999     FORMAT(A10)
        OPEN(UNIT=11,FILE=DATAS)
        READ(11,*)N,M
        IX=23
        READ(11,457)NCUT
        READ(11,460)(NC(I1),I1=1,NCUT)
457     FORMAT(1G)
460     FORMAT(5G)
        READ(11,463)FORM
        READ(11,463)TITLE
        DO 111 J=1,M
111     READ(11,FORM)IDNUM,(X(I,J),I=1,N)
        DO 5 I=1,NCUT
5       ITOT=ITOT+NC(I)
        DO 6 I=1,60
6       NTR(I)=2
        DO 10 I=1,ITOT
        DO 10 J=1,I
        I111=0
        DO 15 K=1,N
        I111=I111+1
        Y(I111)=X(K,J)
        I111=I111+1
        Y(I111)=X(K,I)
15      CONTINUE
        CALL ACRDAN(Y,N,NTR,TM,WTV,S,GM,NDF,IER)
        ICNT=ICNT+1
C       TREAT SSQ
        S(1)=S(1)/NDF(1)
C       BETWEEN SSQ
        S(2)=S(2)/NDF(2)
10      R(ICNT)=(S(1)-S(2))/(S(1)+S(2))
D       WRITE(6,78)IER
78      FORMAT('  IERR  ',I10)
        ITOT=M*(M+1)/2
        K1=1
        IT=0
        M1=1
```

```
D        WRITE(6,*)N,M,NCUT,NC
D        WRITE(6,463)TITLE
D        WRITE(6,463)FORM
 463     FORMAT(20A5)
         WRITE(12)TITLE
         METH=2
         WRITE(12)METH
         WRITE(12)N,M,IDNUM
         WRITE(12)NCUT
         WRITE(12)(NC(IO),IO=1,NCUT)
 456     FORMAT(20A5)
         DO 200 J=1,M
         WRITE(12)(R(I),I=M1,K1)
         WRITE(15,1111)(R(I),I=M1,K1)
 1111    FORMAT(F5.2)
D        WRITE(6,467)(R(I),I=M1,K1)
 467     FORMAT(16F5.2)
         IT=IT+1
         M1=M1+IT
         K1=M1+IT
 200     CONTINUE
         END
```

Supplement I Manuscript "Inverted factor analysis: An evaluation using benchmark data sets"

Craig Edelbrock and Michael L. Reed

Inverted Factor Analysis: An Evaluation Using Benchmark Data Sets

Craig Edelbrock

Michael Reed

University of Pittsburgh School of Medicine

This research was supported by Grant #1-0260-8-PA-IJ from the Methodology Development Program, National Institute of Law Enforcement and Criminal Justice, to the first author. Correspondence may be addressed to Craig Edelbrock, Western Psychiatric Institute & Clinic, 3811 O'Hara Street, Pittsburgh, PA 15213.

Abstract

Inverted factor analysis was evaluated on 20 previously studied multivariate mixtures. Two methods of determining number of factors and two rotational methods--orthogonal varimax and oblique direct quartimin--were compared. Objects were assigned to groups on the basis of highest absolute factor loadings, with the mimimum loading required for assignment systematically varied. Rotational methods did not differ significantly in either accuracy or coverage of the resulting classifications. Paradoxically, setting the number of factors equal to the number of underlying populations resulted in <u>less</u> accurate solutions than determining the number of factors empirically by Cattell's scree test. More importantly, the inverted factoring technique was found to be as accurate as the best hierarchial clustering algorithms previously tested on these mixtures. Thus, despite the implausibility of the factor analytic model for generating typologies--and numerous other problems and criticisms--inverted factor analysis appears to be a useful taxonomic tool.

Inverted Factor Analysis: An Evaluation using Benchmark Data Sets

Inverted factor analysis, also known as Q-factor analysis, inverse factor analysis, and profile factor analysis, is one of the oldest and most widely used procedures for constructing typologies in the behavioral sciences. The basic rationale of this procedure has not changed since Stephenson introduced the "inverted factor technique" in 1936. In fact, Stephenson's original articles (1936a,b) still provide a lucid introduction to the method. In the past 30 years, inverted factor analysis has been used in numerous studies to identify subtypes of individuals, particularly in the areas of psychiatry and deviant behavior (Butler & Adams, 1966; Collins, Burger & Taylor, 1976; Fleiss, Lawlor, Platman & Fieve, 1971; Guertin, 1952, Katz & Cole, 1963; Monro, 1955; Overall, Hollister, Johnson & Pennington, 1966; Raskin & Crook, 1963). The inverted factor technique has also been discussed in several methodological treatises (Baggaley, 1964; Broverman, 1961; Cattell, 1952; Morf, Miller and Syrotuik, 1976; Overall and Klett, 1972; Ross, 1963; Ryder, 1966; Stephenson, 1953).

Despite its historical precedence and diverse applications, inverted factor analysis has been strongly criticized as a method of generating typologies (Baggaley, 1964; Fleiss et al, 1971; Fleiss, 1972; Fleiss & Zubin, 1969; Jones, 1968; Lorr, 1966). A standard criticism has involved the use of the product-moment correlation index similarity between individuals. The correlation coefficient indexes similarity only in profile shape, not elevation or scatter. Moreover, a correlation of 1.00 does not necessarily indicate that two profiles have identical shape, but only that they are linear functions of one another (see Edelbrock & McLaughlin, 1980; and Fleiss & Zubin, 1969 for more detailed discussions). This criticism is not unique to inverted factor analysis, however, in that a variety of clustering methods can employ the

correlation coefficient as the measure of profile similarity (e.g., Carlson, 1972; Edelbrock, 1979; Lorr, Bishop & McNair, 1965; Lorr & Radhakrishnan, 1967).

Several other criticisms have been raised that are more pertinent to the inverted factoring technique. Fleiss and Zubin (1969), for example, have questioned the appropriateness of the linear model underlying factor analysis to the task of generating typologies of individuals. In particular, they have asked whether it makes any sense to say that an individual represents "X" amount of one type plus "Y" amount of another type, and so on. Lorr (1966) has further questioned the rationale of rotating Q-factors. Even if unrotated factor loadings represent similarity to underlying "types", what is the meaning of transforming such loadings so as to better approximate simple structure?

Fleiss and Zubin have also objected that the number of types one may identify is limited by the number of variables in the analysis. The maximum number of factors that can be extracted from a correlation matrix is equal to the rank of the matrix. For a matrix of Q-correlations, rank is at most p - 1, where p equals the number of variables. Fleiss and Zubin therefore reasoned that the maximum number of types one can identify is equal to the number of variables minus 1. This is obviously a problem when one has few variables with which to work, but seeks to identify several types of individuals.

Achenbach and Edelbrock (1981) have noted an additional problem involving procedures of factor extraction. Since the first factor typically extracts the most variance from the correlation matrix, it will encompass more individuals having high loadings than subsequent factors. This bias towards constructing one large group followed by successively smaller and smaller groups is rarely justified in taxonomic research. Clearly, the relative size of the groups should be determined by the data, not the taxometric procedure. Other methodological problems and issues include: (a) translating factor scores into discrete groups of individuals, (b) determining the appropriate number of factors, and (c) selecting a rotational procedure. The latter two problems also arise in regular R-factor analysis and have been discussed in detail elsewhere (cf. Mulaik, 1972; Harman, 1976).

Given these problems and criticisms, one would expect inverted factor analysis to have been laid to rest long ago--but this is not the case. More than 40 years after its inception, the technique is still in use. Furthermore, it has generated heuristically valuable and predictive typologies. A particularly good example is the nosology of depression constructed by Overall et al (1966). Using an inverted factoring procedure (Overall & Porterfield, 1963), three subtypes of depressed patients were identified, based on scores on the Brief Psychiatric Rating Scale. In a subsequent double-blind comparison, the three subtypes (labelled Anxious, Hostile, and Retarded) were found to differ markedly in terms of response to anti-depressant drugs. The value of inverted factor analysis in taxonomic research has been corroborated by several other recent studies (Collins et al, 1976; Evenson, Altman, Sletten & Knowles, 1973; Kunce, Ryan & Eckelman, 1976; Meyer & Kline, 1977; Raskin & Crook, 1976).

A Reconsideration

There are several compelling reasons for reconsidering inverted factor analysis as a taxonomic tool. For one, fruitful applications of the technique would appear to mitigate any methodological criticisms. Second, some points of criticism are patently wrong. For example, although the number of factors may be limited to $p - 1$, the number of types is not limited to the number of factors. In practice, inverted factor analysis may yield bipolar factors

comprised of both positive and negative loadings. Such bipolar factors have been taken to represent two underlying types manifesting opposite patterns of scores. Carlson et al (1976), for instance, obtained only four factors, but because each was bipolar, eight subtypes were identified. Third, some criticisms are based on dogma, not empirical facts. Some recent studies suggest that long-established psychometric dogma is in desperate need of revision. For example, despite the so-called "superiority" of distance measures for indexing profile similarity (e.g., Eades, 1965; Fleiss & Zubin, 1969: p. 239), recent Monte Carlo studies of hierarchical clustering methods have shown that correlation yields substantially better recovery of underlying mixture populations than Euclidean distance (Edelbrock, 1979; Edelbrock & McLaughlin, 1980). Finally, there have been very few attempts to test the inverted factoring technique empirically against other methods. One exception is the recent study by Blashfield and Morey (1980). Using Monte Carlo procedures, data sets designed to mimic MMPI psychotic, neurotic, and personality disorder patterns were generated then analyzed by inverted factor analysis, Lorr's non-hierarchical clumping procedure (Lorr et al, 1965), a hierarchical clustering algorithm called average linkage, and Ward's (1963) minimum variance technique. Blashfield and Morey concluded that the average linkage method yielded the best clustering solutions. For some data sets, however, the inverted factoring technique resulted in substantially fewer misclassifications than the other three methods.

Purpose of this Study

The purpose of this study was to evaluate inverted factor analysis on a standard set of multivariate mixtures. This research builds on Blashfield and Morey's recent study in the following ways: (a) a broad range of multivariate mixtures differing in number of variables, number of underlying populations, difficulty of solution, etc., were analyzed, (b) two methods of determining number of factors were tested, (c) two rotational procedures--one orthogonal the other oblique--were compared, and (d) the effects of varying the minimum loading required for classification were systematically evaluated. In addition, comparisons between the inverted factoring technique and several hierarchical clustering algorithms were made.

Methods

Data Sets

It has been argued previously (Edelbrock, 1979; Edelbrock & McLaughlin, 1980) that evaluations of taxometric methods should include test on "benchmark" data sets--that is, data sets have been well-characterized, are available to other investigators, and have been used in previous mixture model studies. Such benchmark data sets provide a common standard against which to compare clustering and classification methods and thus increase the generalizability of mixture model tests. With this in mind, 20 multivariate normal mixtures generated by Blashfield (1976) were selected for this study. These mixtures mimic real data in many ways, including (a) representative range of number of variables and populations, (b) quasi-normal distribution parameters, (c) addition of "measurement" error to scores, and (d) varying strength and complexity of the covaviance structure of the underlying populations. These data sets have also been used in previous tests of hierarchical clustering algorithms (Blashfield, 1976; Edelbrock, 1979; Edelbrock & McLaughlin, 1980), so direct comparisons across studies are possible.

Procedures

Data were double-centered according to the rationale and procedure given by Overall and Klett (1972; pp. 203-204). Variables were standardized (mean = 0, sd = 1) and scores were then standardized equivalently across objects.

Each of the 20 (object X variable) data sets was then inverted (i.e., to represent a variable X object matrix) and subjected to principal-components factor analysis using the BMDP4M - program. It is important to note that double-centering the data results in bipolarity of the unrotated factors. However, it does not necessarily result in bipolarity in the rotated factors, which were used here.

Two procedures were used to determine the number of factors. First, for each mixture, the number of factors was set to equal the number of underlying populations. Since the rotated factors were not bipolar, each factor comprised only one group of objects having high loadings in the same direction. Thus, determing the number of factors in this way is tantamount to setting the number of groups (j) equal to the number of underlying populations (k). These 20 analyses are subsequently designated by the notation $j = k$.

Second, the number of factors was determined by examining eigen values. For these data sets, the commonly used "eigen value greater than 1" rule resulted in considerable over-factoring. A few factors having large eigen values were obtained followed by several having eigen values slightly greater than 1.00. This problem was also encountered by Blashfield and Morey (1980). Following their procedure, Cattell's (1966) scree test was used to determine number of factors. In this study, both investigators examined the eigen value plot for each mixture and independently selected the number of factors. Although we agreed for all 20 mixtures, the number of factors indicated by the scree test did not always equal the number of underlying populations. For eight mixtures, the number of factors equalled one more than the number of underlying populations (i.e., $k + 1$). These 20 analyses are subsequently designated by the notation $j \neq k$ (i.e., the number of groups did not necessarily equal the number of populations).

One issue in factor analysis is whether to construct orthogonal (uncorrelated) or oblique (correlated) factors. This is an important consideration when deriving typologies because rotational procedures substantially affect final factor loadings, which are the basis for constructing groups. Most previous applications of inverted factor analysis (e.g., Blashfield & Morey, 1980; Collins et al, 1976; Fleiss et al, 1971; Katz & Cole, 1965) involved the varimax rotation--an orthogonal procedure. In this study, both varimax (orthogonal) ind direct quartimin (oblique) rotations were compared. This yields four analyses of 20 mixtures each: $j = k$ and $j \neq k$ with either varimax or direct quartimin rotation.

A crucial issue that arises in inverted factor analysis involves translating factor loadings into discrete groups of objects or individuals. A common procedure has been to assign individuals to groups on the basis of highest factor loadings (in terms of absolute value). Some investigators have specified a minimum loading required for classification. Fleiss et al (1971), for example, selected a minmum loading of .40. Individuals whose highest loadings were less than .40 were left unclassified. In their Monte Carlo study, Blashfield and Morey (1980) selected a minimum loading of .60, with the additional criterion that an object could not have a loading of .60 or higher on any other factor. These rather stringent criteria reduce coverage substantially, but result in more distinct and homogeneous groups.

In this study, objects were assigned to groups on the basis of their highest loadings. This is a simple procedure for constructing groups, but the coverage of the resulting classification can be manipulated by simply changing the minimum loading required for assignment. A low cutoff point results in the classification of a high proportion of objects into relatively heterogenous groups, whereas a high cutoff point results in the classification of a low proportion of objects into more distinct, non-overlapping groups. This

assignment procedure therefore makes it possible to evaluate classifications at several levels of coverage.

Calculating Accuracy

The accuracy of the inverted factor solutions was defined as the agreement between the obtained groups and the underlying populations in the mixtures. A wide variety of statistics have been used to measure accuracy in mixture model studies; and there is little consensus regarding the "best" accuracy measure. Kappa (Cohen, 1960) and Rand's statistic (Rand, 1971) have been used in many studies (e.g., Blashfield, 1976; Edelbrock, 1979; Edelbrock & McLaughlin, 1980; Kuiper & Fisher, 1975; Milligan & Isaac, 1980; Mojena, 1977; Rand, 1971). Both of these measures have drawbacks. Kappa has the advantage of correcting for chance level of agreement in a cross-classification, but it is appropriate only for square matrices (i.e., $j = k$). Rand's statistic does not require that $j = k$, but the scale is not uniform from matrix to matrix. That is, the lower bound of Rand's statistic is not zero but is determined by the marginal distributions of the cross-classification.

One way to overcome the idiosyncracies inherent in individual measures is to use multiple criteria for evaluating accuracy. Six measures, including kappa, Rand's statistic, asymmetric lambda, tau, Kramer's v, and the contingency coefficient were used in this study. We chose to report our main findings in terms of asymmetric lambda for several reasons. This statistic is appropriate for nominal level cross-classifications, has a range of zero to 1.00, and can be used with either square ($j = k$) or rectangular ($j \neq k$) matrices. The "asymmetrical" aspect of this statistic also seems well suited to the task of measuring accuracy. The term "asymmetrical" refers to the fact that lambda indexes the degree to which one classification predicts another, and not vice versa. In mixture model studies, the underlying populations comprise a fixed or dependent classification, predicted by empirically derived groups that are free to vary.

Although we report our main findings in terms of asymmetric lambda, we also report summary statistics in terms of kappa and Rand's statistic. This permits direct comparisons with previous studies. Finally, it is worth noting that our conclusions regarding the relative accuracy of various methods were identical for all six measures we explored. This is not surprising, since such measures are all founded on the same information extracted from the cross-classification matrix (cf. Hubert & Levin, 1976). Furthermore, in these analysis, the six measures of accuracy correlated >.95 with one another.

Statistical Analyses

For each of the 80 inverted factor solutions, objects were classified according to their highest loadings. Accuracy was then calculated at seven levels of coverage dictated by the following minimum loadings: .0, .4, .5, .6, .7, .8, and .9. These minimum loadings between were selected because: (a) all objects had highest loadings greater than .0, thus a cutoff point of .0 yields 100% coverage, (b) very few objects had highest loadings between .0, and .4 so accuracy and coverage varied little in this interval, and (c) there were too few loadings above .9 to calculate accuracy.

Accuracy and coverage values were analyzed in separate 2 x 2 x 7 analyses of variance representing: number of factors (j = k vs. j ≠ k) rotational methods (varimax vs. direct quartimin), and minimum loading (.0 to .9), respectively.

Results

Main results are portrayed graphically in Figures 1 and 2. These figures

- - - - - - - - - - - - - - - - - -

Insert Figures 1 and 2 here

- - - - - - - - - - - - - - - - - -

show the relations between the minimum loading required for classification and both accuracy (left axis) and coverage (right axis). Figure 1 depicts accuracy and coverage functions for the j = k solutions, whereas Figure 2 despicts results for the j ≠ k solutions. Overall, accuracy and coverage were significantly related to the minimum loading (p <.001), but in opposite ways. Raising the minimum loading uniformly increased accuracy, but decreased coverage to a greater and greater extent. No significant differences (F <1.00) were detected between varimax and direct quartimin rotations for either j = k or j ≠ k solutions. Varimax solutions resulted in consistently higher accuracy and coverage, however.

Paradoxically, j ≠ k solutions resulted in significantly <u>higher</u> accuracy and coverage than j = k solutions (p <.01). This was the case for both rotational methods. Figure 3 portrays accuracy differences between j = k and j ≠ k solutions in a manner that equates them for coverage. Accuracy is shown as a function of coverage, rather than as a function of the minimum loadings as in Figures 1 and 2. At all levels of coverage, j ≠ k solutions resulted in significantly higher accuracy and j = k solutions. Examination

- - - - - - - - - - -

Insert Figure 3 here

- - - - - - - - - - -

of the eight mixtures where j ≠ k confirmed that constraining the number of factors to equal the number of underlying groups substantially reduced accuracy. For these mixtures, higher accuracy was achieved when the number of groups was determined empirically by Cattell's scree test.

Comparisons with Other Methods

In a previous study (Edelbrock & McLaughlin, 1980), 18 hierarchical clustering algorithms were tested on the 20 benchmark mixtures. The algorithms included single, complete, average, and centroid linkage using either Euclidean distance, correlation, or the one-way or two-way intraclass correlation as the similarity measure; Ward's minimum variance technique; and a random algorithm used to establish a baseline control for evaluating methods. Two problems arise when making comparisons between inverted factor analysis and these hierarchical methods. First, the accuracy of each hierarchical method was calculated for $j = k$. That is, the number of clusters alwiys equaled the number of underlying populations. To make direct comparisons, it is necessary to select inverse factor solutions were $j = k$. This is unfortunate because $j = k$ solutions were significantly less accurate than $j \neq k$ solutions. Comparisons are therefore based on a conservative estimate of the accuracy of the inverted factoring technique.

The second problem involves selecting the level of coverage at which to make comparisons. Whereas both inverted factor analysis and the hierarchical methods can yield classifications varying in coverage, this occurs in quite different ways. For inverted factor analysis, coverage depends upon the minimum loading required for assignment. For the hierarchical methods, coverage depends on the selection of the best j clusters at various levels in the hierarchical tree. This difference appears to represent a bias in favor of the hierarchical methods. For each mixture, the accuracy of the inverted factor solution is based on the <u>same</u> set of factors--only the minimum loading is varied. The accuracy of each hierarchical solution, on the other hand, is based on <u>different</u> sets of clusters, selected so as to maximize accuracy at each level in the hierarchical tree. This bias is evidenced by the fact that the accuracy of even the <u>random</u> hierarchical algorithm increases as coverage delcines (see Edelbrock & McLaughlin, 1980: p. 310).

To make comparisons between methods, accuracies of the $j = k$ varimax solutions were calculated at 100% coverage. Focusing on 100% coverage eliminates the biases that can arise at lower levels of coverage. Furthermore, inverted factor analysis and the best hierarchical methods show uniform increases in accuracy as coverage declines. Thus, differences at 100% coverage are likely to be representative of differences at lower levels of coverage.

The mean kappa value for the varimax solutions equaled .65, which compares quite favorably with accuracies previously reported by Edelbrock & McLaughlin (1980: p. 310). Specifically, the inverted factoring technique was substantially more accurate than 10 of the 18 hierarchical algorithms: single and complete linkage using any of the four similarity measures, average and centroid linkage using Euclidean distance, and the random algorithm.

The $j = k$ varimax solutions were also compared with the most accurate hierarchical algorithm--average linkage using the one-way intraclass correlation. Mean values for kappa, Rand's statistic, and asymmetric lambda, as well as paired <u>t</u>-test results, are shown in Table 1. According to all three measures, the average linkage algorithm was slightly more accurate than inverted factor analysis, but not significantly so.

- - - - - - - - - -

Insert Table 1 here

- - - - - - - - - -

As a final test, Rand's statistic was calculated for varimax $j \neq k$ solutions: which represent the highest accuracy attained by inverted factor analysis. Edelbrock and McLaughlin previously used Rand's statistic to evaluate the "best possible" clustering solutions attained by the 18 hierarchical methods they examined. Direct comparisons between methods are therefore possible. The average Rand value for the inverted factoring technique was .862. This is higher than 11 of the 18 hierarchical methods, and not significantly different than the most accurate hierarchical algorithm (see Edelbrock & McLaughlin, 1980: p. 311).

Discussion

Inverted factor analysis is one of the most widely used and widely criticized procedures for constructing typologies in the behavioral sciences. Unfortunately, some critics of the method have simply argued: "It shouldn't work, therefore it doesn't." Few commentators have backed up their criticisms with empirical evidence. In this evaluation, the inverted factoring technique yielded more accurate recovery of underlying populations than many previously studied hierarchical algorithms. Moreover, the inverted factor technique was found to be among the most accurate methods yet tested on these benchmark mixtures. These results agree with the previous study by Morf, Miller and Syrotuik (1976) who, on the basis of an objective comparison, concluded that inverted factor analysis was superior to the complete linkage algorithm in identifying subtypes of individuals. Thus, inverted factor analysis appears to be a useful taxonomic tool--despite the implausibility of the factor analytic model for generating typologies, the "inferiority" of the correlation coefficient as a measure of profile similarity, and numerous other problems (e.g., determining number of factors, assigning objects to groups, etc.).

In terms of recovering underlying mixture populations, differences between rotational methods were minimal. The more crucial methodological problem involved selecting the appropriate number of factors. Determining the number of factors empirically via Cattell's scree test resulted in more accurate solutions than the alternative procedure of setting the number of factors equal to the number of populations. Blashfield and Morey (1980) also reported that the scree test was quite accurate in determining the correct number of populations in their MMPI Monte Carlo data. This is a potentially important finding because the scree test does not depend upon a priori knowledge regarding "true" underlying populations. Thus, this procedure may be useful in determining number of underlying groups in applications to real data. This is a major asset of the inverted factor technique. Hierarchical clustering algorithms, by contrast, do not produce a discrete number of clusters, but rather a hierarchical arrangement of objects and groups. Determining the appropriate number of clusters is an unsolved problem, although some work has been done on developing objective criteria for making this decision (e.g., Mojena, 1977).

The inverted factor technique also embodies a simple mechanism for manipulating the coverage of the resulting classifications. In this study, for example, objects were assigned to groups on the basis of their highest factor loadings. Raising the minimum loading required for assignment decreased coverage, but increased accuracy. The ability to vary coverage may be valuable in research applications. In an epidemiological study, for instance, high coverage may be desirable in order to account for the generality and distribution of phenomena in a population. In other situations, it may be advantageous to construct extremely homogeneous groups. This would dictate low coverage, but the resulting groups would encompass individuals representing relatively "pure types". Future research should explore different methods of translating factor loadings into groups. The dual cutoff criteria used by Blashfield and Morey (1980), for example, appear promising. Such stringent assignment rules result in reduced coverage, but yield more homogeneous and distinct groups. Moreover, such assignment rules may yield typologies that are more predictive of external criteria.

Finally, additional comparisons among clustering and classification methods are needed. There are few standard procedures for constructing empirically based taxonomies and little is known about the relative merits of different methods. Objective comparisons are necessary, not only to combat dogmatic arguments for or against specific approaches, but also to identify those procedures best suited to behavioral research. The results obtained here indicate that inverted factor analysis yields accurate recovery of underlying populations from multivariate normal mixutres. Evaluations on other types of mixtures and evaluations involving other criteria (e.g. replicability, sensitivity to data perturbation, etc.) would be valuable.

Achenbach, T.M., and Edelbrock, C. Taxonomic issues in child psychopathology. In T.H. Ollendick and M. Hersen (Eds.) Handbook of Child Psychopathology. New York: Plenum Press, in press, 1981.

Baggaley, A.R. Intermediate correlational methods. New York: Wiley, 1964.

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

Blashfield, R., & Morey, L.C. A comparison of four clustering methods using MMPI Monte Carlo data. Applied Psychological Measurement, 1980, 4, 57-64.

Broverman, D.M. Effects of score transformations in Q and R factor analysis. Psychological Review, 1961, 68, 68-80.

Butler, E.W., and Adams, S.W. Typologies of delinquent girls: some alterma-tive approaches. Social Forces, 1966, 44, 401-407.

Carlson, K.A. Classes of adult offenders: A multivariate approach. Jouranl of Abnormal Psychology, 1972, 79, 84-93.

Cattell, R.B. The three basic factor-analytic research designs - their interrelations and derivatives. Psychological Bulletin, 1952, 49, 499-520.

Cattell, R. B. The scree test for number of factors. Multivariate Behavioral Research, 1966, 1, 245-276.

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

Collins, H.A., Burger, G.K., and Taylor, G.A. An empirical typology of heroin abusers. Journal of Clinical Psychology, 1976, 32, 473-476.

Eades, D.C. The inappropriateness of the correlation coefficient as a measure of taxonomic resemblance. Systematic Zoology, 1965, 14, 98-100.

Edelbrock, C. Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody. Multivariate Behavioral Research, 1979, 14, 367-384.

Edelbrock, C., and McLaughlin, B. Hierarchical cluster analysis using intraclass correlations: A mixture model study. Multivariate Behavioral Research, 1980, 15, 299-318.

Evenson, R.C., Altman, H., Sletten, I.W., & Knowles, R. Factors in the description and grouping of alcoholics. American Journal of Psychiatry. 1973, 130, 49-53.

Fleiss, J.L. Classification of the depressive disorders by numerical typology. Journal of Psychiatric Research, 1972, 9, 141-153.

Fleiss, J.L., Lawlor, W., Platman, S.R., and Fieve, R.R. On the use of inverted factor analysis for generating typologies. Journal of Abnormal Psychology, 1971, 77, 127-132.

Fleiss, J.L. and Zubin, J. On the methods and theory of clustering. Multivariate Behavioral Research, 1969, 4, 235-250.

Guertin, W.H. An inverted factor--analytic study of schizophrenics. Journal of Consulting Psychology, 1952, 16, 371-375.

Harman, H.H. Modern factor analysis, 3rd Edition. Chicago: University of Chicago Press, 1976.

Hubert, L.J., and Levin, J.R. Evaluating object set partitious: Free sort analysis and some generalizations. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 459-470.

Jones, K.J. Problems of grouping individuals and the method of modality. Behavioral Science, 1968, 13, 496-511.

Katz, M.K., and Cole J.O. A phenomenological approach to the classification of schizophrenic disorders. Diseases of the Nervous System, 1963, 24, 147-154.

Kuiper, F.K. and Fisher, L. A Monte Carlo comparison of six clustering procedures. Biometrics, 1975, 31, 777-783.

Kunce, J.T., Ryan, J.J., and Eckelman, C.C. Violent behavior and differential WAIS characteristics. Journal of Consulting and Clinical Psychology. 1976, 44, 42-45.

Lorr, M. (Ed.) Explorations in typing psychotics. Oxford: Pergamon Press, 1966.

Lorr, M., Bishop, P.F., and McNair, D.M. Interpersonal types among psychotic patients. Journal of Abnormal Psychology, 1965, 70, 468-472.

Lorr, M., and Radhakrishnan, B.K. A comparison of two methods of cluster analysis. Educational and Psychological Measurement. 1967, 27, 531-543.

Meyer, L., and Kline, P. On the use of delegate scores in the G analysis of clinical data: A Q factor method for diagnostic classification. Multivariate Behavior Research, 1977, 12, 479-486.

Milligan, G.W. and Isaac, P. The validation of four ultrametric clustering algorithms. Pattern Recognition, 1980, 12, 41-50.

Mojena, R. Hierarchical grouping methods and stopping rules: An evaluation. Computer Journal, 1977, 20, 359-363.

Monro, A.B. Psychiatric types: A Q-technique study of 200 patients. Journal of Mental Science, 1955, 101, 330-343.

Morf, M.E., Miller, C.M., and Syrotuik, J.M. A comparison of cluster analysis and Q-factor analysis. Journal of Clinical Psychology, 1976, 32, 59-64.

Mulaik, S. The foundations of factor analysis. New York, McGraw-Hill, 1972.

Overall, J.E., Hollister, L.E., Johnson, M., and Pennington, V. Nosology of depression and differential response to drugs. Journal of the American Medical Association, 1966, 195, 162-164.

Overall, J., and Klett, C.J. Applied Multivariate Analysis, New York: McGraw-Hill, 1972.

Overall, J.E., and Porterfield, J.W. The powered vector method of factor analysis. Psychometrika, 1963, 61, 415-422.

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

Raskin, A. and Crook, T.H. The endogenous-neurotic distincition as a predictor of response to antidepressant drugs. Psychological Medicine, 1976, 6, 59-60.

Ross, J. The relation between test and person factors. Psychological Review, 1963, 70, 432-443.

Ryder, R.G. Profile factor analysis and variable factor analysis. Psychological Reports, 1964, 15, 119-127.

Stephenson, W. Introduction to inverted factor analysis with some applications to studies of orexsis. Journal of Educational Psychology, 1936a, 27, 353-367.

Stephenson, W. The inverted factor technique. British Journal of Psychology, 1936b, 26, 344-361.

Stephenson, W. The study of behavior: Q-technique and its methodology. Chicago: University of Chicago Press, 1953.

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

TABLE 1

Comparison between inverted factor analysis (varimax rotation) and the average linkage algorithm

	Method		
Accuracy Measure	Inverted Factor Analysis	Average Linkage	Paired t-value[a]
Kappa	.655	.793	1.49
Rand	.789	.864	1.44
Lambda	.656	.801	1.20

Note: Table entries are mean values for 20 mixtures. [a]df=19. None of the paired t-tests were significant ($p > .10$).

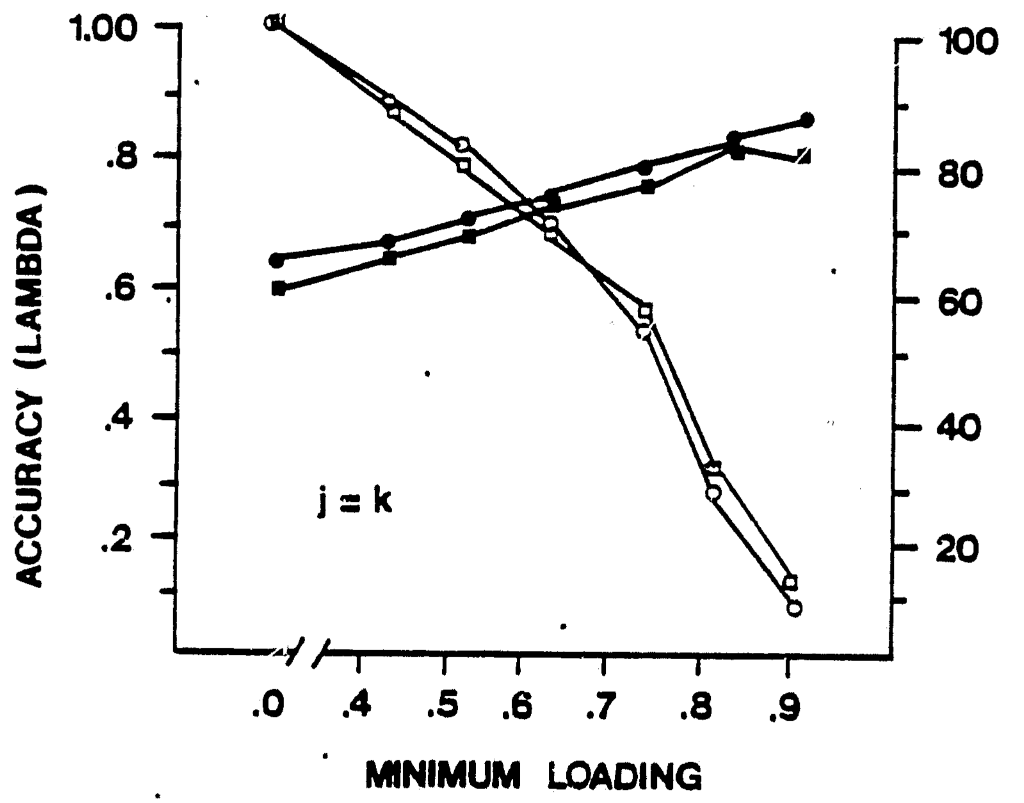

Fig 1. Accuracy and coverage functions for j = k factoring solutions (● varimax accuracy ○ varimax coverage ■ direct quartimin accuracy □ direct quartimin coverage).

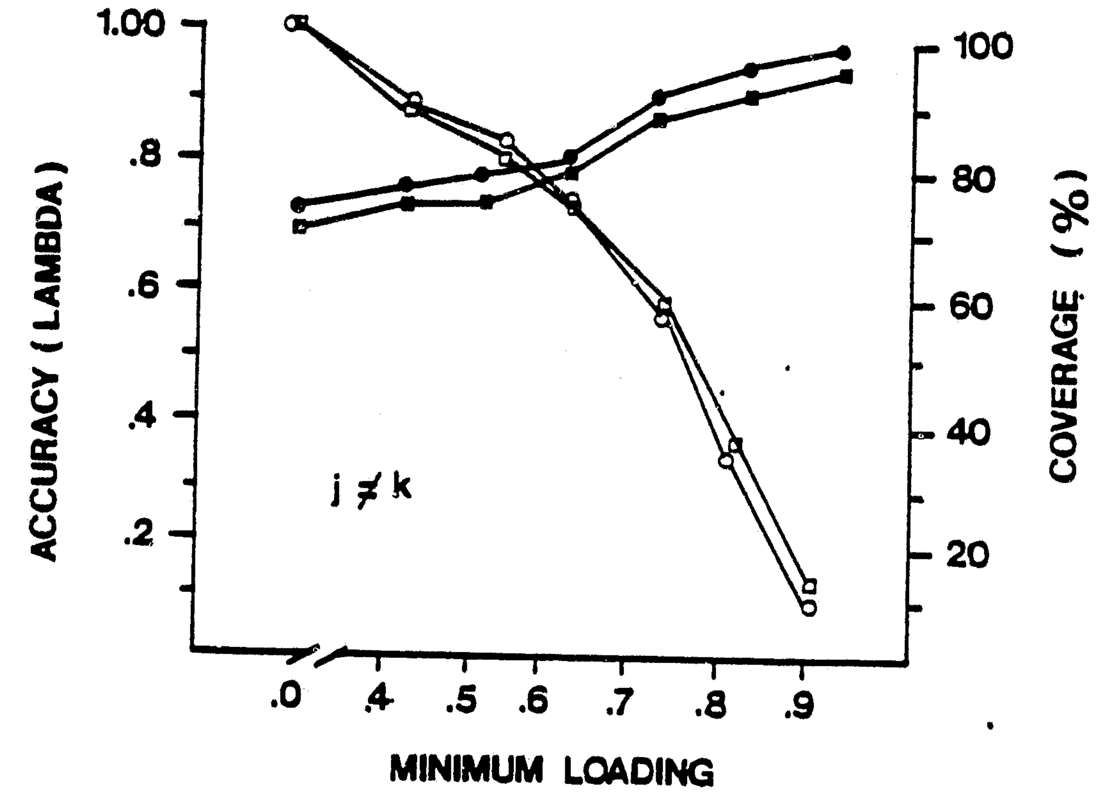

Fig 2. Accuracy and coverage functions for j ≠ k factoring solutions (● varimax accuracy ○ varimax coverage ■ direct quartimin accuracy □ direct quartimin coverage).

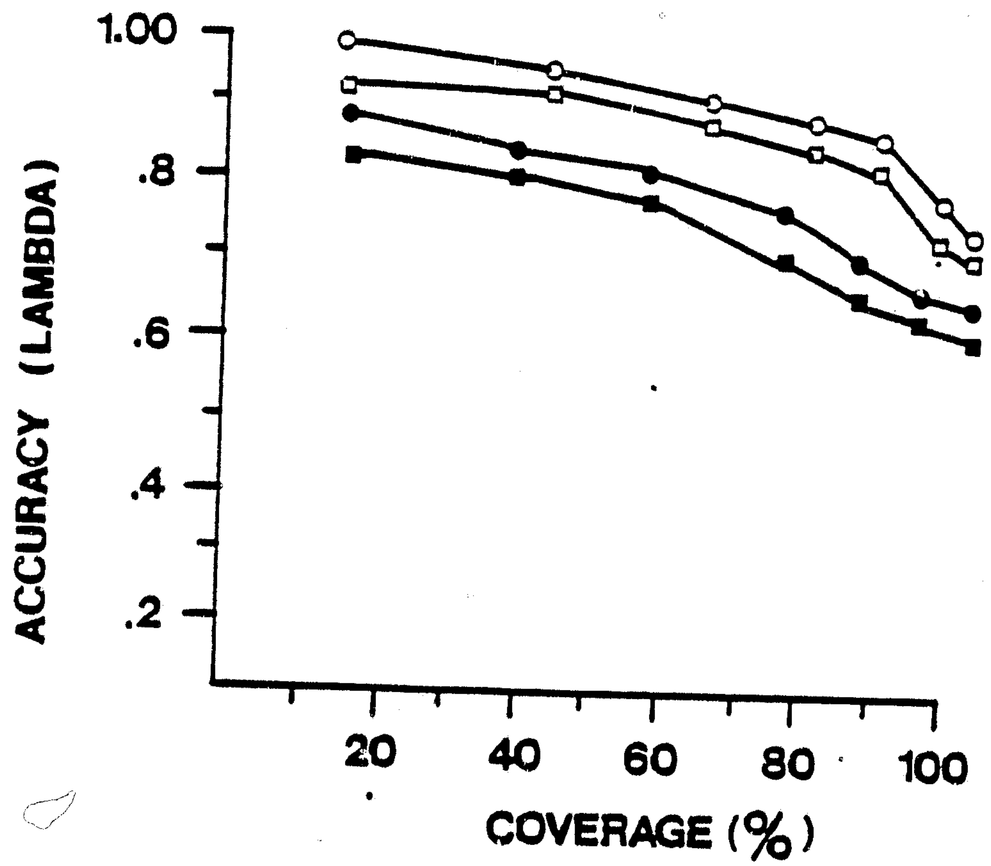

Fig 3. Comparison between $j = k$ and $j \neq k$ factoring solutions (● $j = k$ varimax ■ $j = k$ direct quartimin ○ $j \neq k$ varimax □ $j \neq k$ direct quartimin).

END