Journal of Counseling Psychology
1987, Vol. 34, No. 4, 456–468

Applying Cluster Analysis in Counseling Psychology Research

Fred H. Borgen
Iowa State University

David C. Barnett
Western Connecticut State University

As a research technique that has grown rapidly in applications in many scientific disciplines, cluster analysis has potential for wider use in counseling psychology research. We begin with a simple example illustrating the clustering approach. Topics covered include the variety of approaches in clustering, the times when cluster analysis may be a choice for analysis, the steps in cluster analysis, the data features, such as level, shape, and scatter, that affect cluster results, alternate clustering methods and evidence indicating which are most effective, and examples of clustering applications in counseling research. Although we make an attempt to provide a comprehensive overview of major issues, the reader is encouraged to consult several good recent publications on the topic that are especially relevant for psychologists.

Cluster analysis is a classification technique for forming homogeneous groups within complex data sets. Both the clustering methods and the ways of applying them are extremely diverse. Our purpose in writing this article is to provide an introduction and a road map for applying these techniques productively to research in counseling psychology. The cluster analysis literature is huge, is scattered among many diverse disciplines, and is often arcane. We have made an attempt to cull those aspects most relevant and useful to psychologists from this literature. Most of the discussion in the psychological community about how best to apply cluster analysis to obtain robust, valid, and useful results has taken place within the past 5 years. We seem to be on the verge of a consensus, which has long been needed in an often bewildering field.

In the past 30 years, a number of clustering methods, often with their own vocabulary and approaches, have sprouted within a wide variety of scientific disciplines. The earliest sustained applications were in problems of biological classification, within the field called *numerical taxonomy* (Sokal & Sneath, 1963). Today, clustering is applied to problems as different as the grouping of chemical structures (Massart & Kaufman, 1983) and the classification of helpful and nonhelpful events in counseling (Elliott, 1985). Computerized methods for generating clusters have been developed and made increasingly available over the last decade. Applications of clustering have mushroomed in many disciplines, including the social sciences. In an annual bibliographic search performed by the Classification Society (Day, 1986) 1,166 entries are shown for the 1985 scientific literature alone.

A Cluster Analysis Example

A simple example of the use of clustering in grouping people might be useful to help those unfamiliar with this technique to gain a better sense of what it does. Imagine that you are a college counselor and have been assigned to work with a group of 20 entering freshmen who have not yet declared a major. Your examination of the retention research suggests that it would be helpful to form small support groups of these students, and you would like to group them on the basis of their interests. At orientation, you gave each student the Strong-Campbell Interest Inventory (SCII; Hansen, 1986). Overall, the students' scores on the six General Occupational Theme scales are average to very low, which is not very useful in forming groups. Can small, homogeneous groups be formed by using cluster analysis with other SCII scores? You decide that an interesting experiment would be to form groups that are relatively similar on the SCII Academic Comfort (AC) scale and the Introversion–Extraversion (IE) scale. You surmise that these scales will provide a basis for grouping students on the basis of their comfort in and likelihood of persistence in an academic environment, and their vocational preferences for either ideas and things or for work with people, especially in sales (cf. Hansen, 1986, pp 23–25). Table 1 shows the AC and IE scores for the 20 students. Figure 1 shows the plotting of students' scores on the AC and IE dimensions; the results suggest some natural subgroups. (Although the data are for real counseling clients, we have selected these 20 from a group of 50 to provide a clearcut and simple clustering illustration. A plot of the sample of 50 shows less distinctive subgroups.)

The first step is to calculate the proximity between each pair of students. For this example we use squared Euclidean distance or d^2, which is calculated by finding the difference for each score, squaring the differences, and summing these values over the profile. Thus, the distance between Students 1 and 2 on AC and IE is $(39 - 36)^2 + (71 - 34)^2$, or 1,378. A like index is calculated for each pair of the 20 students and arrayed in a proximity matrix (see Table 2). Once we have a proximity matrix, we can apply a clustering method to divide the group into homogeneous subgroups. For this sample problem, we chose Ward's (1963) widely used clustering method. Ward's method searches the proximity matrix and groups the two persons with the smallest distance value, which in this sample are Students 12 and 13, with a distance of 1.0. Students 3 and 11 have a distance of 4.0, so they are grouped at Step 2; likewise, Students 7 and 10, with a distance of 5.0, are grouped at Step 3. The next closest are Students 18 and

We are grateful to Bruce E. Wampold and two anonymous reviewers for their careful review of earlier drafts of this article.

Correspondence concerning this article should be addressed to Fred H. Borgen, Department of Psychology, Iowa State University, Ames, Iowa 50011.

Table 1
Standard Scores on the SCII Academic Comfort and Introversion–Extraversion Scales for the 20 Students in the Sample Cluster Analysis

	SCII standard scores	
Student	Academic Comfort	Introversion–Extraversion
1	39	71
2	36	34
3	13	56
4	50	44
5	34	40
6	76	49
7	21	64
8	15	57
9	15	66
10	23	63
11	11	56
12	36	61
13	36	62
14	58	52
15	58	40
16	41	60
17	53	50
18	41	39
19	44	37
20	43	40

Note. SCII = Strong-Campbell Interest Inventory.

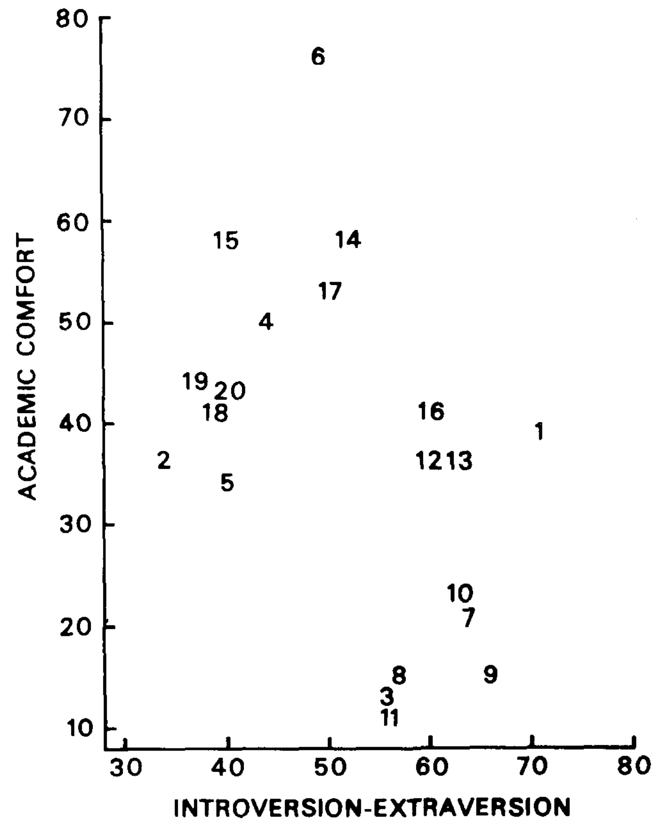

Figure 1. Plot of 20 students on Academic Comfort and Introversion–Extraversion scales.

20, so they are grouped at Step 4. The method continues to merge groups in a way that will keep the within-group variance at a minimum. The process continues as the 20 students are grouped in a hierarchical, treelike structure, and ends when one group has been formed. A tree diagram (or dendogram) is used (see Figure 2) to represent the hierarchical structure of the data. Individual objects (in this case, students) are found at the tips of the branches. As objects are combined, the branches of the diagram become larger and larger until the trunk is reached. Objects on the same branches are thus more closely related to each other than to those on other branches.

Because this example is so simple, with clustering based on just two variables, the reader can examine the two-dimensional space in Figure 1 and readily see how the clustering method proceeds to group together the most similar students. For example, note how Student 6 is an outlier in the visual space, and remains uncombined with others until the four-group stage.

Ward's method provides an index of within-group error at each stage of the grouping. This index can be plotted, as shown in Figure 3, to aid in selection of the best grouping level. When the error index shows a jump upward, it indicates that relatively disparate groups have been combined at that stage.

Thus, Figure 3 indicates a jump in within-group heterogeneity at the four-group stage when the outlier Student 6 has been added. But the major jump occurs when three groups are combined, leading us to conclude that a four-group solution is most appropriate for this example. (Recall that our sample is somewhat contrived for the sake of clarity, so the jump in the error index is more striking than is typically the case.) We can examine the group means for these clusters on the AC and IE scales (Table 3 and Figure 4). We can use these results to assign the students to counseling groups that are either (a) moderate AC, high IE; (b) moderate AC, moderate IE; (c) low AC, high IE; or (d) high AC, moderate IE.

This simplified sample illustrates how clustering can be used to group people (or other objects) into subgroups with different patterns of scores on the original variables. Although this example is limited in complexity, we hope that it will help readers gain a better notion of clustering: the issues that have to be addressed in its use, the choice of methods that it involves, and an understanding of its results.

An Overview of Clustering

Clustering as Alternative Algorithms

A cluster method is essentially a set of rules for dividing up a proximity matrix to form groups of similar objects. This sequence of rules or procedures is referred to as an algorithm. In contrast, most traditional statistical methods make more formal assumptions about a data set and solve the statistical problem through a formal mathematical approach.

Clustering People

Our previous example illustrates psychologists' most common use of clustering to form similar groups of persons (P)

Table 2
Proximity Matrix (Squared Euclidean Distance) for the 20 Students in Sample Cluster Analysis

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	1,378																		
3	901	1,013																	
4	850	296	1,513																
5	986	40	697	272															
6	1,853	1,825	4,018	701	1,845														
7	373	1,125	128	1,241	745	3,250													
8	772	970	5	1,394	650	3,785	85												
9	601	1,465	104	1,709	1,037	4,010	40	81											
10	320	1,010	149	1,090	650	3,005	5	100	73										
11	1,009	1,109	4	1,665	785	4,274	164	17	116	193									
12	109	729	554	485	445	1,744	234	457	466	173	650								
13	90	784	565	520	488	1,769	229	466	457	170	661	1							
14	722	808	2,041	128	720	333	1,513	1,874	2,045	1,346	2,225	565	584						
15	1,322	520	2,281	80	576	405	1,945	2,138	2,525	1,754	2,465	925	968	144					
16	125	701	800	337	449	1,346	416	685	712	333	916	26	29	353	689				
17	637	545	1,636	45	461	530	1,220	1,493	1,700	1,069	1,800	410	433	29	125	244			
18	1,028	50	1,073	106	50	1,325	1,025	1,000	1,405	900	1,189	509	554	458	290	441	265		
19	1,181	73	1,322	85	109	1,168	1,258	1,241	1,682	1,117	1,450	640	689	421	205	538	250	13	
20	977	85	1,156	65	81	1,170	1,060	1,073	1,460	929	1,280	490	533	369	225	404	200	5	10

across a set of variables (V). Often, subgroups of people with relatively homogeneous profiles on a psychological test are identified. The usual practice is to begin clustering with a P × V matrix, with the profiles of the people arrayed in the rows and the variables in the columns. Then clustering is designed to group together the most similar rows (of people) in the matrix. Note that our objective here is to use more than one data point for each person. We want the groupings to reflect the total profiles. In this article we refer to this kind of data as multivariate (cf. Cooley & Lohnes, 1971), to distinguish it from a univariate focus on one variable at a time.

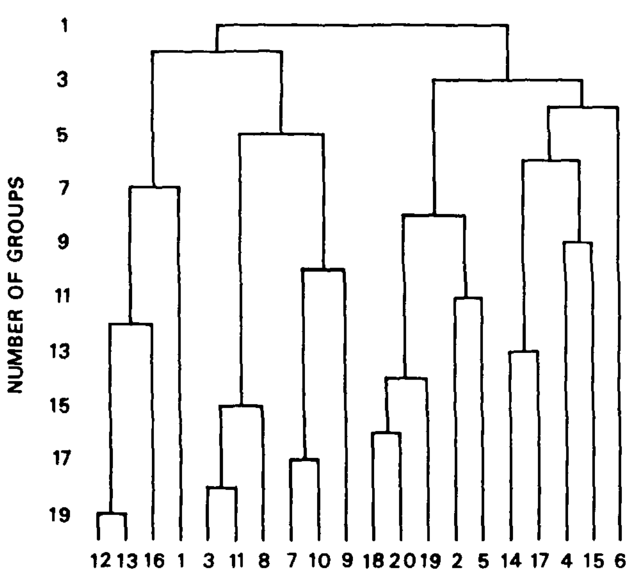

Figure 2. Dendogram showing the clustering of 20 students with Ward's hierarchical grouping method.

Clustering Variables

In a second major clustering application in psychology, researchers take the opposite approach to the P × V data matrix. Here their goal is to form similar sets of variables that show similar patterns or correlations over the group of people. This approach is similar to factor analysis, but differs in the way that variables are assigned discretely to groups. This tactic could be used by researchers to cluster a large number of variables and to form new summary composite variables. Use of this method could produce a new inventory, in which the items were treated as variables and clustered to form scales.

Less Conventional Applications

Because clustering is a general technique for detecting patterns in a data matrix, it need not be restricted to the usual tasks of grouping people or variables. For example, it could

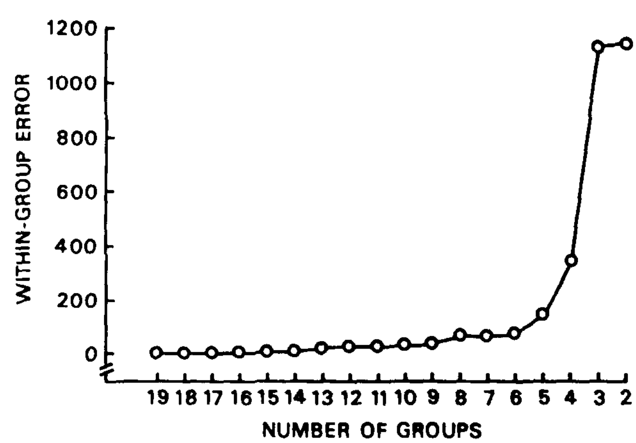

Figure 3. Index of within-group error for clustering 20 students with Ward's hierarchical grouping method.

Table 3
Group Means and Standard Deviations After Clustering 20 Students at the Four-Group Level

		Academic Comfort		Introversion–Extraversion	
Group	*N*	*M*	*SD*	*M*	*SD*
1	4	38.0	2.4	63.5	5.1
2	5	39.6	4.4	38.0	2.5
3	6	16.3	4.7	60.3	4.5
4	5	59.0	10.1	47.0	4.9
Total	20	37.2	17.2	52.1	11.1

be used to identify groups of people who, on a single variable, change in different ways over occasions. Thus, in counseling research we might use clustering to group clients with different patterns of change on a single variable over, say, ten counseling sessions, or even over 5-min segments within a single session. For example, a cognitive therapist might ask clients to complete a depression scale prior to each counseling session. One could group these clients by clustering to identify those with similar patterns of change over a prescribed 15-week protocol. Then one could use other research methods, such as analysis of variance, to understand the differences between client groups responding to therapy in different ways.

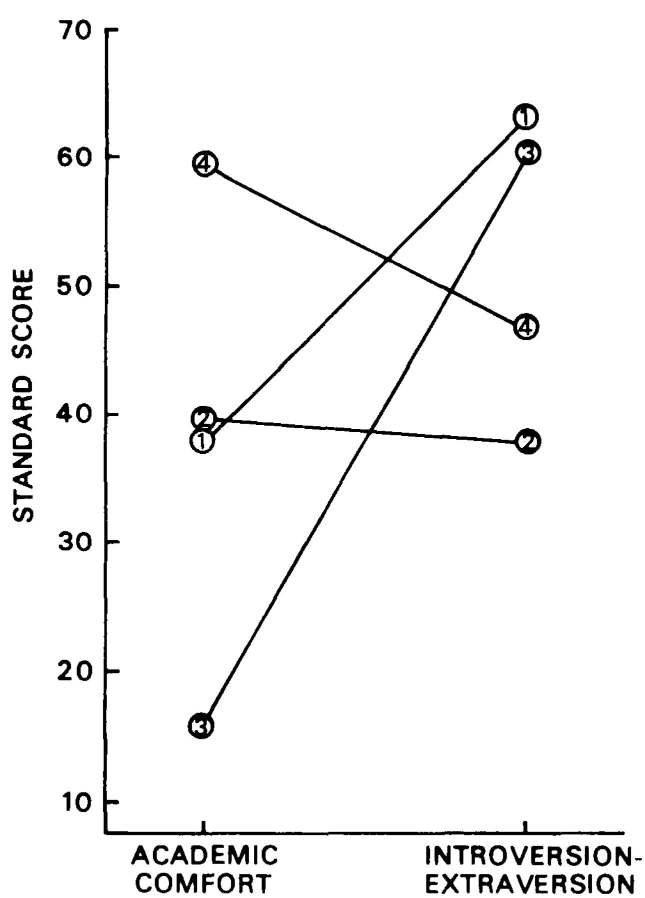

Figure 4. Group means on Academic Comfort and Introversion–Extraversion after clustering at the four-group level.

In addition, other objects besides people or variables might be grouped through cluster analysis. A good example is Elliott's (1985) preliminary taxonomy of helpful and nonhelpful events in counseling. This is discussed more fully later in this article. A different example is Wampold and White's (1985) creative analysis of the structure of research themes in counseling psychology. They focused on the 27 articles on counseling process and outcome that appeared in this journal in 1982. In an examination of the citations in each article, they used overlapping citations as a way of measuring similarity among the 27 articles. The end product of their cluster analysis of this literature was a tree diagram that enabled them to conclude that the primary underlying theme of the 27 articles was the social influence model. Note that Wampold and White were able to form a proximity matrix directly by using the number of overlapping citations for each pair of the 27 articles. They did not begin with the usual P × V or similar data matrix.

Wampold and White's (1985) study illustrates that the researcher does not always begin with the usual P × V matrix with objects measured on a set of variables. Rather it is sometimes possible and desirable to form directly a proximity matrix of the objects on the basis of some characteristic of linkage between the objects. Thus, in Wampold and White's study they first constructed a proximity matrix among the 27 articles by counting the number of overlapping citations for each pair of articles. Similarly, other researchers might wish to create a proximity matrix directly from information linking the objects on a pairwise basis, such as judged similarity of objects or frequency of interactions between subjects, or some other direct measure.

Schematic Overview of Clustering Alternatives

We have attempted to show the major ways in which clustering might be applied. One may focus on grouping objects or on grouping variables. In some cases one may not have a P × V data matrix, but may be able to form a proximity matrix and apply a clustering algorithm directly to it. In all cases, there is a proximity matrix and a clustering method, even though both of these conceptual steps are often embedded with many clustering programs. Schematically, these alternatives to conceptualizing and computing can be shown as follows:

P × V data matrix	→	Proximity matrix of objects (e.g., subjects)	→	Clustering algorithm
P × V data matrix	→	Proximity matrix of variables	→	Clustering algorithm
		Proximity matrix of objects	→	Clustering algorithm

Comparisons With Factor Analysis

In cluster analysis, researchers use multivariate data to assign initially unclassified objects into groups. This feature distinguishes it from other multivariate statistical methods, which also involve multiple measures on multiple objects. A comparison of clustering with factor analysis may provide useful information.

Both cluster analysis and factor analysis are methods for identifying the underlying structure in a multivariate data set. Most typically, factor analysis is used in order to simplify a set of variables, and cluster analysis can also be used for that purpose. Borgen and Weiss (1971) stated that

> The central difference is in the treatment of the variance of a variable: Factor analysis usually partitions the variance among several sources or factors, while cluster analysis assigns the total variance to an underlying "source." Thus, cluster analysis yields results where variables are grouped into discrete sets or "clusters," while the results of factor analysis are typically less clear, with parts of the variance of each variable attributed to each of several "subsets" or factors . . . Choice of cluster versus factor analysis depends, therefore, on the objectives of the study. (p. 583)

Clustering techniques have much in common with factor analysis. Both can be used to reduce a complex data set to its central features. Both require considerable judgment on the part of the researcher, because the final product is dependent on how the analysis is done. In both cases, there is unresolved controversy about the most effective way to perform them, and different researchers have different approaches.

Emerging Guidance for Cluster Analysts

Recent developments make this a good time for a novice to be introduced to the use of cluster analysis. Summaries of cluster analysis that are targeted to social scientists and are written at a relatively nontechnical level and that provide information about the potential misuses of cluster analysis have recently been published. We recommend books by Lorr (1983), Aldenderfer and Blashfield (1984), and Blashfield (1984). These authors are well-respected psychologists with impressive experience in applying clustering methods, especially in classification of psychopathology. We also recommend books by Everitt (1980), Hartigan (1975), Mezzich and Solomon (1980), Norusis (1985), Overall and Klett (1972), and Romesburg (1984).

The availability of comprehensive information on the use of clustering methods in psychology-related situations is especially important because these methods and their related computational packages are evolving so rapidly. In the past 10 years, a few psychologists have been especially active in testing the utility of the commonly used clustering techniques through a series of increasingly sophisticated simulations with Monte Carlo data with known structure. This important work (discussed in the section *Comparative Evaluations of Cluster Methods*), which makes it possible for researchers to confidently recommend the better methods for psychological applications, is covered in these most recent discussions. There are other good introductory materials, but they lack the advantage of having been recently published.

Early use of clustering in research required that researchers have advanced computer programming skills in order to execute the analysis. Now, fortunately, the most popular and recommended clustering routines can be implemented by use of most of the widely available computer packages, including BMDP (Dixon, 1983), SAS (SAS Institute, 1982), and SPSS^{X} (Norusis, 1985; SPSS, Inc., 1986). Programs for clustering procedures are also available for microcomputers (including IBM PC and Apple MacIntosh machines).

In this article, we hope to alert researchers to some of the ways in which cluster analysis can go awry, as well as to its utility. Our treatment should be seen only as a beginning guide. The issues involved in cluster analysis are complex, as in all multivariate methods, and their complexity is multiplied by the diversity of existing clustering methods. Researchers are urged to read the sources recommended here and to seek consultation, when needed, from others more experienced in measurement and clustering issues.

Numerous sources of information on cluster analysis were published in the 1960s and 1970s. Most of these references are available in Aldenderfer and Blashfield (1984), Blashfield (1984), Everitt (1980), and Lorr (1983). Few of these references are cited here, so that we can recommend more recent works about cluster analysis in psychology. Serious researchers should be aware of the earlier sources in order to consult them when they appear to be relevant to their work.

When Should Cluster Analysis Be Considered?

The usual purpose of cluster analysis is to identify homogeneous subtypes within a complex data set. Typically, we do not know a priori the natural groupings or subtypes, and we wish to identify groups within a data set. We wish to form classifications, taxonomies, or typologies that represent different patterns in the data. In Gangestad and Snyder's (1985) phrase, we hope ideally "to carve nature at its joints" (p. 317). Clustering algorithms are used to search a data set to partition the objects (or people) into relatively distinct groups. In research problems in which we have prior knowledge of the types, clustering may not be appropriate, and more traditional techniques such as discriminant function analysis and multivariate analysis of variance, may be more useful to show how the known groups differ.

Choice of Problem and Research Question

The initial conceptualization of a clustering problem is most crucial, because it determines the ultimate usefulness of the classification. For clustering to be productive, it must be applied to problems that exploit the advantages of cluster analysis, such as those with data that are multivariate and fairly complex. There needs to be an advantage in grouping objects with a direct empirical base, rather than on some other basis, such as investigator judgment or a prior established classification system. For example, if the natural grouping is obvious from examination of a small correlation matrix, then it is questionable whether clustering is useful. In such a case, clustering might best be used, if at all, to confirm a grouping derived by other approaches. Several specific purposes for the use of clustering can be delineated (see also Romesburg, 1984). In practice, there may be more than one of these classification goals in a research problem. Three major purposes for clustering can be delineated: exploration, confirmation, and simplification.

Exploration. On the exploratory versus confirmatory continuum, cluster analysis is most often used as an exploratory technique. If the research area is relatively new, clustering may be a productive early step to identify and structure the subgroups that are of potential value in understanding the research problem. Clustering is used by researchers to explore a data set to produce a summary of its structure. There are no generally accepted formal statistical tests in cluster analysis. Rather, the approach is one of seeking structure. When such structure has been identified through clustering exploration, the research process should continue toward a confirmation, testing, and validation of the structure. Elliott's (1985) construction of a taxonomy of helpful and nonhelpful events in counseling is an excellent example of an application used for the purpose of exploration.

Data exploration by means of clustering carries a special caveat, because the choice of method so strongly determines the results of the probe. Moreover, most cluster methods impose a structure on the data, even if no underlying groups exist in the data. Thus, it becomes very important, especially if any generalization is intended, to apply multiple analyses to the same data set. (It is useful to note that even simple conventional statistics have this capacity to mislead. For example, if one obtains only a mean and standard deviation for a distribution that happens to be bimodal, one has obtained a method-bound result that is misleading.)

Most current interest inventories focus on the similarities within occupations rather on the differences that exist. Zytowski and Hay (1984) used interest inventory data from two samples of women in five different occupations, representing five different Holland types. When clustered, neither sample recaptured the membership of the five original groups. The authors interpreted these findings as suggesting that there may be important differences in the interests of persons within the same occupations. Such research questions are appropriate for exploratory cluster analysis, in which researchers take existing groups (in this case occupations) and subject them to a clustering procedure in order to look for previously hidden relations. The notion of homogeneity within occupations assumed in much of the literature is challenged by such attempts. Additional research in which experimenters would look for differences in other occupations and with both genders is advisable.

Confirmation: Testing of prior classifications. If prior knowledge or theory suggests a particular psychological classification, clustering might be used to test the classification. Clustering could be applied to data similar to the original data as a confirmation of a prior approach that might have been theoretically or partly empirically derived. It could be used to replicate a prior cluster analysis, or it could be used to test the generalizability of the taxonomy in a new domain. This use is less common in counseling research at present. Thus, recent examples of it are not easy to cite. Some topics for which this approach should be considered are additional validity studies of the Diagnostic and Statistical Manual of Mental Disorders, 3rd ed. (DSM-III) classification schema, Holland's hexagonal classification, the occupational classification used by the American College Testing Program, and numerous extant classifications for populations such as alcoholics. Paykel (1971, 1972) was an early leader in using cluster analysis to develop and test a classification system for subtypes of depression.

Simplification. Clustering can be an especially powerful technique for simplifying a complex data set. It can be used to group the objects when the use of human judgment would be tedious, subjective, or practically impossible. Wampold and White's (1985) use of clustering to examine themes in counseling research nicely illustrates the use of the technique for this purpose. Instead of presenting the cumbersome 27 × 27 proximity matrix for 27 articles, they clustered this matrix to show its central patterns. Thus they were able to show the core research themes for a potentially confusing maze of data. In addition, the hierarchical clusters were displayed in a tree that revealed the nesting between the research themes. Likewise, our beginning example in this article represents the use of clustering to simplify a data set.

Another example would be the use of Minnesota Multiphasic Personality Inventory (MMPI) profiles to group a large number of offenders (cf. Megargee & Bohn, 1979). Clustering provides an effective alternative to the prohibitively laborious means of examining the frequency of all possible combinations in such a sample. If one were to consider rigorously the concept of similarity of such profiles, the 10 main MMPI scales alone would yield 10! potential categories, yet even this large number of groups does not include profile elevation, shape, and scatter. In using clustering with a distance measure, researchers can consider all three and collapse the offenders into potentially meaningful groups.

College and university counseling centers often have some sort of checklist of problems of concern that new clients complete on an intake form. Presumably, a matrix of the problems by clients could be examined, but it would probably be difficult to decipher any particular patterns. Cluster analysis might be fruitful in identifying patterns of problems that certain subgroups of clients are experiencing. Such identification might enable the counselors to take more proactive measures, such as focused outreach interventions, the development of specialized groups, or more specific in-service training for the staff, to assist clients with these problems.

Steps in Cluster Analysis

When researchers have determined that cluster analysis is a productive data reduction strategy, they can proceed with the specific steps in the analysis. These steps are potentially complex, especially if the implicit decisions in the process that must be made by researchers are recognized. The implicit decisions are especially important because they ultimately determine the kinds of clusters that are obtained and the contribution of those clusters to advancing our knowledge and practice. With the ready availability of computer packages for clustering, it is tempting to choose a clustering method on the basis of the availability of a package and its accompanying range of options rather than on awareness of what information can be obtained from the use of that program and the characteristics of that data set. In that case, crucial decisions are not made, but rather arrived at by default.

1. Measuring Proximity

Cluster methods are used to search for a proximity matrix in order to locate the most similar objects. In typical research, the investigator has a full P × V data matrix. In that case, the first computational step in doing a cluster analysis is to construct a matrix showing the similarity between each object. There are dozens of potential arithmetic measures of proximity that can tap aspects of similarity (or alternatively, dissimilarity) in different ways. Proximity measures are referred to in the literature by a variety of names, such as *association coefficients, measures of dissimilarity*, or *similarity measures*. (In this article we have adopted *proximity* measure as the generic term, in order to be consistent with the usage in the article on multidimensional scaling in this issue. However, readers should be alert to the fact that in the clustering literature *similarity* measure is very often used as the generic term.) In psychology the two most common proximity indexes have been correlation and squared Euclidean distance. The latter, like all distance measures, is strictly speaking a dissimilarity measure; similar objects will have a small distance between them. Correlation, on the other hand, is a direct similarity measure, and the similar objects will have a large correlation coefficient. Of course, the clustering algorithm must in some way "know" whether grouping should be based on small values (distances) or large values (say, correlations). The standard clustering packages typically handle this either by having a particular clustering command linked with a given proximity measure or by asking the user to specify the kind of proximity measure. However, it is a computing issue to be particularly alert about when a clustering method is applied directly to a proximity matrix as the primary data input.

Clustering subjects versus clustering variables. In the remainder of this section, we sketch some of the most important issues related to implicit features of the data that ultimately determine the resulting clusters. As a preface to those comments, we wish to emphasize that these issues are usually relevant whether one is clustering subjects or variables. Some of the conventional language used here may naturally lead a reader to think of either grouping subjects or grouping variables. But if we think of the usual beginning P × V (persons × variables) data matrix, these issues about the mathematical features of the matrix are similar whether we group by rows (persons, subjects, etc.) or by columns (variables). In one case, we have simply rotated the data matrix 180°, either conceptually or in actual fact, depending on the needs of the computer program. In cases in which one does not have an initial data matrix but rather is attempting to cluster directly from a proximity matrix, these issues may not have the same import, although their relevance should at least be considered by the researcher.

Level, shape, and scatter. For psychologists proximity has been historically treated under the rubric of profile similarity. The classic paper by Cronbach and Gleser (1953) continues to be required reading for cluster analysts. Cronbach and Gleser showed that profile or multivariate data can implicitly have elements of level (elevation), shape, and scatter. The choice of proximity index, because of its arithmetic features, directly determines which of these components are the basis for clustering (see also Skinner, 1978). Thus, correlation implicitly standardizes the data to remove level, and reflects only profile shape and scatter. Squared Euclidean distance, on the other hand, reflects all three elements of level, scatter, and shape. Clearly, in this basic example, the choice of index can make a difference in the kind of clusters obtained. If cluster analysts are to be informed, they must know how a proximity index reflects the important features of their data. In our hypothetical example with undeclared majors, we decided to attend to level differences as well as scatter and shape and hence chose squared Euclidean distance as our measure. Certain clustering methods (or computer programs) may also dictate which measures of proximity can be used with them.

These issues also apply to the clustering of variables, despite our tendency to think of clustering subjects because of the language used. For example, if we are grouping variables, the level issue might come into play if some of the variables are arbitrarily scaled with large numbers and other variables are arbitrarily scaled with small numbers. If we use Euclidean distance to cluster these variables without thinking through the implications, we will merely create two artifactual clusters of "large" variables and "small" variables that have no psychological meaning. Rather, in this case we must first do something to remove the arbitrary level differences across the variables, such as standardizing the variables to a common mean, or perhaps using correlation, which does not tap level differences.

Correlation. There are other more complex issues about the structure of data that affect the proximity measure. One issue is the correlation among the variables in the case in which people are clustered. For example, if five of the variables are highly intercorrelated, and the remaining two variables are uncorrelated, there are three underlying factors in the data set. If all seven dimensions are used in the proximity index, the first factor, represented five times, would be weighted nearly five times more in the index than would each of the remaining underlying factors. This may not make sense for a researcher's problem. In such a case the initial data should first be reduced by a method such as factor analysis, so that the underlying factors are equally weighted in the proximity index.

These correlation issues and underlying factor structure also apply in the case of clustering variables, and here their presence is not as likely to be recognized because of our usual ways of thinking about correlation. In this case we have the analogue of "people factors" because of the implicit correlations among people in our sample. If our sample contains five people who are very similar to each other and three people who are relatively uncorrelated with anyone else, those five similar people will be given the greater weight in the determination of which variables are clustered. This may or may not present a problem for our results. If the sampling of people represents the universe we wish to represent, then we probably do not have a problem. If they do not, then we need to rethink how we have sampled.

Standardization. Another decision that often needs to be made in clustering subjects is whether the data should first be standardized within each variable to remove large effects due to arbitrary differences in the standard deviations or means of the variables. Standardization is necessary when the level or scatter of the variables is an artifact of the scaling unit and not something that should be permitted to affect cluster differences. Likewise, when clustering variables, we may have differences between the people (level, scatter) that are artifacts and unrelated to the psychological construct we are studying. In that case, it may be appropriate to standardize the data within each person.

This is a sketch of some of the crucial issues related to preparing the data set and choosing a proximity measure. Decisions made at this stage of analysis are at least as important as those made about clustering per se. There is no substitute for the researcher's giving thoughtful consideration to the multivariate features of the data that are to be represented in the proximity matrix. Most of the texts on cluster analysis address these issues in more detail than is possible here.

2. Choice of Clustering Method

The options for selecting a clustering method are numerous. Later we enumerate some of the methods that have been widely applied in psychology and that have performed best in comparative tests. This should help to narrow the array of possibilities. Another general choice is between hierarchical and nonhierarchical methods. Hierarchical methods are the most widely used, and yield the potential asset of a clustering tree (see Figure 1), with the smaller clusters also arrayed successively within larger superclusters. The hierarchical approaches may take the form of the successive splitting off of groups into more numerous but ever smaller clusters or of the successive joining of similar individuals or clusters into smaller numbers of groups.

Nonhierarchical methods, on the other hand, are of particular use when there is prior knowledge about the likely number of clusters. These may sometimes be appropriate to use when testing a prior classification.

3. Evaluating and Generalizing the Clusters

Too often researchers have merely stopped at the second step mentioned after deriving a set of clusters. There is increasing recognition that a third step is needed both to establish the reliability of the clusters for a given data set and to embed the clusters in a program of construct validation that ultimately demonstrates the scientific and practical importance of the classification (see Aldenderfer & Blashfield, 1984; Blashfield, 1980; Gangestad & Snyder, 1985; Skinner & Blashfield, 1982).

Although most clustering techniques are quite powerful in dissecting a multivariate data set into homogeneous groups for that sample and those particular variables, the aware researcher will not assume too much for that particular cluster result. The remaining questions familiar to psychologists involve the issues of problem–method fit, sampling, generalizability, reliability, and validity. Just as these concepts often have no unique meaning or operational measure in psychometrics, they also have multiple implications in cluster analysis. During the 1980s several guides were presented, often by psychologists, for the most useful and lasting clustering research. These include the books written by Aldenderfer and Blashfield (1984), Blashfield (1984), Everitt (1980), Lorr (1983), and Romesburg (1984), and articles by Blashfield (1980), Lorr (1982), and Skinner and Blashfield (1982).

Many of these approaches are designed as hedges against the current unknowns about clustering methods. Because there are few ironclad guidelines about central issues such as the effect of data structure on results and how to select the final number of groups, it is advisable to practice methodological diversity with multiple samples. That is, the clusters that first emerge should not be accepted as the final solution, but should be tested and triangulated in a variety of ways. Clustering can be a valuable exploratory scientific tool when it is used in this fashion.

Alternative Clustering Methods

Major Hierarchical Clustering Methods

Hierarchical clustering methods form groups from a proximity matrix by successive steps so that the final result is a hierarchical structure with the groups linked in a tree or dendogram (see Figure 1). The most common hierarchical methods are agglomerative, that is, they start with the most specific objects and build them step-by-step into larger and larger nested clusters. Thus, with n objects the process typically starts with the location of the two most similar objects to form the first cluster, and then proceeds similarly for $n - 1$ steps until a full tree structure is formed. This is the approach used in our hypothetical example of clustering college students with undeclared majors. We started with an n of 20 students. The two most similar were combined into a cluster. At the next step, two other similar students were included. At the third step, a third student was merged into the first group. More and more individuals or groups were merged step-by-step until all subjects were recombined into a single group.

The much less commonly used divisive hierarchical methods proceed in the opposite direction, in which the n cases are first segmented into two groups and proceed in a stepwise way until all n cases have been separated into groups.

There are several features that are relatively specific to the hierarchical methods: (a) They construct clusters in a treelike system that provides knowledge of both specific and general clusters and of their relative relations. (This is analogous to first- and second-order factor analysis.) (b) Because the results appear in a tree with successive groupings forming $n - 1$ to 1 final clusters, the number of clusters for a final solution is not specified. On the plus side, this means that the researcher can cut the tree at various cross-sections to have alternate classifications with different numbers of groups. On the negative side, it is often not clear how many clusters should be selected. This problem of determining the number of final groups can be the most problematic one in cluster analysis. Although there are various rules of thumb that often work well in practice, this problem is a limitation that has not been fully resolved by methodologists (cf. Everitt, 1979; Milligan & Cooper, 1985; Mojena, 1977). (c) Most of the hierarchical methods are *noniterative*. Once a grouping has occurred at a given step it is fixed and not reconfigured at a later step. That means that the later grouping may not be completely optimal, as defined by the clustering algorithm at a specific step.

Hierarchical methods are much more commonly used than nonhierarchical methods (Blashfield, 1976; Day, 1986; Romesburg, 1984). Romesburg's bibliographic retrieval of several thousand articles on cluster analysis in the last decade "showed that applications of hierarchical cluster analysis outnumber applications of nonhierarchical cluster analysis by more than ten to one" (p. 3). Among the reasons for the more extensive use of hierarchical cluster analysis are the added information obtained through hierarchical methods, the fit of hierarchical methods to the goals of many cluster analysts, and the wider availability and tradition of use of hierarchical methods.

The most commonly used agglomerative hierarchical techniques are single linkage, complete linkage, average linkage, and Ward's (1963) minimum variance technique. The first three of these techniques were first introduced in biology by the numerical taxonomists (Aldenderfer & Blashfield, 1984; Sokal & Sneath, 1963). Each method differs in the way that proximity matrix information is used at the step of combining groups to form a new group. Average linkage and Ward's method are generally considered the best of these methods. Readers are referred to Aldenderfer and Blashfield (1984) and Lorr (1983) for more information about these methods, which we briefly describe below.

Single linkage. As the first of the modern clustering methods, single linkage was introduced independently by Sneath (1957) in biology and McQuitty (1957) in psychology. It is also called the *nearest neighbor method* (Lance & Williams, 1967) or Johnson's (1967) *minimum method.* Like other agglomerative clustering methods, it entails an algorithm (computing rule) that sequentially searches the proximity matrix and builds clusters based on some definition of nearness of the objects. The single linkage rule, as stated by Aldenderfer and Blashfield (1984), is that "cases will be joined to existing clusters if at least one of the members of the existing cluster is of the same level of similarity as the case under consideration for inclusion. Connections are thus based solely upon single links between cases and clusters" (p. 38). The method does no averaging of within-cluster distances, but true to its name of single linkage, merges clusters according to the shortest link between them at each stage. It thus tends to link clusters to "nearest neighbors" at the edge of the existing cluster. This results in the major drawback of the method, namely its tendency to chain and to put the majority of objects in one elongated cluster. Because of this chaining artifact, the single linkage method does not usually capture the natural structure of data, so it cannot be recommended for general use.

Complete linkage. Sokal and Michener's (1958) complete linkage method, which takes the opposite approach to single linkage, is also called the *furthest-neighbor method* (Lance & Williams, 1967) and the *maximum method* (Johnson, 1967). Here the grouping rule specifies that any new candidate for addition to a cluster must be within a certain level of nearness to *all* members of that existing cluster. In effect, this means that the distance between clusters is defined as the distance between their most distant pair of objects. This is a stringent rule that tends to impose clusters that are dense and spherical and composed of objects that are highly similar. Because clusters in real data often do not have this very orderly natural structure, the complete linkage method has often not performed well in simulation studies of its capacity to detect known clusters. Thus, it cannot be recommended for most clustering applications.

Average linkage. Average linkage is also referred to as the group average method or UPGMA (unweighted pair-group method using arithmetic averages). The method has had wide general use, and is the most frequently used method in biology (Lorr, 1983, p. 88). Only recently has it been more widely used in the behavioral sciences (Aldenderfer & Blashfield, 1984, p. 41). Its computational strategy was developed by Sokal and Michener (1958) as a compromise between the single linkage and complete linkage methods, and thus it minimizes the biases of both. It does so by computing the average similarity of an object under consideration with all the objects currently in a cluster. Then the next linkage in clustering is formed from the pair of candidates with the lowest average similarity. (There are some variations in the literature in precisely how this average similarity is calculated; thus, one should be alert to the fact that there are some variations in computing methods.) In the most recent comparative studies this method has performed as well or better than alternative methods. It is thus one of the methods to be given strong consideration when one chooses a clustering method.

Ward's minimum variance method. Ward's (1963) method has been widely used; the bibliography on clustering prepared by the Classification Society shows 55 citations of Ward's method for 1985 alone (Day, 1986). The method has been especially widely used in the behavioral sciences, in which it was introduced in the 1960s (Ward & Hook, 1963; Veldman, 1967). This is the method we used in our example at the beginning of this article. Applied to a proximity matrix of Euclidean distances (d^2), Ward's algorithm is designed to minimize the variance within clusters at each stage of grouping. The approach proceeds by merging those single objects or groups of objects that result in the least increase in the within-groups sum of squares (or error sum of squares). Although this approach optimizes within-cluster homogeneity at each stage of grouping, it does not ensure optimum homogeneity of the final clusters, because once objects have been joined they are not separated at later stages of grouping. This does not, however, constitute a major practical difficulty with Ward's method.

Comparative studies now suggest that Ward's method is one of the more effective methods for recovering underlying structure, being equivalent or slightly inferior to average linkage. Recent research has pointed to possible biases in the method to which the researcher should be alert. Because the method is usually applied only with squared Euclidean distance as proximity measure, it tends to produce clusters that are heavily influenced by level differences. If the researcher wants to avoid this, the data should be standardized to remove level differences before Ward's method is applied. The method is also biased toward forming spherical clusters in multivariate space; it may not be the method of choice in situations in which the natural clusters are likely to be elongated or otherwise oddly shaped. Finally, there is evidence that the accuracy of the method falls off in the presence of statistical outliers, that is, a few cases of large distances from the general cluster space. It is recommended that data be prescreened for outliers (cf. Comrey, 1985) prior to routine use of Ward's method.

Other Clustering Methods

One of the major nonhierarchical methods is the *k*-means iterative partitioning method. This approach requires the user to specify the expected number of clusters for the program. On the basis of this initial "seed" information, the method calculates centroids for a set of trial clusters, places each object in the cluster with the nearest centroid, and then recalculates the centroids and reallocates the objects. This process iterates until there are no changes in cluster membership. In recent evaluative studies the *k*-means procedure has performed well if fairly accurate information was provided about the correct number of groups. Thus, it should be considered a method of choice when the context of clustering provides reliable a priori information about the number of clusters and when a nonhierarchical solution is desired. When such information is lacking, Milligan's (1980) work suggests that a good approach with the *k*-means method is to first analyze the data with the average linkage method to obtain an appropriate starting point for the number of clusters. The reader is referred to Aldenderfer and Blashfield (1984, pp. 45–49) for a more complete discussion of the *k*-means procedure. They also present a good summary of the additional varieties of clustering methods (pp. 50–53), including density search, clumping, and graph theoretic methods. These are newer methods, and although promising, they have not been widely used in the behavioral sciences.

Grouping by Inverse Factor Analysis

When used in one specialized way, factor analysis can, in principle, be used to form clusters (e.g., Fleiss, Lawlor, Platman, & Fieve, 1971). This grouping of objects by factor analysis has a long history, especially in psychology. When used for this specialized purpose it is called *inverse* or *transposed factor analysis* to denote that the usual data matrix is inverted (transposed) so that correlations are calculated between pairs of people (or other objects) across the set of variables. (This approach is the opposite of conventional factor analysis, in which the correlations are between pairs of variables.) Following Cattell's (1966) usage, inverse factor analysis is also referred to as Q-type factor analysis. When this correlation matrix is factored, the factors represent types of people (or other objects). Then to assign the people discretely to groups the researcher must examine the factor loadings and make assignments on the basis of the highest loadings. This step can constitute the greatest drawback of inverse factor analysis. If a person loads about equally on two factors, then the researcher must be somewhat subjective or arbitrary in assigning the groups. Moreover, this process of examining the loadings can be extremely tedious when the clustering problem is large. Because of these limitations, many modern writers are not enthusiastic about using inverse factor analysis for clustering purposes (e.g., Everitt, 1980). Also, because subjectivity is necessary for inverse factor analysis, it has not been included in the large computer-based Monte Carlo studies of grouping methods discussed in the next section. Therefore, we do not have a large data base for evaluating the accuracy of the method in uncovering underlying group structure. Aldenderfer and Blashfield (1984, p. 49) and Everitt (1980, pp. 72–73) discuss other potential problems in the use of inverse factor analysis.

Comparative Evaluations of Cluster Methods

In the past decade, psychologists have been at the forefront of increasingly more sophisticated evaluations of major clustering methods. The ultimate purpose of these studies has been to determine how effective a method is for identifying the natural structure in data. Does the method adequately detect groups when they are present?

The common approach in these evaluations has been to construct artificial data sets with known group structure and some random error, and then to test the method's capacity to detect or recover these known groups. Several studies by psychologists using this strategy of Monte Carlo simulation have recently been published, notably in *Multivariate Behavioral Research* and *Psychometrika*. Readers who wish to keep current with ongoing evaluations of clustering methods will want to consult journals such as these, which are likely to publish studies in the future.

For this fairly extensive and evolving literature in which clustering methods are compared, here is our synopsis of the current consensus on the most recommended clustering methods. Although the Monte Carlo data and expert opinion provide no clear consensus that one clustering method is always best, some general guidance can now be given from the dozen or more good Monte Carlo studies of the past decade. Specifically recommended hierarchical methods are Ward's method and average linkage (cf. Blashfield, 1984, p.

255). Using both of these methods with the same data set would be a good strategy. If similar clusters are obtained, good evidence would be available for the cross-method stability of the clusters. If the researcher can specify the probable number of clusters in advance, then it is recommended that the *k*-means method also be applied (cf. Blashfield, 1984, p. 255).

In the remainder of this section, we mention some of the historical and technical highlights in the simulation literature that evaluate clustering methods. We believe this information will be helpful to readers who wish to follow future studies of this kind. Readers with little interest in these more technical details may wish to skip over the remainder of this section.

The first consideration about such simulations is that all common clustering techniques will recover the cluster structure if the number of groups is small and the underlying group separation is very clearcut. Therefore, the progressive focus of these comparative evaluations has been to evaluate the performance of the methods when various sources of noise (e.g., random error in data) or cluster structure are introduced into the data. The intent has been to simulate the varieties of atypicality likely to occur with real psychological data. Among these sources of noise are overlap among underlying groups, sampling error, number of groups, unusually shaped groups, presence of statistical outliers, and so forth. Additionally, there has been a focus on the effects of using alternate measures of proximity.

Providing the most thorough review of the Monte Carlo validation studies of cluster methods, Milligan (1981) showed that the foci and conclusions of these studies can be divided into three chronological periods: pre-1975, 1975–1978, and 1979–1981. As the sophistication and complexity of the simulation designs have increased over these periods, Milligan's chronological review shows how the appraisal of the major methods has been reassessed. In the early phase, Ward's method appeared to be distinctly superior (cf. Blashfield, 1976). With methodological refinements and a greater diversity of data types, the group average method has emerged as at least equivalent to Ward's method (Aldenderfer & Blashfield, 1984; Blashfield, 1984; Edelbrock, 1979). Finally, the nonhierarchical *k*-means method has also emerged as a strong contender, along with Ward's and the group average methods (Milligan, 1980). Good recent surveys of these empirical studies are also provided in Lorr (1983, chapter 7); Blashfield (1984, chapter 8); and Aldenderfer and Blashfield (1984, pp. 59–61). A synopsis of salient conclusions from these comparative studies follows.

The first comparative evaluation to have major visibility and impact, especially in psychology, was Blashfield's (1976). He created 50 artificial data sets with known clusters and applied four of the major hierarchical grouping methods. One measure of accuracy commonly used in cluster research is the kappa statistic (Cohen, 1960). It ranges from zero to one, with the higher values representing greater accuracy of agreement. The order of accuracy for these four methods, from high to low, was: Ward's minimum-variance method (median kappa of .77); complete linkage (.42 kappa); average linkage (.17 kappa); and single linkage (.06 kappa). Subsequently, Edelbrock (1979) pointed out that Blashfield's (1976) and other Monte Carlo studies were limited by the way they required all subjects to be grouped. Edelbrock introduced a methodological refinement to handle this limitation. With this refinement, the average-linkage method performed well in Edelbrock's reanalysis of Blashfield's (1976) data set.

In a definitive article, Milligan (1980) published an extensive Monte Carlo study with a larger range of data conditions and 11 different clustering methods. Blashfield (1984) has characterized this as "the best design of any study to date" (p. 253). Milligan confirmed Edelbrock's (1979) demonstration of the good performance of the average linkage method. In addition, he showed that the nonhierarchical *k*-means iterative partitioning method performed well over a range of conditions. This method requires the researcher to specify the probable number of groups in advance, so its use is limited to special cases in which the researcher has considerable a priori understanding of the problem.

In our opinion, there is something inherently misleading about the apparent results of the recent comparisons of proximity measures. The trend of recent studies has been to suggest that choice of proximity measure has relatively little effect on accuracy of cluster recovery (e.g., Milligan, 1980; Scheibler & Schneider, 1985). As Aldenderfer and Blashfield (1984) put it, "these effects seem to be *swamped* by factors of cluster structure, degree of coverage required, and overlap of clusters" (p. 61, italics added). We think that this statement pinpoints the essential problem in increasingly complex Monte Carlo studies with wide varieties of data. They show in theory the performance of the methods for a wide range of data. But in practice, the researcher has just one data set. The methods selected either are or are not capable of detecting the particular underlying structure. The researcher still has a choice about the kind of proximity method to be selected. Because it is a mathematical truism that a given proximity measure either reflects or does not reflect the separate elements of data elevation, scatter, or shape, the choice of proximity measure for a specific data set must make a difference. Thus, correlation cannot reflect data elevation. If elevation differences exist in the data and elevation is of potential importance to the researcher, then there is no justification for using correlation. Similarly, squared Euclidean distance is known to tap all the data features of elevation, scatter, and shape. If the researcher does not want elevation to be present in the resulting clusters, then squared Euclidean distance is an inappropriate choice of proximity measure. This is a case where our heads should prevail and we should begin to recognize the limits for generalizing from large Monte Carlo studies, despite their apparent scope. As they have been used to probe more deeply, their theoretical scope has become so large that the immediate practical implications for a researcher with a specific problem may have become obscured.

Recent Clustering Applications in Counseling Research

Although the overall use of clustering in counseling research is limited, two notable examples of its use were reported in recent issues of the *Journal of Counseling Psychology*. In evaluating these examples, we focus on how successfully clustering has been used to advance knowledge.

Elliott (1985)

The most prominent and potentially important example is Elliott's (1985) use of clustering to form an empirical taxonomy of helpful and nonhelpful events in counseling, as perceived by clients. In a counseling simulation, students acting as counseling volunteers identified and described 86 helpful and 70 nonhelpful counselor responses. Judges then rated the similarity of these responses to yield similarity matrices for the helpful and the nonhelpful events. Average linkage cluster analysis was applied to these similarity matrices. The helpful events were grouped hierarchically in two superclusters related to task and interpersonal aspects of counseling. The largest cluster within task orientation was new perspective, and the largest interpersonal cluster was understanding. Six types of nonhelpful events were identified.

Elliott (1985) followed two important recommendations of Blashfield (1980) to verify the robustness of the clustering results. First, stability of the results across methods was checked with an alternate clustering method (maximum linkage). Second, validity of the clusters was evaluated through associations of the clusters with counselor action variables.

Elliott's (1985) work is important as an empirical effort to classify counseling events in terms of their immediate therapeutic impact. His central question was "What types of events matter in counseling?" (p. 307). Clustering was appropriately applied to this classification task. The raw data base was one with inherent counseling significance, notably students' perceptions of counseling events and trained judges' perceptions of the similarity of these events. Initial steps were taken to verify the stability and validity of the clusters. Elliott acknowledges that the taxonomy should be considered tentative, for a variety of methodological reasons. For example, "the categories obtained are a function of both the judges who did the sorting and the clustering procedures used" (p. 319). We will make the best advances in our science by paying attention to the exploratory aspects of Elliott's work. His taxonomy should not be taken as a given, without further testing, honing, and replication.

Berven (1985)

Berven (1985) studied the use of case management simulations with counselors. As one part of his analysis, he clustered the counselors into three performance groups by applying Ward's method to six performance measures. Subgroups were clearly differentiated by performance styles (see his Figure 1), and were also significantly differentiated by experience level. This is a straightforward example of an appropriate use of cluster analysis that effectively reduced a multivariate data set into a meaningful set of subgroups. The graphic presentation of the subgroup performance profiles is helpful in showing the group differences, which are based both on performance level and profile shape. Such differences would probably not have been discovered by use of correlation as a measure of similarity.

Clustering as a Tool

At this stage, the exploitation of clustering as a valuable resource in the psychological researcher's toolbox has just begun. Because the technique is now accessible to most researchers, their task is to familiarize themselves with this method. Despite the need for caution in trying it out on any data set at hand, performing a clustering procedure on extant data sets is a good way for researchers to explore this technique. As researchers grapple with issues such as what could be learned by clustering these data (Is this exploratory, testing a prior classification scheme, or simplification?), the aspects of the data that are of most interest (Do I need to consider elevation, size, or shape, or all of them when choosing a measure of proximity?), and how these data might be analyzed (With which method or methods?), the concepts outlined in this article may make more sense or seem more relevant. After researchers have developed confidence by becoming familiar with this approach, we hope that they will consider new questions about and new uses for clustering.

References

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis.* Beverly Hills, CA: Sage.

Berven, N. L. (1985). Reliability and validity of standardized case management simulations. *Journal of Counseling Psychology, 32,* 397–409.

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

Blashfield, R. K. (1980). Propositions regarding the use of cluster analysis in clinical research. *Journal of Consulting and Clinical Psychology, 48,* 456–459.

Blashfield, R. K. (1984). *The classification of psychopathology: Neo-Kraepelinian and quantitative approaches.* New York: Plenum.

Borgen, F. H., & Weiss, D. J. (1971). Cluster analysis and counseling research. *Journal of Counseling Psychology, 18,* 583–591.

Cattell, R. B. (1966). The meaning and strategic use of factor analysis. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology.* Chicago: Rand McNally.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Comrey, A. L. (1985). A method for removing outliers to improve factor analytic results. *Multivariate Behavioral Research, 20,* 273–281.

Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate procedures for the behavioral sciences.* New York: Wiley.

Cronbach, L., & Gleser, G. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50,* 456–473.

Day, W. H. E. (Ed.). (1986). *Classification literature automated search service, 1985, Vol. 14.* Classification Society of North America.

Dixon, W. J. (Ed.). (1983). *BMDP statistical software.* Berkeley: University of California Press.

Edelbrock, C. (1979). Comparing the accuracy of hierarchical clustering algorithms: The problem of classifying everybody. *Multivariate Behavioral Research, 14,* 367–384.

Elliott, R. (1985). Helpful and nonhelpful events in brief counseling interviews: An empirical taxonomy. *Journal of Counseling Psy-*

chology, 32, 307–322.

Everitt, B. (1979). Unresolved problems in cluster analysis. *Biometrics, 35*, 169–181.

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

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

Gangestad, S., & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review, 92*, 317–349.

Hansen, J. C. (1986). Strong Vocational Interest Blank/Strong-Campbell Interest Inventory. In W. B. Walsh & S. H. Osipow (Eds.), *Advances in vocational psychology. Volume I: The assessment of interests* (pp. 1–29). Hillsdale, NJ: Erlbaum.

Hartigan, J. (1975). *Clustering algorithms.* New York: Wiley.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32*, 241–254.

Lance, G. N., & Williams, W. T. (1967). A general theory of classificatory sorting strategies. I: Hierarchical systems. *Computer Journal, 9*, 373–380.

Lorr, M. (1982). On the use of cluster analytic techniques. *Journal of Clinical Psychology, 33*, 461–462.

Lorr, M. (1983). *Cluster analysis for the social sciences.* San Francisco: Jossey-Bass.

Massart, D. L., & Kaufman, L. (1983). *The interpretation of analytical chemical data by the use of cluster analysis.* New York: Wiley.

McQuitty, L. L. (1957). Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. *Educational and Psychological Measurement, 17*, 207–229.

Megargee, E. I., & Bohn, M. J. (1979). *Classifying criminal offenders: A new system based on the MMPI.* Beverly Hills, CA: Sage.

Mezzich, J., & Solomon, H. (1980). *Taxonomy and behavioral science: Comparative performance of grouping methods.* New York: Academic Press.

Milligan, G. W. (1980). An examination of the effects of six types of error perturbation on fifteen clustering algorithms. *Psychometrika, 45*, 325–342.

Milligan, G. W. (1981). A review of Monte Carlo tests of cluster analysis. *Multivariate Behavioral Research, 16*, 379–407.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50*, 159–179.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: An evaluation. *Computer Journal, 20*, 359–363.

Norusis, M. J. (1985). *SPSS^X advanced statistics guide.* New York: McGraw-Hill.

Overall, J., & Klett, C. (1972). *Applied multivariate analysis.* New York: McGraw-Hill.

Paykel, E. S. (1971). Classification of depressed patients: A cluster analysis derived grouping. *British Journal of Psychiatry, 118*, 275–288.

Paykel, E. S. (1972). Depressive typologies and response to amitriptyline. *British Journal of Psychiatry, 120*, 147–156.

Romesburg, H. C. (1984). *Cluster analysis for researchers.* Belmont, CA: Lifetime Learning Publications.

SAS Institute. (1982). *SAS user's guide: Statistics.* New York: Author.

Scheibler, D., & Schneider, W. (1985). Monte Carlo tests of the accuracy of cluster analysis algorithms: A comparison of hierarchical and nonhierarchical methods. *Multivariate Behavioral Research, 20*, 283–304.

Skinner, H. (1978). Differentiating the contribution of elevation, scatter, and shape in profile similarity. *Educational and Psychological Measurement, 38*, 297–308.

Skinner, H. A., & Blashfield, R. K. (1982). Increasing the impact of cluster analysis research: The case of psychiatric classification. *Journal of Consulting and Clinical Psychology, 50*, 727–735.

Sneath, P. H. A. (1957). The application of computers to taxonomy. *Journal of General Microbiology, 17*, 201–226.

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin, 38*, 1409–1438.

Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of numerical taxonomy.* San Francisco: W. H. Freeman.

SPSS, Inc. (1986). *SPSS^X user's guide, Edition 2.* New York: McGraw-Hill.

Veldman, D. J. (1967). *Fortran programming for the behavioral sciences.* New York: Holt, Rinehart & Winston.

Wampold, B. E., & White, T. B. (1985). Research themes in counseling psychology: A cluster analysis of citations in the process and outcomes section of the *Journal of Counseling Psychology. Journal of Counseling Psychology, 32*, 123–126.

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

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

Zytowski, D. G., & Hay, R. (1984). Do birds of a feather flock together? A test of the similarities within and the differences between five occupations. *Journal of Vocational Behavior, 24*, 242–248.

Received May 26, 1986
Revision received March 12, 1987
Accepted March 12, 1987 ■