After a brief review of organization classification, multivariate approaches toward the empirical identification of such classifications are discussed. Ten guidelines for conducting future multivariate classificatory studies are presented, supported by a critique of two published large-scale empirical studies. Taxonomic development is viewed as a critical element in the future health of organization science.

A fundamental element in the development of a scientific body of knowledge is the availability of a widely accepted and usable classification scheme (Hempel, 1965; Haas and Drabek, 1973). Existing organizational classifications are not comprehensive enough to classify organizations into scientifically useful groupings, primarily because they ignore many important attributes. Multivariate analytical approaches purported to derive classificatory schemes empirically have been suggested as alternatives to the mental inductive method used in the past by organizational classification theorists. But the usefulness and validity of these methods in aiding the induction of classificatory schemes depend on how unbiased they are by the limitations of the already existing classifications. This article sets forth some guidelines for using multivariate analyses to aid the development of classifications and to evaluate the results of examples of such usage by Haas, Hall, and Johnson (1966), Pugh, Hickson, Hinings, McDonald, Turner, and Lupton (1963), Pugh, Hickson, Hinings, and Turner (1968), and Pugh, Hickson, and Hinings (1969).

THEORETICAL BACKGROUND

Organizational Classification

The study of organizational classification is at such a primitive stage that there is not even agreement about terms, let alone agreement about a theory of classification. Haas, Hall, and Johnson (1966) and Pugh, Hickson, and Hinings (1969) used the term ‘‘taxonomy’’ to characterize the results of empirical multivariate—inductive—analyses, reserving the term ‘‘typology’’ to characterize the deductive or intuitive approaches. Marney and Smith (1964), Eilon (1968), Laufer (1968), and Sagasti (1970) on the other hand used ‘‘taxonomy’’ in reference to the deductive approaches. In addition Deutsch (1966) and Burns (1967) used ‘‘taxonomy’’ broadly to include any kind of classification or typology. Vickers (1970) used ‘‘classification’’ to describe deductively developed groupings, but others such as Gordon and Babchuk (1959), Etzioni (1961), Blau and Scott (1962), Katz and Kahn (1966), and Perrow (1967) employed the term ‘‘typology.’’ Samuel and Mannheim (1970) considered the terms ‘‘classification’’ and ‘‘typology’’ to be interchangeable. The semantic confusion is an indication that little systematic development of organizational taxonomy and classification methodology has taken place over the past 15 years.

To avoid further confusion, Hempel’s (1965) usage is followed here: the term ‘‘taxonomy’’ is reserved to apply to the theory of classification and his definitions of type concepts, which are concepts for arranging phenomena into categories, are used. A classificatory type concept is an either-or notion, with
objects classified in a given category or not in it. Thus, many objects are cats or not cats, metals or not metals. An extreme type concept is the formation of categories by labelling the poles or extremes of orderings or rankings. Organizations are thus categorized as formal or informal, even though formalization appears in many gradations. An ideal type concept, such as Weber’s (1947) concept of bureaucracy, is used primarily for theory generation with little expectation that objects meeting all criteria of the ideal type will be found empirically.

The classificatory type concept is the most useful one for the scientific activity of inducing theories about certain phenomena and deducing testable hypotheses. It provides the basis for explanation, prediction, and scientific understanding by identifying uniformities in the phenomena about which laws and principles may be formed (Hempel, 1965: 146); identifies conditions under which hypotheses might be expected to be valid (Haas, Hall, and Johnson 1966: 158); and aids in the handling of complex sets of variables or attributes by identifying a more parsimonious set of constructs (Mechanic, 1963: 158). The extreme and ideal type concepts were developed by social scientists in an attempt to develop classifications of phenomena not falling neatly into either-or categories. The extreme type concept is difficult to use, either when the ordering is normally distributed, because most cases fall between the extremes, or when the ordering is rectangularly distributed, because then several categories are needed in addition to the two extremes to obtain reasonably homogeneous groupings, and this leads to a less parsimonious classification. The ideal type cannot be used in empirical research because it results in theoretical categories not usually found empirically.

Parsimony of classification is achieved when a scheme contains as few nonoverlapping classes or orthogonal dimensions as is possible. Parsimonious classification of organizations is especially difficult because there are so many important attributes and because most of these attributes are ordering rather than either-or concepts. Sells (1964) suggested as many as 500 variables might be necessary to describe organizations fully. Stogdill (1966) used some 68 categories in his theoretical model, while Haas, Hall, and Johnson (1966) identified 210 variables and Pugh, Hickson, Hinings, and Turner (1968) used 64 variables in their empirical investigations. Given the large number of potential classificatory variables, there is no escaping the need for a theory which suggests a much smaller set to be used in classification. Without criteria for identifying this subset of variables, classifications are apt to include an unmanageable number of classes with too few organizations in each class. The effect of the many ordering type variables is to magnify the problem of too many classes, because each ordering variable may result in several classes.

Until now, classification theorists have overemphasized parsimony at the expense of scientific usefulness. Brown (1945), Parsons (1956), Etzioni (1961), Blau and Scott (1962), Katz and Kahn (1966), Perrow (1967), Thompson (1967), and Vickers (1970) have suggested very simple classificatory schemes based on one or two attribute dimensions, without first working out an underlying theory identifying a comprehensive set of classificatory variables. Consequently, their schemes have
Guidelines

not provided the kinds of classifications useful in empirical research (Burns, 1967; Hall, Haas, and Johnson, 1967). Theorists have tended to test their taxonomies deductively by seeing whether the classes identified by the theory appeared in empirical settings. The problem is that there are countless ways of grouping organizations, and therefore almost any taxonomy can be supported by pointing out these groupings. Taxonomies can be tested best by seeing if they produce homogeneous groupings from which theories about the behavior of organizations can be formed, and with which hypotheses about such behavior can be tested.

Multivariate Approaches

The development of taxonomy is no different than the inductive process used to create any general theory—the theoretician attempts to set forth broadly applicable principles and laws to explain the behavior of the empirical objects observed. Merton (1957) and Glaser and Strauss (1967) have argued that theories of the middle range and theories well grounded in empirical reality are likely to be the most fruitful kind, as opposed to very narrow or very grand theories. Glaser and Strauss (1967) presented an extended discussion showing how theory development may be grounded in either qualitative or quantitative data. The grounding of taxonomy requires some means of identifying possible categories and possible classificatory variables in a way that keeps the discovery of the empirical categories independent from preconceived ideas about taxonomy and classification.

Investigators such as Haas, Hall, and Johnson (1966), the Ohio group, and Pugh, Hickson, Hinings, and Turner (1969), the Aston group, have attempted to deal with the problem of inadequate classes by using multivariate analyses. These analyses did not result directly in taxonomic theory, they only resulted in a means of forming possible groupings, according to certain attributes that could be measured empirically. They were not substitutes for the mental inductive process, but instead were methods of improving the empirical grounding of theory generation.

Both the Ohio and Aston groups began their studies with the observation that multivariate approaches offered a chance to come up with a fresh set of organizational groupings, based on larger samples than had been used before, and free from the influence of the existing, rather limited taxonomies (Haas, Hall, and Johnson, 1966: 161; Pugh, Hickson, and Hinings, 1969: 115). The idea was to avoid what Schutz (1953) called prior scientific or common-sense conceptions of reality. Many methodologists would argue that it is impossible to collect data so raw as to be free from the influence of previous conceptions. In multivariate approaches based on questionnaires, such as those used by the Ohio and Aston groups, it is especially difficult to make any claim that one’s data are truly raw. All sorts of conceptions typically influence the development of questionnaires.

It is impossible to avoid all influence of previously held conceptions. Even the decision to use a five- rather than a seven-point response scale imposes some kind of prior conception on the resulting data. But there are ways to avoid the nar-
rower influence of previous taxonomies in the construction of questionnaires and in the multivariate analyses themselves.

In the following section 10 guidelines are suggested which, if followed, have the effect of keeping multivariate analyses free from the influence of previous taxonomies. Some guidelines may be unreasonable. Nevertheless, they are choice-points that need to be identified. If the investigator chooses not to follow a particular guideline, the results will be influenced by previous taxonomic theory. In many cases this is acceptable, as long as the results are suitably qualified. The use of the guidelines will encourage investigators to be much more self-conscious about critical decisions they make.

Even with the use of multivariate analyses, most investigators need to reduce the scope of their studies because of limitations on funding and access to organizations. The guidelines most apt to be considered unreasonable are the first 6, which suggest the use of random sampling to narrow a study, rather than the use of previous classifications. Since the empirical development of classes is not likely to be a one-time-only affair, successive random samplings can be expected to fill in whatever gaps may have occurred in the initial sampling.

GUIDELINES WITH AN EVALUATION OF THE ASTON AND OHIO STUDIES

There are several reasons for presenting the guidelines in the context of an evaluation of the Aston and Ohio studies. The presentation will be less abstract and more lively, and the places where the guidelines affect multivariate studies will be more apparent. In addition, since the results of these two studies have become increasingly visible as a result of extensive publication, reprinting, discussion in texts, and use in succeeding studies, it is appropriate to evaluate their results in light of the original purposes of the investigators and the general concern for high quality knowledge.

Guideline 1: Define the broadest possible population of organizations or, if a delimitation is unavoidable, base it on a significant cultural unit. Nothing is so basic as the need to identify clearly and define the population of organizations a classification is to include. Prerequisite to the population definition is a definition distinguishing organizations from other phenomena. In the absence of a widely agreed upon technical or scientific definition (Hall, 1972: 5–9), a conservative approach is to take as organizations any object that nonscientists think of as organizations, a commonsense approach similar to that of March and Simon (1958: 1). Given that organizations are artificial devices (Etzioni, 1964: 58–59) whose form emerges from conceptions held by their owners, managers, and members, the cultural patterns affecting these conceptions are important, and therefore boundaries separating one cultural form from another become significant demarcations in organizational form. Inasmuch as the interaction between scientist and practical manager most likely is highest within cultural boundaries, especially if bolstered by geographical and language boundaries, the development of taxonomy by cultural units is in keeping with the practical goals of science as well as those of intellectual curiosity (Hempel, 1965). Most important, the cultural demarcation allows a re-
Guidelines

duction in the scope of classificatory studies without introducing the influence of previous thinking about classes.

The Aston study limited the population it was concerned with to work organizations (Pugh, Hickson, Hinings, McDonald, Turner, and Lupton, 1963). According to the sample actually drawn, work organizations were “firms making motor cars and chocolate bars, municipal departments repairing roads, and teaching arithmetic, large retail stores, small insurance companies,” but not voluntary organizations, since they did not pay their members (Pugh, Hickson, Hinings, and Turner, 1968: 67). Other delimitations were that only organizations in the Birmingham, England, area and only organizations with 250 or more employees were included. Child (1972) pointed out that 20 of the 46 units sampled were not independent organizations, but instead were subsystems of other organizations, which led him to question some of their conclusions. The implications of this for classificatory purposes, however, are not clear. What is clear is that the Aston sample was not drawn from a population defined in a way so as to be free from previous classificatory concepts, such as kind of output, size, and employee relationship.

Did the population delimitation coincide with a culturally important boundary? Possibly, although the Aston group did not so indicate and the Birmingham area is not obviously a culturally significant unit to many non-Britishers. Can a classification based on the Birmingham area reasonably be generalized to other parts of England, the United States, or non-English-speaking parts of the world? Again, it is possible, but no data or arguments were offered. Given the questionable generalizability of the study and the questionable worldwide significance of the Birmingham area, even if it was a cultural unit, it is disconcerting to find that none of the results in Pugh, Hickson, Hinings, and Turner (1968) or in Pugh, Hickson, and Hinings (1969) were properly qualified.

The population defined by the Ohio group was limited to a cultural area coincident with the boundaries of the United States, by any reckoning a significant cultural area. Within this cultural unit, the defined population was inclusive. The Ohio study’s population definition was free from any influence of previous classification concepts. From the perspective of the worldwide community of scientists, the Ohio study’s population definition encompassed a more significant cultural area and within that cultural area defined a more inclusive population than did the Aston study. The Ohio study thus meets guideline 1, whereas the Aston study does not.

Guideline 2: Use a probability sampling plan without any stratification for selecting a sample of organizations. The reason for insisting on a probability sampling plan is to assure that all members of the population have an equal probability of being selected. This reasoning is sound practice no matter what kind of study one has in mind, but it is especially important if the influence of previous classificatory thinking is to be avoided. Any decision to draw a stratified random sample is, of course, based on prior thinking about what the significant strata are and about the possible consequences of leaving some members of the population undersampled and thus underrepresented.
The Aston group drew a “random sample stratified by size and product or purpose” (Pugh, Hickson, Hinings, and Turner, 1968: 67). It ended up with a sample of 46 out of 293 large employing units in the Birmingham area. The Ohio group’s sample was also stratified by size, structure, goals, and type of organizational activity. It did not actually draw a random sample since it claimed that no complete listing of organizations in the United States was possible. Instead, it chose cases representative of several points along one or more attribute dimensions.

Given the taxonomic objectives of both studies, the samples’ stratifications were more of a problem than their lack of representativeness. The Aston sample was representative, but at the possible expense of an overly narrow population definition. It is likely that the Ohio group’s sample of 75 out of perhaps a million or more organizations in the United States was not representative, the result of too ambitious a population definition. By choosing to stratify their samples, both groups presumably argued that some attributes were more important than others for classificatory purposes and, further, that it was important to make sure the full ranges of these dimensions were represented in their samples. Neither group could decide on which dimensions to base their stratifications without thinking in terms of previous classifications. They both chose size and goals—the Ohio group also included structure and work activity—all elements of previous classificatory thinking. It is a question why other dimensions were not included, dimensions available around 1962 when the two groups started collecting data, such as age and complexity (Anderson and Warkov, 1961), compliance (Etzioni, 1961), technology (Woodward, 1958), or managerial beliefs (McGregor, 1960). In the absence of a widely accepted classification scheme, their stratification decisions were quite arbitrary. Given that the purpose of their studies was to identify dimensions that could later be used in stratified sampling plans, the use of stratification plans in generating a taxonomy was self-defeating. Neither study thus satisfies guideline 2.

Guideline 3: Define as inclusive a population of organizational attributes as is possible. Decisions to place organizations in the same or different classes depend in large part on what attributes are considered. A college and a chemical manufacturer could be in the same class if only size was considered, but few observers would classify them the same if only technology were considered. Because of this, the effects of prior thinking about classifications of attributes can have a definite influence on the classification of organizations. To avoid this influence, investigators must avoid using previous classifications of attributes. Because multivariate methods require inputs from structured interviews or questionnaires and thus a decision about attributes is unavoidable, the only recourse is to define the population of attributes to include all known organizational attributes considered to be properties of the collectivity rather than of individual members or subunits (see Lazarsfeld and Menzel, 1969, for a distinction between collective and individual properties).

The Aston investigators defined their population attributes in terms of seven dimensions: specialization, standardization, formalization, flexibility, centralization, configuration—made
up of several subdivisions, and traditionalism, which they said were drawn from that portion of the literature saturated with the Weberian view of bureaucracy (Pugh, Hickson, Hinings, McDonald, Turner, and Lupton, 1963; Hinings, Pugh, Hickson, and Turner, 1967; Pugh, Hickson, Hinings, and Turner, 1968). What is their rationale for such a narrow population definition? They started by saying they wished to investigate the interdependence of three distinct levels of analysis: “(1) organizational structure and functioning, (2) group composition and interaction, and (3) individual personality and behavior” (Pugh, Hickson, Hinings, McDonald, Turner, and Lupton, 1963: 292).

Initially, they chose to limit themselves to analyzing structure and functioning. This is an example of how a prior classificatory scheme was used to delimit the study. It is possible they thought items (2) and (3) were not organizational attributes, but they did not suggest this, and it would be a hard argument to support. Perhaps such a delimitation in the population of attributes was necessary to reduce the scope of the project to a manageable size, a reasonable design decision. In this instance the investigators did properly qualify their findings by pointing out that theirs was a classification of structures rather than of organizations (Pugh, Hickson, and Hinings, 1969: 115–116).

Even if there was a good reason for limiting the initial attempt to organizational structure and functioning, it is not clear that the Aston group appropriately sampled the population of attributes, even so narrowly defined. It claimed that its sampling of attributes within the Weberian tradition was supported by Evan (1963) and Hage (1965) (Pugh, Hickson, Hinings, and Turner, 1968: 66). This claim was not altogether true. Evan (1963) suggested three dimensions, only two of which were comparable to the dimensions of the Aston study. Evan’s dimension, hierarchy of skills, related to the Aston dimension, specialization, and Evan’s hierarchy of authority related to Aston’s centralization. The third Evan dimension, reward structure, was not included in the Aston study. The Hage (1965) study defined the organizational population of attributes in terms of eight dimensions: complexity, centralization, formalization, stratification, adaptiveness, production, efficiency, and job satisfaction. The only overlap between the Hage and Aston studies included complexity, centralization, and formalization. The Aston group claimed their findings corroborated the Evan and Hage studies and yet, of the 10 distinct variables mentioned by Evan and Hage, only 4 were picked up in the Aston study. There was no explanation why the other 6 were ignored. The seven dimensions the Aston group selected are even narrower when compared to the original 32 attributes Weber suggested (1947: 329–340).

It is also not clear why the Aston group chose to limit itself to the Weberian view of structure and function. Before the group started its investigation, Haire (1959) had published an anthology that contained several suggestions of alternative organizational forms and various attributes of organizations by Bakke, Cyert and March, Argyris, White, Likert, Dubin, and Haire. In light of the population of attributes suggested by these authors, the choice of the Aston group to limit its population of attributes to a subset of those suggested by Weber seems arbitrary.
The Ohio researchers emphasized "official (formal) structure" and "performance structure" (Haas, Hall, and Johnson, 1966: 162). Theirs was a broader population of attributes than the Aston group's, but it was not inclusive when compared to available knowledge at that time (Haire, 1959). The scope of the Ohio study was limited, in part, because the researchers included in their questionnaire only those items of information which could be collected in interviews with top executives. Thus, they were precluded from gathering information about individual and group characteristics and the noninstitutionalized structure of the organization. Since both studies applied classificatory schemes in narrowing their populations of attributes, neither of them meets guideline 3.

**Guideline 4: Use a probability sampling plan for selecting a sample of organizational attributes.** There are a large number of organizational attributes. Because of limited amounts of time, funding, and cooperation from managers, most investigators limit the scope of their study by limiting the number of attributes they measure. A probability sampling plan is the only way to choose a subset of attributes without the intrusion of previous thinking about the classification of attributes. Because most studies are reduced to manageable proportions, organizational scientists have become accustomed to not getting the whole picture all at once; there is an incremental accumulation of results. Guideline 4 proposes, for taxonomic and classification purposes at least, that the increments be based on probability samples rather than samples formed by applying some prior classification of variables.

Because both the Aston and Ohio groups defined narrow populations of attributes, their samples were no smaller than their populations; thus, guideline 4 cannot be applied to their studies.

**Guideline 5: Define the population of observers of organizational attributes to be as inclusive as possible.** Guideline 6: Use a stratified probability sampling plan for selecting observers. A third way prior thinking about organizations may enter into the empirical grounding of taxonomy is through the use of schemes classifying observers into one or another category.

One way to avoid using observers is to concentrate on objective measures, but there is some doubt that what is objective covers all important attributes (Rice and Mitchell, 1973) and, further, it is possible the objective measures do not bear on actual organizational life, but only as it is institutionalized in documentary form (Pugh, Hickson, Hinings, and Turner, 1968). A conservative approach is adopted at the start to avoid decisions that classify measures of attributes as objective or subjective or as direct or requiring the use of observers. Because organization scientists usually have not had an opportunity to measure directly all organizational attributes, they have relied on others as sources of data, most typically top executives (Haas, Hall, and Johnson, 1966; Lawrence and Lorsch, 1967; Pugh, Hickson, Hinings, and Turner, 1968; Blau and Schoenherr, 1971; Child, 1972; Khandwalla, 1973; McMillan, Hickson, Hinings, and Schneck, 1973). Any classification of observers results in a classification of attributes, because all observers do not have an equal opportunity to be
Guidelines

in contact with the same characteristics of organizations, consequently such a classification is to be avoided. The population of observers has included high- and low-ranked members and nonmembers. The boundary between members and nonmembers is a relative concept, depending on the perspective of the scientific observer (Haas and Drabek, 1973). For the most accurate information, observers having the best opportunity to come in contact with a particular attribute ought to be used.

Inasmuch as an inclusive definition of the population of observers includes all members of the organization as well as such nonmembers as suppliers, customers, or clients, most investigators will have to use a subset of them to reduce the study to manageable proportions. With observers, a stratified random sampling plan is appropriate. All views of the organization should be represented, but there is no logic to support the idea that some members should have more impact than others, as might be the case with lower members, because there are usually more of them.

Both the Aston and Ohio studies defined the population of observers to include only top administrators, often the chief executive. Both groups admitted that as a result of this, their selection of attributes was curtailed. Clearly, neither study meets guideline 5. Since the population of observers was limited to one or a few administrators, there was no need to sample. Guideline 6 therefore does not apply.

Guideline 7: The sample of attributes must be no larger than the input capacity of the multivariate analysis program or else an iterative procedure of analyses based on randomly selected overlapping subsets of the sample should be used. It is easier to be able to analyze all attributes sampled at one time, but it does not make sense to limit the size of the sample just because an adequate multivariate analysis program is not available. Since the analysis is based on a sample of attributes, on the assumption other studies will follow, there is nothing unusual in running several analyses on subsamples of the data actually collected. It is necessary to make sure the subsamples overlap. Variables may be dropped from the analyses once it is clear they add nothing to the results of the investigation.

The Aston people started with a large pool of items which were operational measures of six of the seven dimensions. One dimension, flexibility, was deleted because they felt they could not properly measure it (Pugh, Hickson, Hinings, and Turner, 1968: 67–72). As a result of various item analyses, these items were segregated into 64 multiple item measurement sets—scales—each of which was treated as a variable. For reasons not made clear, the Aston group chose to base its analysis on a subset of 16 variables, some of which were composites of some of the 64 variables and some of which were not. These 16 were those that were most representative of the 64 or were most distinctive (Pugh, Hickson, Hinings, and Turner, 1968: 82). This was an unfortunate strategy, since the multivariate method it chose, factor analysis, was especially designed to suggest distinct groupings of similar variables. The Aston study does not satisfy guideline 7.
The Ohio study meets guideline 7 in that all 99 variables measured were included in the multivariate analysis.

Guideline 8: Each attribute must not be overrepresented in the input stream of the multivariate program and must be independently measured. Given the objective of carrying out empirical analyses free from the influence of previous classificatory thinking, it must be assumed that each attribute has equal potential as a classificatory concept. It must also be assumed that each attribute concept is a theoretical entity independent of all others in definition and capable of being operationally measured in some fashion. It follows that each attribute must be operationally measured in a way that does not compromise its independence, and leaves it equally represented in the input stream as a potential classificatory concept.

The Aston group used many items—operational measures—to build multiple item measures of more than one attribute, thus compromising their independence as operationally measured. For example, the multiple item measure of autonomy contained 22 of 37 items that made up the measure for centralization. The measure of standardization-selection contained 18 of the 68 items used for the overall measure of standardization. All of the traditionalism items came from other measures. Altogether, 4 of the 16 variables eventually included in the factor analysis were not independent because of a significant number of overlapping items. Other variables were not independent of the formalization variable—defined as amount of written rules, regulations, documentation—because the Aston group insisted on documentary support of all of its variables. Thus, the more documentary corroboration of a measure, the more it correlated with formalization. The consequence of including non-independent variables as inputs to the factor analysis was to produce a significant amount of artifactual common variance. Since the resulting factor structure reflected the effects of the artifactual variance, the results of the Aston study could be quite misleading.

In the Ohio study, tape recorded interviews were transcribed and then coded to prepare the data for computer processing. Damage to the independence of the items of this study, because of the intrusion of coder familiarity with prior classification schemes, was much more subtle than in the Aston study, if in fact such intrusion occurred. The authors did not indicate any steps specifically taken to avoid such intrusion, however.

Neither study included a preanalysis stage during which there was an attempt to weed out unanticipated, but obviously overlapping, interdependent measures which were inadvertently included among the attributes to be analyzed.

In the Aston study tendencies toward unequal weighting of the measures, because of unequal variances, were taken care of through the use of product-moment correlations as inputs to the factor analysis. In the Ohio study, there was considerable unequal weighting of the measures as a result of the researchers’ method of reducing scales containing several intervals to the either-or scales required by their computed program. Because the Ohio group reduced all of its variables
Guidelines

to a binary form, the unequal weighting did not appear as differences in the variance of the item distributions. Instead, each ordering type variable was translated into a different number of either-or subvariables, depending on the number of intervals in its response scale. For example, the measure “number of hierarchical levels” had a seven-point response scale and was translated into seven yes and no type subvariables. Thus, attributes having many either-or subvariables as measures had more impact on the final solution than variables that had only a few subvariables.

In sum, neither study wholly satisfies guideline 8; the Aston study included a large proportion of interdependent operational measures and the Ohio study did not have all its attribute concepts equally weighted at the operational level.

Guideline 9: Criteria guiding unavoidable decisions in using multivariate analysis must be publicly described and consistently applied. Multivariate procedures are not totally objective and replicable by others unless the key decisions are described in enough detail so that an independent observer may replicate the analysis, given the same data. Unfortunately, in multivariate procedures such as factor analysis the most critical decisions are either subjective or are based on objective criteria not universally agreed to be valid. These decisions are critical because they determine the number of classes and the number and kind of attributes used to identify each class.

The Aston group used the principal component method of extracting raw factors and a graphic rotation procedure (Pugh, Hickson, Hinings, and Turner, 1968; Levy and Pugh, 1969). According to Harman (1967) the principal component label meant unities were used as communality estimations; the Aston people did not indicate otherwise. Different communality estimations could have a significant impact on the loadings of the items in the factor analysis, especially if there were a relatively small number of variables.

One of the principal indeterminacies of factor analysis is how many factors to accept and rotate (Harman, 1967). This decision is especially critical in a classificatory study in which the number of factors accepted determines the number of classes ultimately accepted. Nowhere did the Aston group indicate what criteria led them to decide four factors should be rotated rather than two, three, or five. Although the Aston people first reported four factors (Pugh, Hickson, Hinings, and Turner, 1968; Levy and Pugh, 1969), in the later paper that presented the empirical classification (Pugh, Hickson, and Hinings, 1969), they only used three factors as their basic dimensions. No reason was given as to why the fourth factor was not included as a dimension in their classification, and yet this was probably the single most important decision of their entire study, from a classificatory perspective.

Assuming one has made a sound decision about how many factors to accept, there still remains a choice of whether to (1) use estimates of factor scores, thereby assuring orthogonality but requiring the use of all items, though they are differentially weighted (Harman, 1967), or (2) use only the core set of items loading highest on each factor, thereby using a smaller set of items but consequently needing to
make an independent confirmation of orthogonality and needing the specification of how the core items are selected.

For reasons not stated, the Aston group did not use estimated factor scores, a good approach since most studies of this sort do not meet the required assumptions. Instead, it took the second alternative and used subsets of the items loading on three factors. It assumed that these subsets were orthogonal because it used an orthogonal rotation solution (Levy and Pugh, 1969: 208), but this assumption is almost never warranted and should be tested. The Aston group included no data supporting the orthogonality of its resulting classificatory dimensions.

The criteria it used for deciding which items were actually included in defining and operationally measuring the three classificatory dimensions were not specified. Its decisions were somewhat arbitrary and confusing. For example, in Figure 2 of the Pugh, Hickson, Hinings, and Turner (1968) paper, the Aston group distinguished between the first three factors which it actually called "factors" and the fourth one which it labelled "percent of non-workflow personnel." Presumably, the reason it did not call the fourth one a factor in the diagram, and ultimately did not include it as a classificatory dimension, was that "percent of non-workflow personnel" only had one item with a loading higher than .6, although the Aston group never explicitly said this was its criterion for including items. The other three factors each had some items loading .6 or higher, and in the Pugh, Hickson, and Hinings (1969: 116) paper, it was clear that the variables chosen to measure each dimension had loadings of .6 or higher. But in some cases the researchers had not included all variables with loadings of .6 or more. In the first factor there were three variables with loadings above .6 which they did not include, and in the second factor there was one such variable not included. Furthermore, they included recording of role performance on factor three, even though it had a slightly higher loading on factor one. Why? Since this variable was factorially complex, with high loadings on more than one factor, and it added to the interdependence between the two factors, why did they not drop it out altogether? Finally, why include factor three? It has only one item with a .6 loading on that factor and no substantial loadings on any other factor. Factor four also had only one loading of .6 or better, but it was dropped.

The reasoning was clouded and the decisions inconsistent. Inasmuch as these subsets of core items ultimately were the operational basis of the classificatory dimensions used as dependent variables in later papers (Pugh, Hickson, Hinings, and Turner, 1969; Hickson, Pugh, and Pheysey, 1969; Inkson, Pugh, and Hickson, 1970; Hinings and Lee, 1971; and Pheysey, Payne, and Pugh, 1971), these seemingly small decisions undermined the validity and usefulness of the subsequent papers.

Instead of searching for underlying dimensions with which to classify organizations, the Ohio group took a different approach. Its computer program was designed to search through all of the organizations and find the organization which was most typical (see Haas, Hall, and Johnson, 1966, for
Guidelines
details on their method). After the organizations were ranked in terms of their similarity to the most typical organization, a cutoff point was selected, and organizations above the cutoff point were considered to be in the first cluster formed, that is, organizations most similar to the most typical organization. Organizations below the cutoff point were retained in the main body of organizations and then the procedure started over again looking for the most typical organization remaining. This iterative process continued until all organizations were clustered.

The entire configuration of the Ohio group’s classification scheme was determined by the choice of the cutoff point; it determined the size of each cluster and ultimately how many clusters were retained. The more organizations included in a cluster, the fewer the attributes they had in common (see Table 3 in Haas, Hall, and Johnson, 1966: 171). Despite its pivotal importance, the Ohio group did not discuss the pros and cons of different cutoff points that would result in fewer or more of what they called “prime classes of organizations.” Nor did they discuss what criteria might be used to evaluate results based on one cutoff value versus results based on another.

Neither the Aston nor the Ohio groups discussed the criteria and issues surrounding the most critical decisions in their multivariate methods, decisions that had a direct effect on the configuration of their classificatory schemes. Consequently, neither study meets guideline 9.

Guideline 10: Classificatory breaks in ordering type concepts should come at points optimizing parsimony and intraclass homogeneity. Inasmuch as most attributes of organizations are ordering type concepts, there remains the problem of translating them to fit the either-or classificatory type concept. Often organizations can be expected to have a normal distribution of any given attribute. With the usual bifurcation decision dividing organizations into two groups of equal size, most of the organizations are found bunched just on either side of the dividing line. Consequently, separation into the two categories is quite arbitrary for the many organizations right at the mode of the distribution. Many, often the majority, of the organizations in one class are very similar to those just across the dividing line in the other. A classification is most useful if it consists of a parsimonious set of very homogeneous classes. To maintain as much homogeneity within a class as possible, one would like to be able to organize the classes around one or more dominant modes along a given attribute dimension, with each class encompassing the modal organizations and a few on either side of the mode. The fewer the number of dominant modes along each attribute dimension, the more parsimonious the classification will be. Ultimately, it may be that taxonomists will accept for classification purposes only those attribute dimensions having a few dominant modes. In the meantime, it makes sense simply to identify the kinds of distributions associated with each attribute.

Judging from the distributions shown in Figures 1 and 2 (Pugh, Hickson, and Hinings, 1969), the three dimensions found in the Aston study, structuring of activities, concentra-
tion of authority, and line control of workflow, were clearly normal distributions. For all three dimensions, the Aston group placed the dividing line at the mean, right at the point where most of the organizations were. Thus, the distinctiveness of each class was obscured. The results would have been clearer if they had checked for bimodality and, having found several modes, had placed a dividing line between the modes or, having found only one mode, had broken the dimension down into an odd number of categories. In this way the resulting classes would have been more homogeneous and distinctive.

The Ohio group took a different approach to the task of working up a classification. After settling on a set of variables, many of which were ordering type concepts, it chose to reduce all variables to either-or codes. There was no discussion about the pros and cons of doing this. The consequences of translating ordering dimensions into classificatory type concepts is difficult to unravel in the Ohio group’s method. Seemingly, the most direct effect was on parsimony. In its approach, each point of an ordering type attribute became an either-or attribute. Thus, the variable number of hierarchical levels actually resulted in seven subattributes or subvariables. The Ohio group did not say how it would handle a variable such as the Aston dimension structuring of activities, which was a many-point ranking. The researchers would not consider having some 50 subattributes to cover one attribute concept. They did not say how many subattributes they actually ended up with, but supposing that each of their 99 variables resulted in 5 attributes, it could have been as many as 500 attributes. Theirs was hardly a parsimonious solution and yet there were many kinds of variables which they left out. Such a lack of parsimony is intolerable if the classification is ever to be used. Where would forestry be if it took 500 attributes to tell a pine tree from a spruce?

With the Aston method the choice was to base the analysis on multi-point dimensions and then attempt to find a way to reduce the number of dimensions to a few either-or classificatory type concepts. With the Ohio method the choice was to start with either-or dimensions, but run the risk of having an excessive number of attributes as a result of the way of handling variables having a multi-point response scale. Both sets of investigators ran head-on into the problem of attempting to produce a parsimonious set of classes from what were mainly ordering dimensions. In neither case was there much discussion of the consequences of their particular method of resolving the difficulty. Neither study meets guideline 10.

Summary

The degree to which the Aston and Ohio studies meet the ten guidelines and thereby minimize the intrusion of prior thinking about classification is summarized in Table 1.

The table indicates that the Aston study does not meet any of the guidelines and that the Ohio study meets two. Consequently, neither study comes close to meeting its stated objective of producing a classification drawn directly from empirical analysis which was free from the influence of prior classificatory concepts and schemes.
Guidelines

Table 1

<table>
<thead>
<tr>
<th>Guideline Number</th>
<th>Aston Study</th>
<th>Ohio Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Broad population of organization defined</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>2. Probability sampling of organizations defined</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>3. Inclusive population of attributes defined</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4. Probability sampling of attributes</td>
<td>No*</td>
<td>No*</td>
</tr>
<tr>
<td>5. Inclusive population of observers defined</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>6. Stratified probability sampling of observers</td>
<td>No*</td>
<td>No*</td>
</tr>
<tr>
<td>7. Appropriate input stream to multivariate program</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>8. Equal representation and weighting of attribute measures</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>9. Decision criteria described and consistently applied</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>10. Classificatory breaks optimizing parsimony and homogeneity</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

* This guideline does not apply since the population definition was arbitrarily attenuated at the start.

CONCLUSION

Organization science, and especially the application of its findings to the problems of organizations and managers, is not likely to emerge with viable laws and principles until substantial progress is made toward an acceptable taxonomy and classification of organizations. The basic inductive-deductive process of science does not work without the phenomena under investigation being divided into sufficiently homogeneous classes. Managers cannot use the fruits of science unless they first can discern which of all the scientific findings apply to their situation. The recent flourishing of contingency approaches (a critical review is given by Moberg and Koch, 1975) is in fact a grassroots response to the absence of useful classifications. If a usable classification existed there would be no need for contingency theory. Biologists do not need contingency theories because their taxonomy and classification scheme makes it clear that one does not apply findings about reptiles to mammals unless one is dealing broadly with the subphyla level of vertebrates. Organization and management researchers need contingency theories because there is no taxonomy to make it clear that one does not, for example, and only for example, apply findings from small British candy manufacturers to large French universities. It may be that it is best to look for an incipient classificatory scheme in the contingencies of the contingency theorists, were it not for their oversimplification and operational confusion (Moberg and Koch, 1975).

There is little doubt that the deductive classifications of theorists such as Parsons (1956), Etzioni (1961), Blau and Scott (1962), Katz and Kahn (1966), Perrow (1967), and Thompson (1967) are useless for empirical research (Burns, 1967; Hall, Haas, and Johnson, 1967). History will show the
Aston studies (Pugh, Hickson, Hinings, McDonald, Turner, and Lupton, 1963; Pugh, Hickson, Hinings, and Turner, 1968; Pugh, Hickson, and Hinings, 1969) and the Ohio studies (Haas, Hall, and Johnson, 1966) to be major turning points in the development of organization science, not because their first attempts at taxonomy produced usable results, but because they identified the problem and worked out two alternative methods of solving it. Not often do researchers have the insight to recognize what it is that is causing a science to drown in its own watery findings.

Blessed with 20-20 hindsight, this article reconfirmed the importance of taxonomy and set forth some guidelines to assure that inductive, empirically based approaches toward finding a useful classificatory scheme will not be inadvertently biased by unnecessary reliance on prior thinking about taxonomy and classification. The critique of the Aston and Ohio studies was presented to make the presentation of the guidelines more lively and relevant and to make it clear that the actual results of these studies did not satisfy their stated objectives. They are best viewed as illustrations of what can be done rather than as usable findings to be accepted and carried forward in subsequent research. The usefulness of this article rests more on the use of the guidelines in future multivariate approaches toward the problem of classification than it does on a critique of studies started well over a decade ago.

Bill McKevelley is an associate professor of management and organization design, Graduate School of Management, University of California, Los Angeles.

REFERENCES

Anderson, Theodore, and Seymour Warkov

Blau, Peter, and Richard A. Schoenherr

Blau, Peter M., and W. Richard Scott

Brown, Alvin

Burns, Tom

Child, John

Deutsch, Karl W.

Eilon, Samuel

Etzioni, Amitai


Evans, William M.

Glaser, Barney G., and Anselm L. Strauss

Gordon, C. Wayne, and Nicholas Babchuk

Haas, J. Eugene, and Thomas E. Drabek

Haas, J. Eugene, Richard H. Hall, and Norman J. Johnson

Hage, Jerald

Haire, Mason (ed.)

Hall, Richard H.

Hall, Richard H., J. Eugene Haas, and Norman J. Johnson
Guidelines


Physey, Diana C., Roy L. Payne, and Derek S. Pugh 1971 "Influence of structure at organizational and group levels." Administrative Science Quarterly, 16: 61–73


Khandwalla, Pradip N. 1973 "Viable and effective organizational design of firms." Academy of Management Journal, 16: 481–495

