

A Nontechnical Introduction to the Taxometric Method

John Ruscio

*Department of Psychology
Elizabethtown College*

Ayelet Meron Ruscio

*Department of Psychology
Pennsylvania State University*

Although social and behavioral scientists often presume that individual differences underlying measured variables consist of differences in degree rather than differences in kind, the distinction between taxonic (categorical) and dimensional (continuous) latent structure poses an empirical question with important implications for basic and applied science. In this article, we present a nontechnical introduction to the taxometric method for assessing latent structure. We outline unique features of the general approach and then describe and illustrate specific taxometric procedures, emphasizing the conceptual logic of each analytic technique, factors that can influence results and their interpretation, and decisions that must be made to implement each procedure most appropriately and powerfully. We present a number of refinements and extensions to taxometric methodology including a useful interpretive aid based on the parallel analysis of simulated taxonic and dimensional comparison data. We focus on practical suggestions for taxometric research, concluding with a checklist of 5 general questions that we believe should be thoughtfully considered and explicitly addressed when reporting or evaluating any taxometric investigation.

taxometric method, latent structure, coherent cut kinetics, taxon,
classification, consistency tests, MAMBAC, MAXCOV, MAXEIG,
L-Mode

Social and behavioral scientists have long preferred to conceptualize most constructs as continuous—rather than discrete—in nature. There appears, however, to be a growing recognition that the fundamental structure of some constructs may be categorical rather than continuous and that there is a need to empirically evaluate the structure of each construct of interest (Haslam & Kim, 2002). For example, in setting an agenda for the next century of research on psychopathology, Widiger and Clark (2000) noted that “the challenge facing the developers of [the *Diagnostic and Statistical Manual of Mental Disorders*] *DSM-V* may not be to differentiate more clearly between normal and pathologic expressions of behavior; rather, it may be to determine *whether or not a qualitative distinction can in fact be made* [italics added]” (p. 950). Making this determination poses a significant analytic challenge. Fortunately, developments pioneered by Meehl and his colleagues (e.g., Golden & Meehl, 1979; Grove & Meehl, 1993; Meehl, 1995a, 1999; Meehl & Golden, 1982; Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998) resulted in a statistical approach called the *taxometric method* that allows investigators to powerfully test the latent structure of constructs. The utility of this method paired with increasing emphasis on the importance of empirically evaluating (rather than presuming) structure has resulted in a marked increase in taxometric investigations in recent years (Haslam & Kim, 2002).

Although the taxometric method has the potential to serve as a powerful tool for basic and applied science, many researchers remain unfamiliar with the approach. This state of affairs may be at least partly attributed to the highly technical nature of most descriptions of taxometrics. Although the mathematical underpinnings of taxometrics are important, the absence of a more accessible introduction to the method continues to prevent many researchers from attaining a working knowledge of its procedures. Those unfamiliar with its technical aspects may find it difficult to critically evaluate taxometric investigations conducted in their areas of research. Currently available treatments of the taxometric method are also limited to relatively few procedures and tend to offer little or no practical guidance to those who wish to implement the method or evaluate its results. Finally, a number of questionable conventions have appeared in the applied taxometric literature but have not been directly addressed in existing sources.

It is with the interested nonspecialist in mind that we offer this nontechnical introduction to the taxometric method. After contrasting the general approach and basic principles of the taxometric method with those of alternative statistical techniques, we describe and illustrate a variety of taxometric procedures within the method. Because our primary aim is to facilitate conceptual understanding of taxometrics among the producers and consumers of this research, we articulate the key decision points in taxometric analysis and provide readers with the tools to make and evaluate these decisions in a reasoned manner. We emphasize practical guidelines, focusing on the logic by which the procedures operate and their effective implementation. We also introduce a number of methodological advances and

refinements including techniques based on the parallel analysis of simulated comparison data to test the suitability of data for a planned analysis and to facilitate interpretation. To promote informative applications of taxometric methodology and critical evaluations of taxometric research, we conclude with a brief checklist of general issues that we believe should be carefully considered and explicitly addressed in any taxometric study.

STRUCTURAL TERMINOLOGY AND THE TAXOMETRIC PROBLEM

Before delving into taxometric methodology, a handful of relevant terms must be defined to help clarify the nature of the problem that the taxometric method was designed to address. We begin with an overview of two important distinctions: that between manifest and latent structure and that between *taxa* and dimensions.

Latent structure refers to the fundamental nature of a construct, the underlying structure that exists regardless of how one might choose to conceptualize or measure it. *Manifest structure*, in contrast, refers to characteristics of observable measures of a construct, the surface structure that depends—among other things—on how the construct is conceptualized and assessed. Thus, the manifest and latent structures of a given construct may differ. A mixture of latent types can easily appear continuous at the manifest level, and latent continua may give rise to manifest types through threshold effects, sampling error (particularly when samples are small), selective sampling from the extremes of a continuum, observer bias, and a variety of other causes (Grayson, 1987; Haslam, 1999; Murphy, 1964).

The terminology most often used in reference to latent structure, particularly in the taxometric literature, involves the distinction between dimensional and taxonic latent structures. The term *dimensional* is used to refer to constructs along which individuals differ quantitatively from one another (i.e., along continua, scales, or factors) such that any groups that might be formed are arbitrary. Clear examples of dimensional constructs include barometric pressure and temperature, which can be scaled along the continua of millimeters of mercury and degrees centigrade. “High pressure” weather and “hot” objects are not naturally occurring types but are instead distinctions superimposed on dimensions for pragmatic purposes (J. Ruscio & Ruscio, 2002, in press).

By contrast, the term *taxonic* is used to refer to constructs in which individuals are separated into nonarbitrary groups (i.e., types, classes, categories) at the latent level. That is, one or more boundaries “carve nature at its joints,” with individuals either belonging or not belonging to these groups regardless of an observer’s beliefs or preferences. Two latent groups, or *taxa*, are traditionally referred to as the *taxon* (by convention, the higher scoring group on the indicator variables) and its *complement* (the lower scoring group on these variables). J. Ruscio and Ruscio (in

press) discussed more complex latent structures such as multiple latent taxonic boundaries superimposed on one or more latent dimensions.

Distinguishing latent taxa from latent dimensions poses an important empirical challenge. One of the most fundamental goals of science is to organize understanding of the world not only according to laws of cause and effect but also according to underlying structures. Indeed, fruitful systems have been developed to organize and classify aspects of the universe from its smallest to its largest detectable features. Although the development and refinement of a valid classification scheme can be extremely challenging in any science, in social and behavioral science this task must contend with a number of particularly thorny problems. Our measures are inevitably fallible, possessing some degree of measurement error that makes underlying structures difficult to detect. Even when taxa exist, they are likely to be difficult to distinguish using available variables and to contain considerable within-group variation. The fundamental challenge of structural research in social and behavioral science can be observed even in the deceptively simple task of distinguishing two latent taxa from a single latent dimension. Both of these relatively simple latent structures can give rise to similar distributional and correlational patterns in a data set, making it difficult to draw accurate structural inferences from the data.

The mathematical model that forms the basis for taxometric analysis illustrates the difficulties of inferring latent structure from such distributional or correlational patterns. The general covariance mixture theorem (GCMT) represents the expected covariance between two variables that each validly distinguish two latent taxa (Waller & Meehl, 1998):

$$\text{cov}(xy) = P \text{cov}_1(xy) + Q \text{cov}_2(xy) + PQ(\bar{x}_1 - \bar{x}_2)(\bar{y}_1 - \bar{y}_2), \quad (1)$$

where $\text{cov}(xy)$ is the covariance between indicators x and y in the total (mixed) sample; $\text{cov}_1(xy)$ and $\text{cov}_2(xy)$ are the covariances within the taxon and complement, respectively; P is the base rate of the taxon; $Q = 1 - P$; and \bar{x}_1 and \bar{x}_2 are the means of the two latent classes on indicator x . Given that variance is the covariance of a variable with itself, the expected variance of two variables that validly distinguish two latent taxa is

$$\sigma_x^2 = P\sigma_1^2 + Q\sigma_2^2 + PQ(\bar{x}_1 - \bar{x}_2)^2, \quad (2)$$

where σ_x^2 is the variance of indicator x in the total (mixed) sample and σ_1^2 and σ_2^2 are the variances within the taxon and complement, respectively.

Thus, a latent class model that allows for intraclass variance and covariance (as well as manifest overlap between classes) can explain the variances and covariances among a set of indicators. In fact, any variance-covariance matrix can be fit or reproduced equally well by either a latent class model like this one or by a latent dimen-

sional model such as one that models observed variances and covariances through loadings of all indicators onto a single underlying factor. More formally, data generated using a dimensional model with i latent dimensions can be perfectly fit using a taxonic model with $i + 1$ latent classes (Bartholomew, 1987). The classification problem—determining whether and how individuals are separated by taxonic boundaries as opposed to dimensional variation—clearly poses significant statistical challenges. To the extent that indicators are fallible (i.e., possess measurement error with respect to the construct under investigation), the differentiation of taxonic and dimensional models will be even more difficult for any statistical method.

However, the payoff from empirical evaluations of latent structure makes the effort worthwhile. Indeed, progress in many basic and applied scientific domains can be facilitated by research that tests the latent structure of relevant constructs (J. Ruscio & Ruscio, 2002, in press). There are strong theoretical reasons to conduct such research such as establishing more empirically defensible typologies and factor models, narrowing the list of potential etiological factors, and guiding future research agendas (see Haslam, 1997; Meehl, 1992; J. Ruscio & Ruscio, in press). In addition, there are practical payoffs such as improved research economy, more powerful research designs, better identification of appropriate populations for research and treatment, increased statistical power, and greater reliability and validity of measurement (see Fraley & Waller, 1998; A. M. Ruscio & Ruscio, 2002; J. Ruscio & Ruscio, 2002). Because all of these basic and applied benefits depend on the correct identification of latent structure, we turn now to a discussion of statistical approaches that can be used to make this crucial determination. Throughout, we restrict our attention to the most fundamental problem in structural research: distinguishing two latent taxa from a single latent dimension. Suggested approaches to the study of more complex structures are described by J. Ruscio and Ruscio (in press).

CONVENTIONAL APPROACHES TO TESTING LATENT STRUCTURE

Three broad families of analytic techniques have traditionally been used to test for taxonic boundaries: cluster analysis, finite mixture modeling, and latent class analysis. We briefly discuss each of these families before describing the role that the taxometric method may play in this structural arena.

Cluster Analysis

Cluster analysis represents a wide array of statistical procedures that are used to determine how many relatively homogeneous groups of cases can be distinguished in a given sample of data. Because a comprehensive review of cluster analysis is beyond the scope of this article (see Arabie, Hubert, & DeSoete, 1996; Everitt, 1993;

Lorr, 1994; McLachlan & Peel, 2001), we provide only a brief overview of approaches that have been used frequently in social and behavioral research. Several of the most popular clustering techniques involve a process analogous to generating a multidimensional scatterplot and testing whether cases are distributed haphazardly or clump together within this space. For a relatively large number of cases and indicator variables, one cannot feasibly test all distinct clustering solutions. Thus, the many available clustering algorithms serve as heuristic search tools.

Most of the widely used clustering techniques are hierarchical in nature and involve a two-step process. First, a measure of similarity or distance is selected to quantify the pairwise relations between all observations. Second, a mathematical rule is applied to parse the matrix of similarity or distance values into clusters. Agglomerative techniques, which begin by treating each case as a cluster and then fusing similar clusters until all cases have been joined into one cluster, represent the most popular approach (Everitt, 1993; Lorr, 1994).

The most significant problem with using such cluster analyses to empirically assess latent structure is the difficulty inherent in determining the appropriate number of clusters in the data. Because a hierarchical cluster analysis yields a one-cluster solution, a two-cluster solution, and so forth up to an N -cluster solution in which each case is treated as a cluster, the researcher must choose a solution corresponding to the most appropriate number of clusters. When the number of clusters is itself the central research question, this task requires an empirical index or stopping rule that can be shown to determine the correct number of clusters. Unfortunately, cluster analyses often yield spurious taxa because the stopping rules that have been developed are highly fallible (Grove, 1991; Milligan & Cooper, 1985). Stopping rules perform particularly poorly when distinguishing between the presence or absence of a single taxonic boundary (Lorr, 1994). Researchers (e.g., Lorr, 1994; Milligan, 1996) have noted that identifying the true number of clusters remains one of the most challenging problems in cluster analysis, leading Everitt (1993) to recommend that hierarchical cluster analysis be used as a descriptive tool to summarize data or to pose structural hypotheses. These concerns suggest that cluster analysis may have significant limitations as a test between taxonic and dimensional latent structure.

Finite Mixture Modeling

Finite mixture models are designed to determine the parameters of hypothetical subgroup distributions, or components, that reproduce the observed distributions of manifest variables (e.g., Everitt & Hand, 1981; McLachlan & Basford, 1988; McLachlan & Peel, 2001). By systematically varying the number of components and their parameters (e.g., base rate, mean, standard deviation, skew, kurtosis), one can identify a best-fitting mixture model. When this technique is used to assess latent structure, the number of components in the best-fitting model is interpreted as the number of latent taxa. For example, if a two-group model better accounts for the

distribution of manifest scores than a one-group model, this suggests that two taxa underlie the scores.

Finite mixture modeling also has potential limitations as a tool for investigating taxonic boundaries. Univariate mixture models do not allow researchers to test for taxonic boundaries that can only be represented by multiple indicators, although this problem may be solved through the use of multivariate mixture models. More important, as with cluster analysis, it may not be possible to determine the correct number of components in a mixture with sufficient validity. McLachlan and Basford (1988) described and critiqued several quantitative fit indexes that have been proposed for this purpose and noted that this remains a "very difficult problem which has not been completely resolved" (p. 21), a conclusion echoed by Bock (1996) and repeated by McLachlan and Peel (2001, p. 175). McLachlan and Peel demonstrated that many fit indexes tend to overestimate the correct number of components and noted that although normal mixture models can adequately model skewed indicators, the best-fitting model will often include spurious components to accommodate the skew. Given that nonnormal distributions are ubiquitous in psychological research (Micceri, 1989), the tendency of finite mixture models to uncover too many components in the presence of skewed data seems to be an important limitation of these procedures.

Latent Class Analysis

Latent class analysis (e.g., Green, 1951; Lazarsfeld & Henry, 1968; Muthén, 2001; Uebersax, 1999) proceeds much like finite mixture modeling in that responses to a set of indicators (usually categorical in nature, although latent profile analysis extends the approach to accommodate continuous indicators; Muthén, 2001) are modeled using a latent class variable. Competing models consisting of one or more latent classes are tested against one another using any of a number of fit indexes to determine the number of latent classes that best represents the target construct.

Latent class analysis is complicated by a number of factors (Uebersax, 1999). Its assumption of conditional independence, which requires all indicators to be statistically independent within each latent class, may be unrealistic and can affect the fit of many latent class analysis models. Moreover, when calculating model fit using a maximum likelihood algorithm based on a particular set of initial parameter values, or seeds, to represent the putative latent classes, the emergence of local maxima presents a computational challenge in that different seeds may result in widely varying solutions. However, the most fundamental limitation is that like finite mixture modeling, latent class analysis has a tendency to overidentify the number of latent classes (Uebersax, 1999). Whereas latent class analysis can be useful for assigning cases to classes once latent structure has been assessed using more suitable statistical procedures, it has not proven itself sufficiently valid for the initial determination of the number of latent classes.

A final difficulty arises when using any statistical technique—cluster analysis, mixture modeling, latent class analysis, or otherwise—that searches for more than two latent taxa simultaneously using a single set of manifest indicators. For example, a researcher might wish to evaluate the latent structure of major depressive disorder (MDD) by submitting a large number of affective, cognitive, and somatic depressive indicators—including those relevant to a putative MDD taxon and its hypothesized subtypes—to one analysis. This approach appears quite efficient in that it searches for all taxa (e.g., depressive types and subtypes) in a single analysis. However, such simplicity may also pose a conceptual and empirical problem. A single set of indicators is unlikely to be equally valid for detecting multiple, heterogeneous taxa. Taxa best represented by a subset of the indicators may not be detected when irrelevant indicators are included in the analysis or when indicators unique to that taxon are excluded from the analysis. For example, the search for a putative melancholic subtype of depression may yield false negative results if only indicators that distinguish depressed from nondepressed individuals are included or if indicators that do distinguish the hypothesized subtype are diluted by the inclusion of indicators that do not. In short, it may be unrealistic to expect any statistical procedure to validly differentiate more than two latent taxa within a single analysis.

TAXOMETRIC METHOD

Although the term *taxometrics* broadly refers to the entire domain of empirical classification (Meehl & Golden, 1982), it is most commonly associated with the *coherent cut kinetics* taxometric method pioneered by Meehl. The name of this method reflects the process by which many of its procedures work: A summary statistic is calculated repeatedly to determine whether predictable results are obtained (hence coherent) as a cutoff point is moved through a distribution of scores (hence cut kinetics). The coherent cut kinetics framework is because the GCMT (Equation 1) describes the relations between valid indicators of two latent taxa not only in a full sample of data but also in any subsample. Our use of the terms *taxometric method* and *taxometrics* refer solely to Meehl's (1995) coherent cut kinetics approach that searches for predictable relations among indicators across ordered subsamples of cases.

The taxometric method is a hypothesis-testing (rather than exploratory) technique with several noteworthy features. First, taxometric procedures explore the relations among indicators rather than the distributions of individual variables to infer the latent structure of the construct under investigation. Second, each procedure yields estimates of important latent parameters as well as graphical results that can be visually inspected with high reliability. Third, nonredundant lines of evidence are contributed by procedures that operate in mathematically different ways. Fourth, results are inspected for converging evidence within tolerable margins of error rather

than submitted to tests of statistical significance. Fifth, the final determination of latent structure is based in part on the degree of consistency among all obtained results. We discuss each of these general features before turning to a more specific description of the individual procedures that comprise the method.

Relations Among Indicators

Taxometric analysis presents a “bootstraps” problem (Cronbach & Meehl, 1955) in that it uses fallible, manifest variables to evaluate the latent structure of a construct in the absence of a “gold standard” criterion. Whether observable variables (or indicators) of a construct are themselves distributed categorically or continuously, the relationships between these indicators can provide valuable clues about the underlying structure of the construct. Taxometric procedures capitalize on predictable differences in the way that indicators interrelate in the presence of taxa versus dimensions, as derived from the GCMT.

One important benefit of their use of multiple indicators is that taxometric procedures can be used to test the structure of constructs for which no single variable is a necessary and sufficient defining characteristic. For example, schizophrenia is not synonymous with visual or auditory hallucinations—there are a number of cognitive, affective, and behavioral features involved in this disorder (American Psychiatric Association, 1994). Thus, whereas statistical procedures that examine the distribution of one indicator at a time (e.g., finite mixture models with a single variable; McLachlan & Peel, 2001) may be useful for testing the structure of those individual signs or symptoms, taxometric procedures are akin to techniques (e.g., multivariate normal mixture models, cluster analysis, latent class analysis) that incorporate multiple indicators in each analysis to test the structure of a more complexly defined construct. Because we suspect that relatively few substantive constructs can be adequately represented by a single indicator, we recommend that multiple relevant indicators be submitted to procedures such as taxometric analyses to evaluate the latent structure of the construct that they polythetically define.

Graphical Results and Numerical Estimates of Latent Parameters

Taxometric procedures yield graphs that provide clues to latent structure along with numerical estimates of important latent parameters (e.g., the proportions of cases belonging to the taxon and complement). As we show, when taxometric procedures are provided with suitable data for analysis, they produce different curve shapes for taxonic and dimensional data. Taxometric graphs can be visually inspected with a high level of interrater agreement (see Meehl & Yonce, 1994). Moreover, graphical results allow investigators to examine the influence

of potentially problematic factors (e.g., skewed indicators) that might otherwise be obscured in purely quantitative results. Finally, considerable Monte Carlo evidence suggests that estimates of latent parameters yielded by taxometric analyses are quite accurate (e.g., Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998).

Nonredundant Lines of Evidence

The taxometric method includes a set of diverse analytic procedures that assess latent structure in different ways. Each procedure is based on the same formal-numerical definition of taxonic latent structure captured in the GCMT, yet each uses a unique approach to corroborate or refute this model. Because taxometric procedures differ in their theoretical rationales and mathematical operations, they contribute complementary pieces of evidence. A bedrock principle of the taxometric method is that multiple analytic procedures must be employed to provide independent lines of evidence for a given structural solution, thereby reducing the likelihood of an incorrect conclusion. This aspect of taxometric methodology can be contrasted with statistical techniques for which there are no complementary analytic procedures to contribute nonredundant evidence or for which there is little expectation that the available procedures will suggest the same underlying structure.

Converging Evidence Within a Tolerable Margin of Error

The results obtained through taxometric analyses—both graphical and parametric—are evaluated in terms of their consistency with taxonic or dimensional latent structure, not in terms of their statistical significance. There are two reasons why the method does not employ significance tests. First, any conjectured model of latent structure may only imperfectly capture the true state of the construct as it exists in nature. Furthermore, because the mathematical derivations of all statistical tests include simplifying approximations, tests that measure departures from perfect fit are confounded by what is already known: that the approximations are not perfectly accurate. Rather than seeking the unattainable goal of perfect model fit, taxometric procedures seek to reliably and validly distinguish taxonic from dimensional structure.

For example, taxometric procedures are derived using the simplifying approximation of negligible *nuisance covariance* among indicators within taxa. That is, one presumes that if groups exist, indicators will be uncorrelated within them (this is similar to the assumption of local independence required by latent class analysis; Uebersax, 1999). In taxometrics, it is acknowledged that this approximation will never literally be true. However, studies have revealed that within-group correlations as high as .30 ordinarily have little impact on the results of taxometric procedures (Beauchaine & Beauchaine, 2002; Meehl, 1995a). Thus, given its robustness

to deviations from this idealization, the taxometric method is useful even in situations in which fit indexes might otherwise discard useful models.

A second reason that taxometric procedures do not rely on tests of statistical significance concerns the methodological paradox by which researchers can be penalized for working with large samples (Meehl, 1967). Fit indexes will almost always reveal statistically significant departures from perfect fit in large samples, and large samples are required for informative taxometric investigations. Given that no model of latent structure will perfectly specify all latent parameters, the penalty for working with appropriately large samples virtually guarantees the rejection of even highly exemplary models. Although this drawback of fit indexes is lessened when multiple competing models are compared, the informational value of these statistics remains confounded with the degree to which simplifying assumptions are met. That is, these assumptions may be better satisfied by some models than by others, and it may be impossible to disentangle the extent to which models truly fit the data from the extent to which important assumptions of those models are satisfied.

Consistency Checks

In place of null hypothesis significance tests, the taxometric method relies on the convergence of evidence from as many nonredundant sources as possible, with increasing coherence leading to increasing confidence in the structural solution. Because the method consists of several independently derived analytic procedures, each affording a judgment about latent structure independent of the others, the use of multiple taxometric procedures enables these techniques to serve as consistency tests for one another. This system of consistency checks, unique to the taxometric method, guards against the potential misinterpretation of isolated results that may be anomalous and inaccurate.

In addition to the taxometric procedures themselves, estimates of latent parameters (e.g., the base rate of taxon membership, the distribution of taxon and complement members' scores on each indicator, intraclass indicator correlations) yielded by these procedures provide a valuable opportunity to check for consistency. Estimates of the taxon base rate (i.e., the proportion of cases belonging to the taxon) are most commonly used for this purpose. Multiple base-rate estimates can be compared within and across taxometric procedures; close agreement among these estimates often suggests that analyses are converging on a group of cases belonging to a genuine latent taxon. In contrast, clear divergence among base-rate estimates across procedures and analyses suggests the absence of a latent taxon, providing support for a latent dimension.

There are a number of additional ways to check the consistency of taxometric results. For example, each taxometric procedure may be conducted multiple times using all available indicators in a variety of configurations. Likewise, multiple sets

of indicators can be analyzed, each constructed using a somewhat different approach (i.e., differentially stressing empirical vs. theoretical considerations). The comparison of graphical results and base-rate estimates across analyses performed with multiple sets and configurations of indicators provides many opportunities to assess coherence among structural solutions. In a well-conducted taxometric investigation, each analysis adds to the rigor with which latent structure is tested. Whereas any single result may be ambiguous or misleading, confidence in a structural solution is bolstered when evidence from multiple sources converges.

Relative to the alternative approaches described earlier, these features of the taxometric method make it arguably the most appropriate statistical technique for validly inferring the presence or absence of a taxonic boundary. In the following, we review the requirements for performing taxometric analyses and discuss ways in which the challenges posed in meeting these requirements reflect some important limitations of the method.

SUITABILITY OF DATA FOR TAXOMETRIC ANALYSIS

As with any statistical tool, taxometric analysis must be performed with suitable data to yield interpretable and meaningful results. Meehl (1995a) emphasized the need for large samples (e.g., at least 300 cases) containing valid indicators (e.g., at least a 1.25 *SD* separation between putative groups) for a taxometric analysis to be informative. Based on the findings of Monte Carlo studies as well as our own experience in conducting and reviewing taxometric research, we offer several additional suggestions for determining whether data are suitable for taxometric analysis. It is not our intention to establish specific or inflexible criteria for suitability but to identify relevant factors that should be explicitly considered and reported in a taxometric investigation. We advocate an approach to testing data suitability that focuses on the unique characteristics of a given data set rather than relying solely on general rules of thumb derived from Monte Carlo studies or conventional practices. In this section, we discuss issues of sampling and indicator selection and describe a data simulation technique that can be used to empirically evaluate the suitability of a given data set for taxometric analysis.

Sampling

There are two significant issues to consider when assessing the appropriateness of a sample for taxometric analysis. First, the sample must contain enough members of the putative taxon to allow the taxon to be detected. Second, because samples selected according to specific criteria can yield *pseudo-taxa* (i.e., false taxa reflecting methodological artifacts rather than true latent structure), it is preferable to perform taxometric analyses on unselected samples. We discuss each of these issues in turn.

Monte Carlo research suggests that a taxon base rate of at least .10 is needed for the successful detection of a small group (e.g., Meehl & Yonce, 1994, 1996). However, there are several reasons to treat this generalization with caution. First, it is possible that taxa with base rates less than .10 can be detected, especially in very large samples; studies have not yet systematically tested the sensitivity and specificity of taxometric procedures with smaller taxon base rates (Beauchaine & Beauchaine, 2002, provided the only exception, although they evaluated only sensitivity—and not specificity—for just one taxometric procedure, *MAXimum COVariance* [MAXCOV; Meehl & Yonce, 1996]). Second, other characteristics of the data (e.g., indicator validity, nuisance covariance within taxa, indicator skew) may make it more or less difficult for taxometric procedures to detect small taxa, making a single, absolute minimum less appropriate than a standard that takes other data parameters into consideration. Third, it may be true that the absolute number of taxon members is as important as the base rate of taxon membership. J. Ruscio and Ruscio (in press) demonstrated that once a sufficient number of taxon members was present in a sample, the addition of large numbers of complement members did not obscure this taxon, even when its base rate dropped well below .10. This argues against removing putative complement members from a data set to increase the taxon base rate, as this practice may provide only an illusory gain.

Another important sampling consideration is to avoid selecting or constructing a sample in a manner that may yield spuriously taxonic results. Ideally, an unselected sample from the target population of interest should be used. If the putative taxon represents an exceedingly rare class of individuals (e.g., when studying a rare disorder or talent in a community sample), a very large sample may be needed. Although it may be tempting to reduce the number of participants by constructing a sample consisting of more balanced numbers of suspected taxon and complement members, this approach can lead to subsequent interpretive problems. For example, pooling data from 250 disordered individuals and 250 normal individuals, as in case-control designs, can be problematic because individuals who meet some—but not all—diagnostic criteria will be largely absent from this sample. Thus, taxonic results may reflect the artificial nature of the sample (the fact that moderate or borderline cases have been excluded) rather than the true latent structure of the disorder. Grove (1991) discussed more complex problems of “institutional pseudo-taxa” that can result from nonrandom selection along one or more variables relevant to the construct under investigation.

Multiple, Quasi-Continuous Indicators

All of the procedures in the taxometric method were developed for use with continuous variables, and most procedures require the use of at least one indicator with enough variation to allow cases to be reliably sorted or divided into many subsamples. Although there is some evidence that at least one taxometric procedure

(MAXCOV) can be modified to accommodate dichotomous indicators (J. Ruscio, 2000), dichotomous data result in diminished power and altered curve shapes that make interpretation more difficult. The performance of other taxometric procedures with dichotomous indicators or even with indicators that have a small number of values, has not been systematically studied. Thus, caution is warranted when interpreting the results of any taxometric analysis performed with indicators that are not distributed along enough values to reasonably approximate a continuous scale.

Fortunately, there are ways to avoid the problem of insufficient indicator variability. If candidate variables have too few values to permit their independent use, it is often possible to form suitable composites by aggregating variables that represent a similar facet of the construct under study. This approach has the potential added benefit of increasing the reliability and validity of the composite indicator relative to its constituent variables. When forming composite indicators, however, care must be taken to minimize nuisance covariance (correlations within either putative group). One way to do so is to ensure that correlations between variables within a composite are much lower than correlations between variables in different composites. For example, all variables that validly measure one symptom cluster of a mental disorder can be combined to form a composite indicator representing that cluster. We recommend against randomly or haphazardly constructing composites from a pool of variables, as this greatly increases the likelihood of unacceptably high nuisance covariance (J. Ruscio, Ruscio, & Meron, 2004).

Because taxometric procedures work by exploring the relationships among indicators, multiple variables are required to conduct taxometric analyses. As an absolute minimum, two indicators are necessary, although it is far preferable to have at least three or four. However, whereas a large number of indicators is desirable, a point is reached where additional indicators may be highly redundant with those already selected and hence increase nuisance covariance. Likewise, there are only so many variables available in a given data set, and as more and more composites are formed, each will include fewer items and vary along a smaller range of values. Therefore, it is important to recognize these competing goals in indicator selection and to strike a reasonable balance between them.

Indicators Adequately and Uniquely Represent the Construct of Interest

The indicator selection process may be the most critical step in ensuring that a taxometric investigation will yield meaningful results. As noted earlier, the taxometric method is designed to test competing structural hypotheses for a specified target construct, not to search for potential latent taxa in an atheoretical, purely exploratory manner. This construct is represented by one's choice of indicators (Meehl, 1986). In discussing the importance of proper indicator selection, Widiger (2001) stressed that the interpretation of taxometric results will depend on the ex-

tent to which indicators comprehensively assess relevant features of the putative taxon, noting that poor choice of indicators can produce misleading results. For example, a taxometric analysis of a mental disorder should incorporate indicators representing all relevant features of the disorder; the absence of one or more symptom domains would mean that the structure of some other, related construct was being tested. By the same token, analyses may be seriously confounded if a significant number of indicators inadvertently triangulated on another construct.

Thus, to effectively test the latent structure of the target construct, one must grapple with the breadth and specificity of a set of indicators in much the same manner as a researcher who engages in scale development. The content coverage of a scale is an important indicator of the validity with which it assesses the desired construct. If important aspects of a construct are not tapped by the items of the scale or if a significant proportion of the items tap a different or related construct, the construct validity of measurements yielded by the scale will be called into question. In the same way, the interpretation of taxometric results depends on the use of indicators that adequately and uniquely represent the target construct.

Valid and Nonredundant Indicators

Taxometric procedures yield informative results only when they are performed with indicators that are capable of distinguishing the putative taxon and complement. This requires indicators to be sufficiently valid and to possess negligible nuisance covariance. (In the taxometric literature, the term *validity* refers to the separation between groups or the efficiency of classification.) When indicators with questionable validity or nuisance covariance are analyzed, they may produce uninterpretable (or worse, misleading) results. Thus, special care should be taken to ensure that the available indicators are empirically suitable before proceeding with any taxometric analyses.

There are several things that investigators can do to enhance the validity and minimize the nuisance covariance of their indicators. Relevant theory as well as previous research may suggest variables that are most valid for the construct at hand. Likewise, examination of the content of items, scales, and other measures may suggest which potential indicators are likely to covary within putative taxa. Indicators may be selected by first choosing the variables that are hypothesized to be the most valid then combining or discarding those variables that appear to be redundant and are likely to be associated within groups.

This reasoned evaluation of the available data may be supplemented by analyses to estimate the likely validity and nuisance covariance of the indicators. It is often useful to begin by examining the corrected item-total correlations of all candidate variables. If these variables are indicative of the target construct, their total score should itself be a highly valid measure of that construct. Thus, those variables that are most highly correlated with the total score are likely to be among

the most valid indicators of the construct. Rather than simply choosing the variables with the highest corrected item-total correlations (which may inflate nuisance covariance if some of the most valid variables are redundant with one another), variables can be selected with an eye toward minimizing overlap in their content. For example, if the construct of interest is a mental disorder, one might select the most valid indicator from each of its major symptom clusters rather than selecting multiple variables from just one symptom cluster because these happen to be the most valid items overall. Alternatively, multiple valid variables measuring each symptom cluster can be combined to form composite indicators.

In addition to careful consideration of content overlap, nuisance covariance can also be empirically estimated by calculating interindicator correlations within subsamples of cases in the outer regions of the total score distribution (Meehl, 1995b). For example, the upper and lower quartiles are likely to contain relatively pure subsamples of taxon and complement members, respectively (Golden & Meehl, 1979), although smaller subsamples could be used if either group is suspected to be particularly small. If indicators are correlated only weakly (e.g., less than .30) within these high- and low-scoring subsamples, this suggests that nuisance covariance is likely to be tolerably low (Meehl, 1995a).

Given an estimate of the taxon base rate P (which can be drawn from previous research, indicator frequency distributions, a fallible external criterion, an educated guess, or a combination of these approaches), the average manifest correlation between indicators in the total sample (r), and the average degree of nuisance covariance (within-group correlation r_i), validity (d , in standardized units) can be estimated using a simple variant of the GCMT presented in Meehl and Yonce (1996, p. 1146):

$$d = \sqrt{\frac{r - r_i}{PQ(1 - r)}}, \quad (3)$$

where Q is the putative complement base rate ($1 - P$). This formula presumes equal indicator variances and covariances within and between groups, but even such a rough estimate of indicator validity can be valuable for determining whether to proceed with a taxometric analysis—and later for interpreting the results. For example, if investigators fail to establish that their indicators are actually capable of detecting putative taxa, they may be unable to draw confident conclusions if their results appear dimensional.

Simulating Comparison Data

One approach that may prove particularly helpful in evaluating the suitability of data for taxometric analysis is to simulate taxonic and dimensional comparison

data sets and subject them to the taxometric analyses planned for the research data. These data sets can be simulated to very closely match the distributional and correlational properties of the research data while having different, known latent structures. By subjecting simulated data sets to parallel series of taxometric analyses, one can determine whether the planned analyses will be capable of distinguishing taxonic from dimensional latent structure given the unique characteristics of the research data (J. Ruscio et al., 2004). Thus, the suitability of the data and the analysis plan are evaluated simultaneously. If discernibly different results are obtained in analyses of the simulated taxonic versus dimensional data sets, this suggests that a taxometric analysis of the research data will afford a genuine test of latent structure. If results for simulated taxonic and dimensional data sets cannot be distinguished, then either the data or the analysis plan should be revised and retested until results can be distinguished. Used this way, simulated comparison data may afford the most powerful test of the empirical suitability of data for analysis. This technique is illustrated in all of the taxometric analyses we perform later in this article.

PROCEDURES IN THE TAXOMETRIC METHOD

When taxometric analysis is conducted with a sufficiently large sample of data containing multiple, quasi-continuous indicators that adequately and uniquely represent the target construct, distinguish the putative taxon and complement with sufficient validity, and evidence negligible nuisance covariance, the analysis should yield consistent results that clearly point toward either a taxonic or a dimensional structural solution. We now introduce the major taxometric procedures that can be used to perform these analyses and provide guidelines for their successful implementation. We begin by explaining the unique logic of each procedure then illustrate and describe the results of each procedure using a taxonic data set containing four indicators. This data set was simulated so that its actual latent structure would be known.

Rather than using a generic label such as “sample data” to describe our data set, we sought to enhance the clarity of our illustrations by patterning the data after a psychological construct that has been studied extensively using the taxometric method. Schizotypy has been theorized to represent a genetic liability for schizophrenia (Meehl, 1962, 1990), and taxometric studies have supported the existence of a latent schizotypy taxon (Golden & Meehl, 1979; Korfine & Lenzenweger, 1995; Lenzenweger, 1999; Lenzenweger & Korfine, 1992). Our taxonic data set consisted of four indicators of schizotypy named in accordance with Meehl’s (1990) theory: cognitive slippage, social aversiveness, anhedonia, and ambivalence. The simulated schizotypy data set contained 600 cases. To illustrate what taxometric results look like with groups of unequal size, the base rate of schizotypy was set at .25 in our sample. The taxon (schizotype) and complement (nonschizotype) groups were separated by 2 SD s; the indicators were therefore highly valid, although not unrealistically so (see, e.g., A. M. Ruscio, Ruscio, &

Keane, 2002; J. Ruscio & Ruscio, 2000). In addition, a number of “messy” characteristics typical of research data were incorporated into this data set: Indicators were correlated within the taxon and the complement, moderately positively skewed, and distributed along 10-point scales. Thus, these data possessed some attributes favorable for taxometric analysis (e.g., a sufficiently large taxon that was validly separated from the complement) as well as some attributes inhospitable to taxometrics (e.g., moderate nuisance covariance and discontinuous, skewed indicators). Table 1 displays the interindicator correlation matrix in the full sample, the taxon, and the complement. Figure 1 displays the distribution of scores on an indicator in the full sample, the taxon, and the complement (the distributions of the other three indicators were highly similar, differing only due to sampling error).

TABLE 1
Correlations Among the Four Indicators in the Schizotypy Data Set

	Full Sample ^a			Taxon ^b			Complement ^c			
	1	2	3	1	2	3	1	2	3	
2	.60			2	.28		2	.19		
3	.59	.61		3	.26	.22	3	.14	.23	
4	.61	.57	.62	4	.32	.14	.35	4	.19	.21

Note. Correlations for taxonic and dimensional comparison data were compared to the target values in the schizotypy data to ensure that relationships among indicators were reproduced with sufficient precision. Within each sample or subsample, the largest absolute residual correlation was recorded: dimensional comparison data, full sample = .01; taxonic comparison data, full sample = .01; taxon = .02; complement = .04.

^aN = 600. ^bn = 150. ^cn = 450.

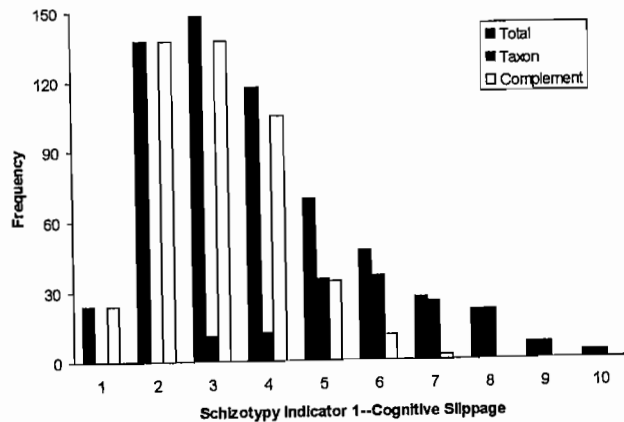


FIGURE 1 Frequency distribution for cognitive slippage, the first indicator in the schizotypy data. Note that there is no hint of bimodality in the full sample.

In what follows, we explain the rationale underlying each taxometric procedure, note the key decisions that must be made to implement this procedure, and demonstrate how it is performed through analysis of the schizotypy data set as well as parallel analyses of taxonic and dimensional comparison data sets simulated to match the distributional and correlational properties of the schizotypy data. Readers interested in more technical discussions of these procedures are directed to the references provided for each. All analyses were performed using our suite of taxometric programs written in the R language (also available in an S+ version); the program code and a detailed documentation file can be downloaded, see J. Ruscio (2003).

Mean Above Minus Below A Cut (MAMBAC): Searching for an Optimal Cutting Score

The MAMBAC (Meehl & Yonce, 1994) procedure takes advantage of the fact that if two groups exist, there must be an optimal cutting score for distinguishing between them. That is, if an indicator validly separates two latent taxa as depicted in the GCMT, there must be a particular score on this indicator that will minimize the number of false positive and false negative classifications of cases into these groups. In the absence of latent taxa, such an optimal cutting score does not exist. MAMBAC requires two valid indicators, one of which is treated as the input and placed along the *x*-axis of a graph. Cases are sorted according to their scores on this input indicator. Then, the mean score on the other (output) indicator for all cases falling below the input cut is subtracted from the mean score of all cases falling above the input cut. This subtraction is repeated for a number of cutting scores along the *x*-axis, and each mean difference is plotted as the corresponding *y* value. Thus, MAMBAC involves plotting mean differences on the output indicator above, minus below, cutting scores on the input indicator. The shape of the resulting curve allows one to make an inference about latent structure: Taxonic constructs yield peaked curves, whereas dimensional constructs yield concave curves that tend to bow upward at one or both ends.

In addition to inspecting the shapes of these MAMBAC curves, one can calculate an estimate of the taxon base rate from each curve (see Meehl & Yonce, 1994). For taxonic data, the location of a peak suggests the relative size of the two groups. Equal-sized groups produce a peak toward the center of the curve because the optimal cutting score lies near the middle of the input indicator range. For a small taxon, the optimal cutting score is higher, and thus, the peak is deflected toward the right. Similarly, for a large taxon, the optimal cutting score is lower, and the peak is deflected toward the left.

Implementation decisions. Three decisions must be made to implement the MAMBAC procedure, each of which can have a significant influence on results. The first decision concerns how to assign variables to the roles of input and output indicators. One can use variables in all possible input–output pairs generating $k(k - 1)$

MAMBAC curves for k indicators. Alternatively, one can remove a single variable at a time to serve as the output indicator and combine all remaining variables (e.g., through summing) to form a composite input indicator, yielding k curves. Finally, one can combine a subset of the available variables to form the output and combine the remaining subset to form the input, yielding an intermediate number of curves. An important benefit of combining variables, particularly in the role of input, is that the resulting indicator contains a larger range of values that provides a more reliable rank ordering of cases (and hence a more stable MAMBAC curve). Combining variables to form composite input or output indicators also allows all data to be included in each analysis, which increases the statistical power of the analysis.

Having determined how to assign variables to input–output indicator roles, one must decide where to place cuts along the input indicator. Here, too, there are many options. Cuts may be made between each successive case, yielding the maximal number of points on the curve for interpretation. However, a smaller number of points often suffices for interpretational clarity and may dramatically reduce the computational demands of the analysis. Methods of selecting a more limited number of cut points include cutting at intact scale values, at fixed standard deviation intervals, or at every n th case along the input. The latter method may be the simplest in that one need only determine the desired number of points to appear on the MAMBAC curve and place evenly spaced cuts to achieve this outcome.

If one or more cuts happens to fall between equal-scoring cases on the input, it can be useful to perform internal replications by reshuffling the equal-scoring cases, recalculating MAMBAC values, and plotting the average values across all replications to produce a curve for a given input–output configuration. With increasing numbers of replications, the obfuscating effect of cutting between equal-scoring cases will be reduced. Also, regardless of the method that is used to locate cutting scores, it is important to place the first and last cuts far from the ends of the input range to stabilize the endpoints of the curve. For example, many researchers have used a minimal n of 25 cases beyond which no further cuts are made.

The third and final decision concerns how the MAMBAC results should be plotted. The interpretation of MAMBAC curves is usually facilitated by using the narrowest range of y values that includes all MAMBAC values. This will draw out the shape of the curve, be it peaked or concave, most clearly. Although a smoothing procedure is seldom necessary to help discern the shape of a MAMBAC curve, particularly when a large number of points have been plotted, smoothing may improve the accuracy of base-rate estimation. This is because the taxon base rate is estimated from only the two endpoints of the MAMBAC curve, and these are the points that are most influenced by sampling error.

Empirical illustration. We now turn to a demonstration of MAMBAC using the schizotypy data set. First, MAMBAC was conducted in the traditional manner by using the four indicators in all possible input–output pairs to generate a total of

12 curves (see Figure 2, left column; to conserve space, only the averaged curves are presented). Second, to boost statistical power, we reran MAMBAC with the input constructed via the summed input indicator method, yielding a total of four curves (see Figure 2, right column).

To provide a more detailed account of how MAMBAC is performed, we focus for the moment on the summed indicator analyses. We generated these four MAMBAC curves by using each indicator once in the output role, with input scores for each curve obtained by summing the three remaining variables. To cre-

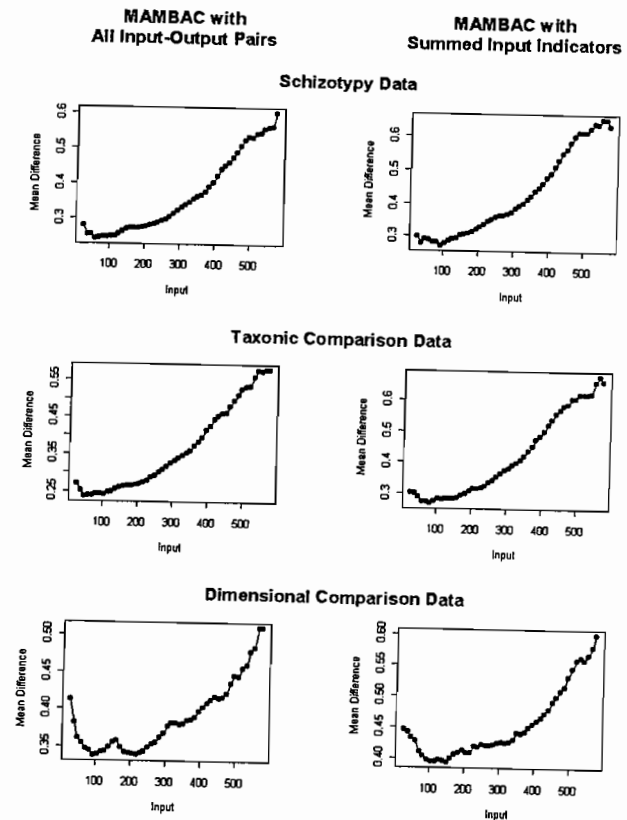


FIGURE 2 Mean above minus below a cut (MAMBAC) plots for the schizotypy data plus the taxonomic and dimensional comparison data. Averaged curves in the left-hand column were generated using the four indicators in all possible input–output configurations ($n = 12$ curves), whereas those in the right-hand column were generated using one indicator as the output and the sum of the remaining three indicators as the input ($n = 4$ curves). Cuts were made at 50 equally spaced points between cases sorted along the input beginning 25 cases from each end. To help stabilize curves, five internal replications were performed and averaged for each analysis.

ate each MAMBAC curve, we sorted all cases according to their scores on the input (from lowest to highest); this input indicator formed the *x*-axis for each graph and was labeled according to case numbers (there were 600 cases in the schizotypy data set). Then we located our first cutting score 25 cases from the lowest value on this summed input indicator. We subtracted the average output indicator score for all individuals falling below the cut from the average for all individuals falling above the cut. Because this low cutting score did not distinguish the schizotypes from the nonschizotypes (most schizotypes fall above the cut, but so do many nonschizotypes), the mean difference was relatively low. We plotted this value along the *y*-axis, then moved the cutting score to the right, calculated another mean difference, and plotted it. This process was repeated until the last of 50 evenly spaced cuts (which yielded easily interpretable curve shapes) was made 25 cases from the highest value on the input indicator, completing the analysis. As the cutting score approached the optimal value for differentiating the taxon and complement—a value relatively high on the input because there were considerably fewer schizotypes than nonschizotypes in this data set—the mean difference steadily increased. Our taxonic schizotypy data yielded MAMBAC curves that peaked at the rightmost end of the graph. Notably, the MAMBAC analysis using the summed input method (upper right panel of Figure 2) yielded a more clearly defined taxonic peak than the traditional MAMBAC analysis (upper left panel), suggesting that this indicator combination approach may improve MAMBAC's statistical power.

TABLE 2
Estimates of the Taxon Base Rate for Each Taxometric Procedure in Each Data Set

Taxometric Procedure	No. of Estimates	Schizotypy Data ^a		Taxpmoc Comparison Data ^a		Dimensional Comparison Data	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
MAMBAC, traditional	12	.32	.02	.32	.03	.45	.03
MAMBAC, summed input	4	.32	.01	.32	.05	.43	.01
MAXCOV	12	.13	.02	.11	.02	.16	.04
MAXEIG	4	.19	.01	.21	.02	.22	.06
L-Mode	3 ^b	.31	.05	.28	.04	.63	.34
<i>M</i> and <i>SD</i> ^c of these procedures' <i>Ms</i>		.25	.09	.25	.09	.38	.19

Note. MAMBAC = mean above minus below a cut; MAXCOV = maximum covariance; MAXEIG = maximum eigenvalue; L-Mode = latent mode.

^aThe actual taxon base rate in the simulated schizotypy and taxonic comparison data was .25.

^bEstimates were calculated from the location of both modes and from the empirical classification of cases. ^cThe consistency of taxon base-rate estimates across procedures may be inflated due to the inclusion of redundant analyses in the table: The variants of MAMBAC do not provide two independent estimates, nor do MAXCOV and MAXEIG provide independent estimates.

Although the estimates of the taxon base rate derived from these curves were slightly overestimated (*Ms* = .32 for both MAMBAC series; true base rate = .25), they were highly consistent across curves (see Table 2).

The second row of Figure 2 shows the MAMBAC curves for the taxonic comparison data. Just as in the schizotypy data set, these data yielded peaked curves, again more clearly when using the summed input indicator method. The third row of Figure 2 shows the MAMBAC results for the dimensional comparison data. Because there were no groups to be distinguished in this data set, no optimal cutting score existed. Hence, the dimensional latent structure produced concave curves with no discernible peaks. The clear contrast between the graphical results yielded by the taxonic and dimensional comparison data sets suggests that the schizotypy data—whose parameters they closely matched—were suitable for these MAMBAC analyses. The comparison data sets also provided a benchmark to aid interpretation: The schizotypy results were clearly more consistent with the results yielded by the taxonic than the dimensional comparison data.

Interpretive issues. Although taxonic and dimensional data typically produce dramatically different MAMBAC curve shapes, a number of cautionary notes are in order. First, as noted previously, taxonic data with a low base rate produce a MAMBAC peak toward the far right end of the curve. However, because the concave curves of dimensional data can also slope upward at one or both ends, it can sometimes be difficult to distinguish these curves from those yielded by a low base-rate taxon. Second, highly skewed indicators can alter the shape of MAMBAC curves (A. M. Ruscio & Ruscio, 2002; J. Ruscio, Ruscio, & Keane, 2003). Whether a construct is taxonic or dimensional in nature, positively skewed indicators of that construct produce curves that slope upward, whereas negatively skewed indicators produce curves that slope downward. This can be seen throughout Figure 2; all of the curves generally sloped upward, whether they were peaked or not. Although taxonic and dimensional data still produce different curve shapes under these conditions, the "tilting" effect of skew can complicate interpretation, particularly when a peak is suspected near either end of the curve. It is noteworthy that even with a taxon base rate as large as .25, the positive skew of the present indicators nearly obscured the taxonic peak; only the increased power of the summed input indicator method enabled the peak to emerge. Third, the taxon base-rate estimates were consistent (i.e., yielded small standard deviations) across the MAMBAC curves for all three data sets, even the dimensional comparison data (see Table 2). Indeed, converging MAMBAC base-rate estimates may be only weakly supportive of taxonic latent structure; highly discrepant estimates, however, would provide comparatively strong evidence against taxonic structure. Fourth, the taxon base rate was overestimated slightly in the schizotypy data as well as the taxonic comparison data. The tendency of MAMBAC to overestimate base rates was documented by Meehl and Yonce (1994) and can be exacerbated by nuisance covariance. Because the base rate is estimated using only

the endpoints of the MAMBAC curve, anything that tends to increase both of these values (e.g., nuisance covariance) will bias the base-rate estimate toward .50, resulting in a potentially substantial overestimation of the size of small taxa.

Fortunately, there are several ways to deal with these difficulties. First, whereas an individual MAMBAC curve may appear ambiguous, the panel of curves representing a series of MAMBAC analyses often exhibits a clearer pattern. Taxonic data consistently produce peaks in similar positions across MAMBAC curves, whereas dimensional data do not (see Meehl & Yonce, 1994). Second, interpretation may be greatly facilitated by the use of appropriate comparison curves (J. Ruscio et al., 2004). Although the MAMBAC curves yielded by the schizotypy data were highly consistent with one another, it was less clear whether they evidenced a right-end peak (suggesting latent taxa) or were merely concave (suggesting a latent dimension). Parallel analyses of comparison data revealed the schizotypy curves to be more consistent with a taxonic than a dimensional solution. Third, the use of additional taxometric procedures often helps to resolve apparent ambiguity in the results of any one procedure. Because factors such as indicator skew do not influence all taxometric analyses in the same way, cross-procedure consistency checks can clarify the process of interpretation. Likewise, the potential upward bias in MAMBAC estimates of the taxon base rate is not present in other procedures, which may instead tend to underestimate the base rate.

MAXCOV and MAXimum EIGenvalue (MAXEIG): Examining the Structural Sources of Covariance

These two highly related taxometric procedures examine interindicator associations within ordered subsamples of cases to test whether the indicators covary due to a mixture of latent taxa or due to the indicators' shared loadings on a latent dimension. Taxonic latent structure produces differing associations across subsamples in a manner consistent with the GCMT. Recall that the GCMT shows that the observed covariance between two indicators can be represented as the sum of two terms denoting the within-group covariances (the first two terms) and a third term denoting the product of the taxon and complement base rates and the indicator validities. Within any subsample that contains a relatively pure group of either taxon or complement members, there should be little association between indicators. This is because negligible nuisance covariance within putative groups is a prerequisite to conducting a taxometric analysis (keeping the first two terms of the GCMT low) and because either the taxon or the complement base rate will approach zero in a homogeneous subsample (keeping the third term low). In contrast, associations among indicators should be high in subsamples that contain a mixture of groups because taxon members tend to score high and complement members tend to score low on the indicators. In this case, the third term of the GCMT remains high because the product of the taxon and complement base rates does not approach

zero. The strength of the association should reach a maximum in the subsample containing an equal mixture of taxon and complement members (i.e., when $P = Q = .5$, which yields the maximal value of PQ). Thus, any measure of the association between indicators should vary in this predictable way across ordered subsamples of cases when latent taxa exist.

By contrast, dimensional latent structure should produce a degree of association between indicators that remains fairly constant across ordered subsamples. This is because in the absence of latent taxa, there is no reason for the strength of this association to systematically vary. For example, a scatterplot depicting the relationship between two indicators of a latent dimension takes the form of an oval-shaped cloud of points. Slicing this scatterplot with a series of vertical lines would produce subsamples of cases across which newly calculated correlations would roughly equal one another. Thus, when the latent structure is dimensional, any measure of the association between indicators should remain fairly constant across ordered subsamples.

The MAXCOV (Meehl & Yonce, 1996) procedure involves the simultaneous use of three valid indicators. One is treated as the input indicator and is placed on the x -axis, and the remaining two are treated as output indicators for the calculation of covariances. Cases are sorted by their scores on the input, which is then divided to form a series of subsamples. Within each subsample, the covariance between the output indicators is calculated and plotted as the corresponding y value. The shape of the resulting curve provides information about latent structure: Taxonic constructs yield peaked curves, whereas dimensional constructs yield nonpeaked (but not necessarily flat) curves.

The location of the point of maximum covariance along a MAXCOV curve hints at the size of the taxon base rate in much the same way as does the location of a peak in a MAMBAC curve. A peak toward the center of the MAXCOV curve indicates that the two latent groups are of equal size. For a small taxon, the peak will be deflected toward the right, whereas for a large taxon, it will be deflected toward the left (for details on base-rate estimation in MAXCOV, see Meehl & Yonce, 1996).

MAXEIG (Waller & Meehl, 1998) is a powerful multivariate extension of MAXCOV that works on the same fundamental principles. As in MAXCOV, one variable is selected to be the input indicator; however, all the rest are used as output indicators. Whereas MAXCOV is concerned with the covariance between two output indicators, MAXEIG is instead concerned with eigenvalues, the multivariate analog of covariances. MAXEIG measures the association between indicators as the first eigenvalue of the indicator covariance matrix (with variances along the diagonal replaced by zeros to leave only covariances). How is this eigenvalue comparable to a covariance? If the indicators are strongly related to one another, they will tend to load highly on the first principal factor, resulting in a high eigenvalue. If the variables are relatively unrelated to one another, they will not load highly on this factor, leading to a low eigenvalue.

Other than the substitution of eigenvalues for covariances, MAXEIG proceeds much like MAXCOV. Subsamples are formed according to their scores on an input indicator, and the association between the full set of output indicators is computed within each subsample. MAXEIG plots are interpreted in the same way as MAXCOV plots: Taxonic structure yields peaked curves, whereas dimensional structure does not, and the location of a peak can be used to estimate the taxon base rate and other latent parameters.

Implementation decisions. There are three significant implementation decisions to consider when performing MAXCOV or MAXEIG analyses. The first concerns how variables should be allocated to the required roles of input and output indicators. There are a number of ways to do this, each similar to a method described previously for MAMBAC. For example, one could use the variables in all possible input–output–output indicator triplets or remove a pair of variables to serve as outputs and combine the remainder to form the input. At least two output indicators must be available for each curve. If there are only two, one can calculate either covariances or eigenvalues; if there are more than two output indicators, only eigenvalues can be calculated.

The second decision concerns how cases should be divided into subsamples along the input indicator. This problem is similar to the location of cutting scores in a MAMBAC analysis but with the additional concern that sampling error within each subsample may obscure the shape of a MAXCOV/MAXEIG curve and reduce the accuracy with which latent parameters are estimated. Thus, a balance must be struck between the number of subsamples (more subsamples result in better delineation of curve shape) and the amount of sampling error (more subsamples result in fewer cases per subsample for covariance or eigenvalue estimation). Subsamples may be formed on the basis of intact scale values, standard deviation units, or fixed numbers of cases. If fixed numbers of cases are selected, subsamples can be constructed using either nonoverlapping intervals or overlapping windows. For example, a sample of 600 cases may be divided into six intervals of $n = 100$ (Cases 1 to 100, 101 to 200, ... 501 to 600) or into 51 windows of $n = 100$ that overlap 90% with adjacent windows (Cases 1 to 100, 11 to 110, 21 to 120, ... 501 to 600); one can also vary the degree of overlap between windows.

Traditionally, MAXCOV has been performed using intervals, and MAXEIG has been performed using windows. However, there is nothing to prevent researchers from choosing between these approaches for either MAXCOV or MAXEIG. Windows afford far more points on a curve with no increase in sampling error within subsamples, providing a no-cost increase in the interpretability of curve shape. As in MAMBAC, if the dividing lines between subsamples fall between equal-scoring cases, internal replications can be performed to further improve the interpretability of the curve and the accuracy with which latent parameters are estimated.

One unique approach to dividing cases into subsamples is to repeat the analysis for a given indicator configuration using increasing numbers of subsamples. Waller and Meehl (1998) introduced this technique and dubbed it the "Inchworm Consistency Test." This procedure is particularly helpful in dealing with the challenge of distinguishing a small taxon from a latent dimension whose indicators are positively skewed when it can be exceedingly difficult to determine whether a right-end cusp represents a taxonic peak or merely the increasing curve associated with indicator skew (A. M. Ruscio & Ruscio, 2002; J. Ruscio, Ruscio, & Keane, 2003). As the number of windows is increased in a MAXCOV or MAXEIG analysis, taxonic structure should yield an increasingly well-defined peak. This occurs for the following reason: Whereas the members of a small taxon may be outnumbered by complement members in even the highest scoring of large subsamples, an increase in the number of subsamples leaves fewer cases in each, eventually allowing even a small taxon to outnumber complement members in the uppermost subsamples. As taxon members come to outnumber complement members in the highest scoring subsamples, the resulting downward slope of the curve will cause the taxonic peak to become more clearly defined. MAXCOV or MAXEIG can therefore be especially informative when they are performed with increasing numbers of subsamples and when these results are compared across analyses to aid interpretation. At the same time, it is important to note that sampling error resulting from too many (and hence too small) subsamples will limit the size of a taxon that can be detected. Internal replications can mitigate against this loss of sensitivity, but there nonetheless must be some lower limit to the size of a detectable taxon.

The third decision concerns how results should be plotted. Although we advocate a narrow scaling of the y -axis for MAMBAC analyses to facilitate the distinction between concave and convex curves, this may be inappropriate for MAXCOV or MAXEIG, as it can exaggerate fluctuations in covariances or eigenvalues due to normal sampling error and create the appearance of a peaked curve. On the other hand, an overly broad range of y values may artificially flatten even a genuinely peaked curve. Thus, care should be taken to select a reasonable y scale that does not distort the results. To do so, researchers may need to override the default procedures of computerized graphing programs, which often apply too narrow a range of y values.

A related presentational issue involves curve smoothing, which is controversial because of its potentially marked impact on the graphs from which interpretations are drawn. When a MAXCOV or MAXEIG analysis is carefully performed with suitable data, using appropriate techniques to allocate variables to input–output roles and to divide cases into subsamples, interpretable curves can almost always be obtained without smoothing. However, smoothed values may yield more accurate taxon base-rate estimates. If the choice is made to use a smoothing procedure, care should be taken to use an appropriate method that evens out but does not distort the shape of the curve (e.g., Hartwig & Dearing, 1979; Tukey, 1977). In addi-

tion, we suggest presenting both smoothed and unsmoothed curves so that readers can evaluate both for themselves.

Empirical illustration. To illustrate these procedures, we performed MAXCOV and MAXEIG analyses of the schizotypy data. MAXCOV was conducted using the four indicators in all input-output-output triplets ($n = 12$ curves), with the input indicator divided into subsamples consisting of 10 equal-sized intervals (i.e., deciles). MAXEIG was conducted using one indicator at a time as the input and the remaining three indicators as outputs ($n = 4$ curves), with the input indicator divided into subsamples consisting of 50 windows overlapping 90% with one another.

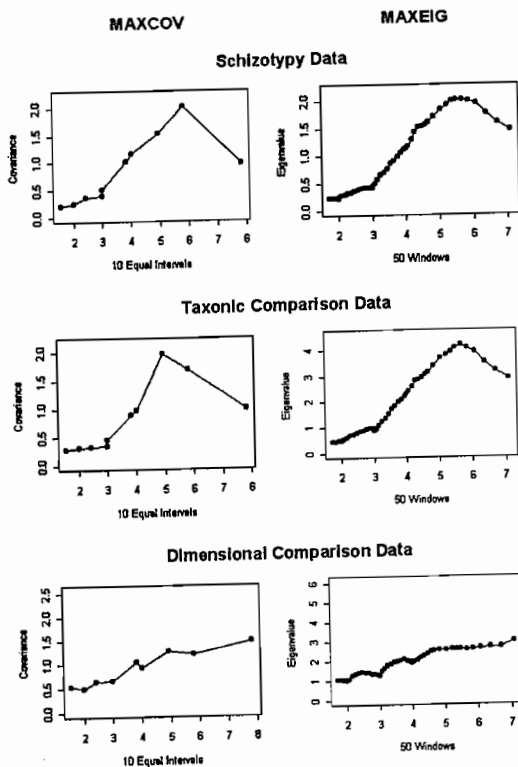


FIGURE 3 Maximum covariance (MAXCOV) and maximum eigenvalue (MAXEIG) plots for the schizotypy data plus the taxonic and dimensional comparison data. MAXCOV was conducted using the four indicators in all possible input-output-output triplets ($n = 12$ curves), with the input subdivided into deciles. MAXEIG was conducted first using one indicator as the input and the remaining three indicators as outputs ($n = 4$ curves), with the input subdivided into 50 windows that overlapped 90%. To help stabilize curves, five internal replications were performed and averaged for each analysis.

Figure 3 shows the MAXCOV (left column) and MAXEIG (right column) results for the schizotypy data (top row), taxonic comparison data (middle row), and dimensional comparison data (bottom row). Whereas all curves increased from left to right due to positive indicator skew, results for the taxonic comparison data were easily distinguished from the results for the dimensional comparison data. It is noteworthy that the Inchworm Consistency Test was not needed in this case. Although MAMBAC seemed strained to the limits of its resolving power by the parameters of these indicators (e.g., their positive skew), both MAXCOV and MAXEIG easily differentiated taxonic from dimensional structure using the original number of intervals or windows included in the analysis. Thus, the schizotypy data were considered suitable for these analyses, and results of the schizotypy analyses were judged to be taxonic.

The taxon base rate was underestimated for the schizotypy data set and estimates were quite consistent within and between all three data sets (see Table 2). Thus, whereas MAMBAC base-rate estimates exhibited an upward bias, MAXCOV and MAXEIG's base-rate estimates were biased downward in the presence of positive skew, which deflects taxonic peaks to the right. As these analyses illustrate, an average base-rate estimate computed across MAMBAC and MAXCOV/MAXEIG procedures may be more accurate than the estimate yielded by any individual procedure not only because random errors of estimation will tend to cancel out, but because slight biases in the procedures may also tend to cancel out. Finally, it is noteworthy that taxon base-rate estimates were once again highly consistent within and between all data sets, including the dimensional comparison data. As with MAMBAC, consistency among MAXCOV or MAXEIG base-rate estimates may provide only weak support for a taxonic conclusion, whereas consistency with other procedures may provide stronger evidence for taxonicity (see the standard deviations in the bottom row of Table 2 for an index of cross-procedure coherence).

Choosing a procedure. Because MAXCOV and MAXEIG are premised on a similar rationale, the evidence that they contribute to a taxometric investigation is largely redundant. Rather than choosing between MAXCOV and MAXEIG, it may be more useful to regard these procedures as variants of one common procedure that is most appropriately implemented by working through the decision points outlined previously. That is, one must decide how to allocate variables to the required input and output indicator roles, whether to calculate covariances or eigenvalues, how to divide cases into subsamples along the input indicator, and how to plot the results. Depending on the choices that one makes, the resulting technique might resemble a traditional MAXCOV analysis, a traditional MAXEIG analysis, or a blend of the two. Refinements and extensions such as the summed input indicator method and internal replications may further improve the ability of the MAXCOV and MAXEIG procedures to distinguish taxonic from dimensional latent structure.

Latent Mode (L-Mode): Exploring the Distribution of Factor Scores

The final procedure that we discuss differs from those presented previously in that it is not based on the coherent cut kinetics approach to taxometrics. Instead, L-Mode is a taxometric procedure based on factor analysis that tests latent structure in a manner more closely resembling finite mixture models. Traditionally, factor analysis has been used almost exclusively to explore the dimensionality of psychological constructs, most often to resolve the number of latent dimensions underlying variation in responses to a large number of items. In contrast, L-Mode seeks to differentiate taxonic and dimensional latent structure by plotting the distribution of individuals' scores on a single latent factor calculated through factor analysis (constrained to a one-factor solution). To the extent that indicators covary due to the mixture of latent taxa, factor scores will more validly separate a taxon and complement than will any

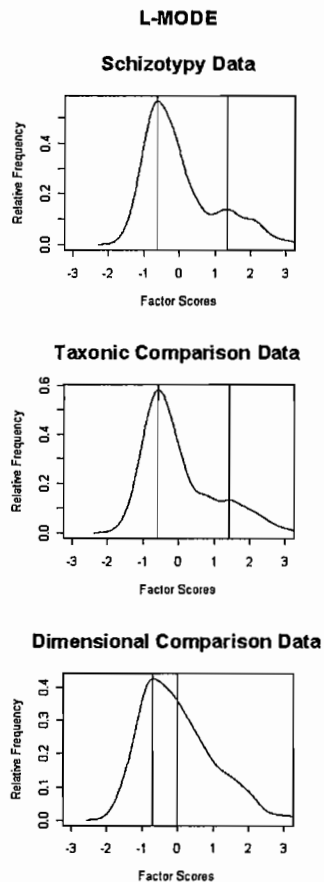


FIGURE 4 Latent mode (L-Mode) plots for the schizotypy data plus the taxonic and dimensional comparison data. Vertical lines mark visually apparent modes to the left and right of $x = 0$.

individual indicator. With enough indicators of sufficient validity and negligible nuisance covariance, taxonic data will yield a bimodal distribution of factor scores, whereas dimensional data will yield a unimodal distribution. Because L-Mode automatically involves all available indicators in the calculation of one frequency distribution, there are no significant decisions required to implement the procedure.

The multivariate power of L-Mode can be seen by comparing the manifest distributions of the schizotypy indicators (the first of which is shown in grey in Figure 1) with the distribution of factor scores generated using L-Mode (Figure 4, top graph). Whereas the schizotypy indicators were distributed unimodally, the L-Mode plot for schizotypy clearly revealed the presence of two groups as did the plot for the taxonic comparison data (middle graph). In contrast, the plot for the dimensional comparison data (bottom graph) was unimodal. The bimodality of taxonic L-Mode graphs becomes more striking as the quantity and validity of indicators increase (Waller & Meehl, 1998).

L-Mode also provides several methods for estimating the taxon base rate (Waller & Meehl, 1998). First, one estimate can be derived from the location of each estimated latent mode (one to the left of $x = 0$ and one to the right). If these two estimates agree well with one another, they can be averaged to provide a single value for subsequent use. Second, based on his or her profile of scores on all available indicators, each case in the sample can be classified as a likely taxon or complement member. By calculating the proportion of cases classified into the taxon, an additional base-rate estimate can be obtained. In this example, the taxon base-rate estimates for the schizotypy and taxonic comparison data were accurate and consistent, whereas those for the dimensional comparison data were markedly inconsistent (see Table 2).

Because L-Mode distinguishes taxonic from dimensional latent structure in a manner more akin to finite mixture models than to coherent cut kinetics procedures, it may be susceptible to some limitations of the former. Nonetheless, L-Mode does provide independent estimates of the taxon base rate that can be usefully compared to those obtained from other taxometric procedures. More research is needed to explore the strengths and limitations of L-Mode.

A PRACTICAL EXAMPLE: CLASSIFYING CASES INTO THREE GROUPS

In the preceding discussion, we described the use of taxometric procedures to distinguish taxonic from dimensional latent structure, but the method can do more than this. To provide a brief illustration of the power of taxometric procedures to resolve more complex structures and accurately assign cases to multiple latent classes, we created a data set based on a simple hierarchical structure: A taxon ($n = 100$) that contains two equal-sized subtypes (Type 1 and Type 2) is distinct from its

complement class ($n = 200$). By drawing random normal deviates from specified population values, three indicators were created to validly distinguish the taxon ($\mu = 20, \sigma = 5$) and complement ($\mu = 10, \sigma = 5$). Three additional indicators were created to validly distinguish the two subtypes within the taxon ($\mu = 10$ for Type 1; $\mu = 20$ for Type 2; $\sigma = 5$ for each) and take on intermediate values in the complement ($\mu = 15, \sigma = 10$). The sample size is relatively small for taxometric analyses, particularly the distinction between the two subtypes within the taxon. However, the challenge that this poses is mitigated by the high indicator validity ($d = 2.00$) and absence of nuisance covariance.

Because cluster analyses are often used to study data structures like this, we began by performing hierarchical, agglomerative cluster analyses in a traditional

manner: We submitted all six indicators to analyses using squared Euclidean distances and either the average linkage method (Sokal & Michener, 1958) or Ward's (1963) method. Because stopping rules are poor guides to the appropriate number of clusters, we evaluated the classification accuracy of two-, three-, and four-cluster solutions. The results are shown in Figure 5.

Whereas the average linkage method repeatedly lumped nearly all cases into a single large cluster, the two-cluster solution using Ward's method achieved a reasonably accurate distinction between the overarching taxon and complement classes. Because only two clusters were formed, this solution could not make the additional distinction between the two subtypes within the taxon. Allowing Ward's method to form additional clusters did not help matters, as the accuracy of the three- and four-cluster solutions suffered from the well-known bias of this method toward equal-sized clusters: A large number of complement members was diverted into Type 1 or Type 2. Thus, none of the cluster analyses adequately captured the true latent structure of the data. This is a discouraging showing, as the indicators do in fact differ by a substantial amount across fairly large groups within a simple structural arrangement.

The poor performance of these cluster analyses may be attributed in part to a potentially unreasonable demand placed on them: simultaneously differentiating multiple latent classes using indicators that are each relevant to one latent boundary but not another. Whereas clustering algorithms are ordinarily set to work using all available indicators in a single analysis, the taxometric method requires a more careful selection of indicators that represent one putative boundary per analysis. Thus, we performed taxometric analyses (MAMBAC, MAXEIG, and L-Mode) in a two-stage process. First, each procedure was conducted using the three indicators that distinguish the overarching taxon and complement. As can be seen in Figure 6, all of these results were clearly taxonic. The average estimate of the taxon base rate was used to assign the cases scoring the highest on the sum of these three indicators to the taxon until a proportion of cases equal to its estimated base rate was reached. Second, each procedure was conducted within the subsample of cases that had been assigned to the taxon in the first stage, this time using the three indicators that distinguish the subtypes of the taxon. Once again, all results were clearly taxonic, and cases were assigned to the two subtypes as described previously. Not only did each taxometric procedure successfully capture the true latent structure of these data, but the average taxon base-rate estimates perfectly reflected the actual group sizes (200, 50, 50), and an impressive accuracy was achieved in classifying cases (95.3%). This accuracy level was greater than that achieved by applying Bayes' theorem using the results of MAXEIG analyses (90.7%) but less than that achieved by using Waller and Meehl's (1998) profile similarity technique based on the results of L-Mode analyses (96.7%).

Finally, whereas cluster analysis is ordinarily conducted using all available indicators in a single analysis, we also performed it in two stages to enable com-

CLASSIFICATION RESULTS FOR CLUSTER ANALYSES

AVERAGE LINKAGE METHOD, ONE-STAGE

	Cluster 1	Cluster 2	Total
Comp.	197	3	200
Type 1	50	0	50
Type 2	50	0	50
Total	297	3	300

N = 197 correct (65.7%)

	Cluster 1	Cluster 2	Cluster 3	Total
Comp.	197	2	1	200
Type 1	50	0	0	50
Type 2	50	0	0	50
Total	297	2	1	300

N = 197 correct (65.7%)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Comp.	197	2	1	0	200
Type 1	50	0	0	0	50
Type 2	49	0	0	1	50
Total	296	2	1	1	300

N = 198 correct (66.0%)

WARD'S METHOD, ONE-STAGE

	Cluster 1	Cluster 2	Total
Comp.	196	4	200
Type 1	3	47	50
Type 2	8	42	50
Total	207	93	300

N = 243 correct (81.0%)

	Cluster 1	Cluster 2	Cluster 3	Total
Comp.	137	59	4	200
Type 1	2	1	47	50
Type 2	6	2	42	50
Total	145	62	93	300

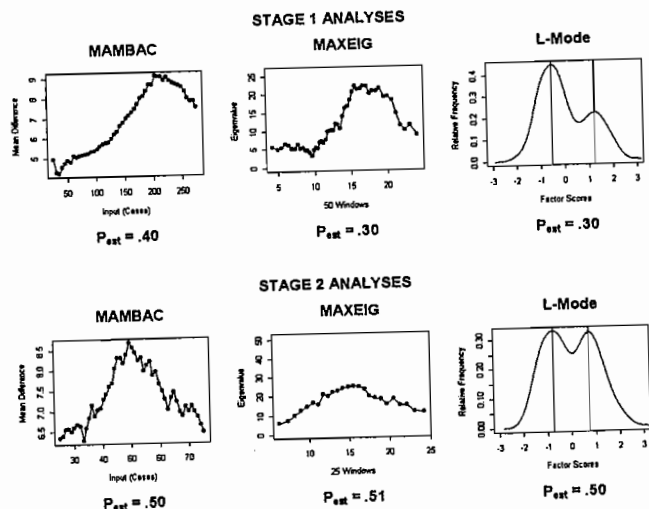
N = 186 correct (62.0%)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Comp.	73	59	64	4	200
Type 1	1	1	1	47	50
Type 2	2	2	4	42	50
Total	76	62	69	93	300

N = 124 correct (41.3%)

FIGURE 5 Classification tables for cluster analyses using average linkage or Ward's method. Numbers in bold print represent correct classifications. Comp = complement class.

CLASSIFICATION RESULTS FOR TAXOMETRIC ANALYSES



CLASSIFICATION RESULTS				
	Group 1	Group 2	Group 3	Total
Comp.	196	2	3	200
Type 1	2	46	2	50
Type 2	3	2	45	50
Total	200	50	50	300

N = 288 correct (96.3%)

FIGURE 6 Classification tables for taxometric analyses. Stage 1 analyses used three indicators to make the distinction between the complement (Comp) and the taxa, and Stage 2 analyses used three new indicators to make an additional distinction among cases that were classified into the taxon in the Stage 1 analysis. Numbers in bold print represent correct classifications.

parisons with the two-stage taxometric approach. Analyses using the average linkage method and Ward's method accurately revealed the structure of these data, reasonably reflected the three groups' sizes (205, 47, and 43 for average linkage; 208, 49, and 43 for Ward's method), and achieved classification results (94.3% correct for each method) that approached the accuracy of the taxometric analyses (see the bottom of Figure 6). It is noteworthy that cluster analyses are capable of impressive performance when they are conducted in successive stages using only the indicators relevant to one particular boundary between two latent classes at each stage. We believe that the taxometric method's emphasis on testing one latent boundary at a time represents a significant strength in that it prevents researchers from submitting a large, heterogeneous set of indicators to a single analysis and expecting the statistical procedure to successfully reveal a complex structure and accurately classify cases.

IMPLEMENTING TAXOMETRIC PROCEDURES

We have demonstrated the power of the taxometric method and outlined some of the most important decisions that must be made to successfully implement each taxometric procedure. In doing so, we aimed to offer a few pieces of general advice without suggesting that any particular strategies are universally applicable or acceptable. There are in fact so many relevant decisions to make when conducting a taxometric investigation that a "one-size-fits-all" approach would be inappropriate. It would also be impractical, if not impossible, to conduct Monte Carlo studies that systematically test every realistic configuration of data parameters and analytic choices that researchers might encounter. Therefore, we urge researchers to simulate taxonomic and dimensional comparison data and to subject these data to multiple taxometric analyses, varying the choices that are made for each procedure and noting what influence, if any, each choice has on the results. The goal is to determine how best to configure each taxometric procedure so that its results best discriminate the taxonomic from the dimensional comparison data. After the researcher has determined how best to set up each analysis to maximally differentiate the two latent structures given the specific parameters of the available research data, these data can be subjected to precisely that analysis to see which structural solution is best supported. By intentionally separating the analysis of simulated comparison data from the analysis of research data, one can circumvent the capitalization on chance that occurs when many analyses are performed on a set of research data while avoiding the temptation to select the results most favorable to the hypothesized latent structure. For further details on the simulation and use of comparison data, see J. Ruscio et al. (2004).

Consistency Tests

A cornerstone of the taxometric method is the evaluation of agreement of results across as many procedures and consistency tests as possible. In addition to employing multiple taxometric procedures on the available data—an essential requirement of any taxometric investigation—users of the method can evaluate consistency in several other ways. These include the repeated application of each taxometric procedure with different indicator configurations as well as the comparison of base-rate estimates (or other latent parameters) within and across procedures. We discuss each of these approaches along with two consistency tests that may provide less evidentiary support than is often presumed.

Repeated Application of Each Taxometric Procedure

As was mentioned earlier, most taxometric procedures can be performed multiple times by using the available indicators in all possible combinations. The resulting

panels of curves can be compared and interpreted as a whole as can the multiple taxon base-rate estimates yielded by each set of analyses. Moreover, each procedure can be performed on multiple sets of indicators derived from the research data.

Comparison of Base-Rate Estimates

Large numbers of base-rate estimates can be readily summarized and evaluated for consistency. Table 2 displays the base-rate estimates yielded by all analyses of the schizotypy and comparison data sets. Because some procedures tend to yield consistent estimates regardless of latent structure (e.g., when indicators are skewed), it may be that the standard deviation of taxon base-rate estimates across procedures is a more useful measure of consistency than the standard deviation of estimates within a procedure. For the schizotypy data set, the agreement of base-rate estimates can be most parsimoniously explained as denoting an actual latent taxon. Moreover, the average estimate of .25 matched the true taxon base rate. In analyses of the taxonic comparison data, the estimates were likewise consistent across procedures and highly similar to the base rate that had been simulated, whereas in analyses of the dimensional comparison data, the estimates were often consistent within procedures but much less so across procedures. The most parsimonious explanation for this discrepancy is that no real groups were detected by analyses of the dimensional data.

Additional estimates of the taxon base rate can be obtained through reanalysis of one's data following a targeted removal of cases (see J. Ruscio, 2000). This approach is based on the logic that the removal of specific cases from the data set should result in a predictable change in the taxon base rate. For example, if taxometric analysis in a full sample of data yields a taxon base rate estimate of .20, reanalysis after removing the lower quartile of cases (which presumably contains virtually no members of the putative taxon) would be expected to yield a new base rate estimate of $.20 / .75 = .27$ in the reduced sample. A result that diverges markedly from this expected value would argue against taxonic structure. Finally, one can also compare base-rate estimates derived from taxometric analysis to independent, externally derived estimates. For example, when evaluating the structure of a mental disorder, one might use its rate of diagnosis in the sample (based on clinical interviews or other data) as a plausible base-rate estimate to compare to those derived from taxometrics (e.g., A. M. Ruscio et al., 2002; J. Ruscio & Ruscio, 2000).

Distribution of Bayesian Probabilities

Another frequently used consistency test involves examining the distribution of Bayesian probabilities of taxon membership. To perform this test, one calculates the probability of taxon membership for each case in the data set using Bayes' theorem, then inspects the distribution of these probabilities. The rationale is that if cases pile up near probabilities of 0 or 1, with few cases scoring at intermediate val-

ues, this supports a taxonic interpretation (Waller & Meehl, 1998). Unfortunately, a U-shaped distribution of Bayesian probabilities is readily obtained even with dimensional latent structure. The reason for this is that Bayes' theorem—as it is presented and utilized in taxometric research—involves the assumption of conditional independence. This requires that variables be unrelated within groups, an assumption that is violated with dimensional structure because there is only one group (the full sample) within which the indicators covary. Because of these interindicator correlations, individuals scoring above the threshold on one indicator will also tend to score above the threshold on many or most other indicators (and vice versa for those below threshold). Thus, probabilities can accumulate near 0 and 1 regardless of latent structure.

FREQUENCY DISTRIBUTIONS OF BAYESIAN PROBABILITIES OF TAXON MEMBERSHIP

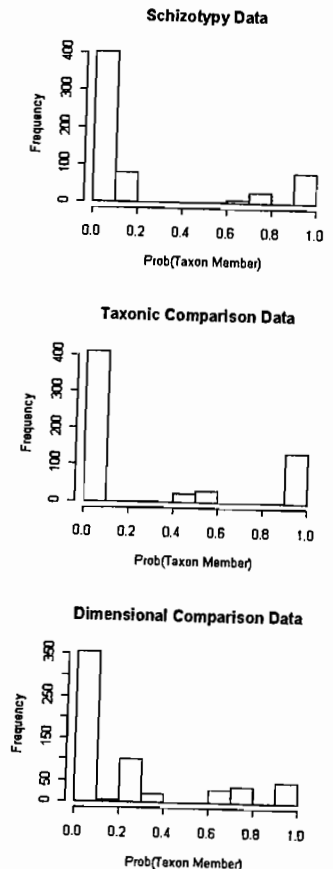


FIGURE 7 Distributions of Bayesian probabilities of taxon membership for the schizotypy data plus the taxonic and dimensional comparison data. Note that all three distributions are U shaped and would likely be interpreted as evidence supportive of taxonic structure.

To illustrate the potentially low specificity of U-shaped distributions to taxonic latent structure, results from the MAXCOV analyses reported earlier were used to generate the distribution of Bayesian probabilities of taxon membership for the schizotypy data as well as the taxonic and dimensional comparison data (see Figure 7). All three distributions appeared U shaped, with many more cases' probabilities falling close to 0 than close to 1 because of the low taxon base rate in the sample. Even the distribution for the dimensional comparison data might have led to an inference of taxonic latent structure. Although there may be differences in the extent to which these distributions can be characterized as U shaped or in the frequency with which such distributions are obtained from taxonic versus dimensional data, there is clearly the potential for ambiguity and misinterpretation when using this consistency test. Until research clarifies the sensitivity and specificity of this test, we suggest that researchers avoid it or interpret its results with caution.

GFI

One final consistency test is worthy of mention because its use appears to be on the rise in taxometric investigations despite recent research calling its utility into question. Waller and Meehl (1998) introduced a Goodness-of-Fit Index (GFI) that tests the fit of output yielded by the MAXCOV procedure to a taxonic structural model. After conducting a small-scale Monte Carlo study, Waller and Meehl reported that taxonic data usually yielded GFI values above .90, whereas dimensional data seldom did so. Unfortunately, no information was provided regarding the parameters of the data employed in the study (e.g., sample sizes, base rates, indicator validities). Based on subsequent Monte Carlo research (Cleland, Rothschild, & Haslam, 2000; Haslam & Cleland, 2002), two tentative conclusions can be reached. First, the GFI poorly discriminated taxa from dimensions. Second, no particular cutoff was found to be universally optimal for differentiating these two structures. For some parametric configurations, both taxonic and dimensional data tended to yield GFI values below .90, whereas for other configurations, both types of data tended to yield values above (sometimes well above) .90. Although several recent taxometric investigations have relied on this GFI as a source of structural evidence, Monte Carlo research suggests that it may be best to suspend the use of this index unless it is refined and demonstrated to validly distinguish competing structures.

INFORMATIVE TAXOMETRIC INVESTIGATIONS

Many fields of study stand to profit from empirically evaluating the latent structure of the constructs that they study. Knowledge of latent structure has the potential to simplify communication, identify more accurate taxonomies, improve the reliability and validity of assessment devices, and increase the statistical power of research. In these and many other ways, latent structure poses an empirical question

fundamental to both theory and application, and the taxometric method provides a powerful tool to test it.

Our primary goal in this nontechnical introduction has been to promote a conceptual understanding of the taxometric method and the analytic procedures that it subsumes. A working knowledge of taxometrics is important not only for researchers who use the method but also those who wish to be informed readers, reviewers, and consumers of this research. In this spirit, we conclude with a checklist of five general questions that we contend should be thoughtfully considered and explicitly addressed when reporting or evaluating any taxometric investigation. These questions are designed to highlight the potential limitations of taxometrics when certain conceptual and methodological requirements of the method and its constituent procedures are not met.

Question 1: Is a Taxometric Analysis Scientifically Justified?

Haslam and Kim (2002) documented a recent surge in taxometric investigations, many of which are being published in top-ranked psychological journals. As with any statistical tool that begins to enjoy increased attention, the taxometric method runs the risk of becoming a "flavor of the month" analytic approach that is applied to any data set that appears suitable. Investigators must be careful to avoid the temptation of conducting a taxometric study simply because it appears trendy or publishable. As with any other type of research, it is incumbent on investigators to provide a compelling scientific rationale for the appropriateness of the taxometric method for their specific research questions.

Question 2: Are the Data Suitable for Taxometric Analysis?

Before engaging in taxometric analysis, it is the investigator's responsibility to demonstrate that his or her data meet certain basic requirements. Researchers run the risk of reaching unfounded conclusions if they fail to establish prior to analysis that their data are sufficiently valid for the planned taxometric analyses. For example, a latent taxon may go undetected if there are too few total cases of data or too few taxon members in the sample or if the available indicators do not adequately define the construct of interest, are too few in number, vary across too narrow a range of values, are insufficiently valid, or are overly redundant with one another. Likewise, artificially constructed or highly selected samples as well as skewed indicators raise the likelihood of detecting spurious pseudo-taxa and reaching incorrect conclusions about latent structure. We have described many of the relevant considerations for establishing the suitability of data for taxometric analysis, and we encourage researchers to pay careful attention to this important but often overlooked aspect of taxometric research. Without convincing readers that the data can

provide a powerful test between competing latent structures, the results of taxometric analyses may have little theoretical or practical value for the field.

Question 3: Has a Sufficient Variety of Procedures Been Implemented Properly?

Because the taxometric method depends so heavily on consistency checks, any taxometric investigation must include multiple procedures to build confidence in the obtained structural solution. Although the nature of the available data may constrain the analytic options available to researchers, a variety of procedures and consistency tests should always be selected with an eye toward obtaining nonredundant information. For example, once MAXCOV has been conducted, MAMBAC would add some independent evidence, whereas MAXEIG would not.

Also, because most taxometric procedures can be conducted in a variety of ways, it is essential to make informed choices at each decision point. For example, researchers should clearly specify the methods by which variables were assigned to input-output roles, the methods by which cutting scores were located or subsamples formed along the input for each analysis, and so forth. Ideally, researchers would also explain their reasons for having made these choices.

Question 4: Have the Results Been Presented and Interpreted Appropriately?

There are a number of issues to consider when presenting taxometric results. For example, researchers should think through the most appropriate scaling of y-axes and the need for and consequences of smoothing curves. Moreover, although it may go without saying, it is important to fully label all graphical results so that it is perfectly clear what is being graphed. Many taxometric reports have included graphs whose axes have been inadequately labeled or whose figure captions have failed to indicate how a given analysis was performed (e.g., what variables were assigned to be input and output indicators, how cuts were made or subsamples formed). It is still more problematic when the text does not clarify which results are being presented in a given graph, whether the presentation includes all or a subset of the results, and which procedures were used to generate which results.

It is also critical that the full set of taxometric results be interpreted carefully. The nature of the latent construct—which is represented by the manifest indicators—should be considered and discussed along with alternative explanations for apparently taxonic or dimensional results (e.g., selected samples, indicator skew). Parallel analyses of simulated taxonic and dimensional comparison data can serve as a particularly valuable interpretive aid allowing readers not only to judge for themselves whether the research data are suitable for particular taxometric analyses but also whether the proffered structural conclusions appear reasonable. If the

curve shapes, taxon base-rate estimates, and other results yielded by the research data appear much more consistent with those of one comparison data set than of another, readers can readily verify the appropriateness of the authors' conclusions.

Question 5: Are Implications of the Findings Clearly Articulated?

Finally, a researcher's responsibility does not end with the conclusion that a construct's latent structure is best characterized as taxonic or dimensional. To be useful, a taxometric report should spell out the implications of this structural finding for future research and practice. Perhaps the taxometric analysis is part of a broader research program investigating the latent structure of a particular construct or group of related constructs, or perhaps it has implications for extant theories of the target construct or for the manner in which this construct is assessed or studied. Whatever the scientific issues that originally prompted the taxometric investigation, they should be revisited at this stage to consider and build on the advances provided by the taxometric results.

ACKNOWLEDGMENTS

This work was supported in part by National Research Service Award MH12675 to Ayelet Meron Ruscio from the National Institute of Mental Health. We thank Del Ellsworth for his valuable comments on an early draft of this manuscript.

REFERENCES

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Arabie, P., Hubert, L. J., & DeSoete, G. (Eds.). (1996). *Clustering and classification*. River Edge, NJ: World Scientific.
- Bartholomew, D. J. (1987). *Latent variable models and factor analysis*. New York: Oxford University Press.
- Beauchaine, T. P., & Beauchaine, R. J. (2002). A comparison of maximum covariance and k-means cluster analysis in classifying cases into known taxon groups. *Psychological Methods*, 7, 245–261.
- Bock, H. (1996). Probability models and hypothesis testing in partitioning cluster analysis. In P. Arabie, L. J. Hubert, & G. DeSoete (Eds.), *Clustering and classification* (pp. 377–453). River Edge, NJ: World Scientific.
- Cleland, C. M., Rothschild, L., & Haslam, N. (2000). Detecting latent taxa: Monte Carlo comparison of taxometric, mixture model, and clustering procedures. *Psychological Reports*, 87, 37–47.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Everitt, B. S. (1993). *Cluster analysis* (3rd ed.). London: Arnold.
- Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. New York: Chapman & Hall.
- Fraley, R. C., & Waller, N. G. (1998). Adult attachment patterns: A test of the typological model. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 77–114). New York: Guilford.

- Golden, R. R., & Meehl, P. E. (1979). Detection of the schizoid taxon with MMPI indicators. *Journal of Abnormal Psychology, 88*, 217-233.
- Grayson, D. A. (1987). Can categorical and dimensional views of psychiatric illness be distinguished? *British Journal of Psychiatry, 151*, 355-361.
- Green, B. F. (1951). A general solution for the latent class model of latent structure analysis. *Psychometrika, 16*, 151-166.
- Grove, W. M. (1991). The validity of cluster analysis stopping rules as detectors of taxa. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1, pp. 313-329). Minneapolis: University of Minnesota Press.
- Grove, W. M., & Meehl, P. E. (1993). Simple regression-based procedures for taxometric investigations. *Psychological Reports, 73*, 707-737.
- Hartwig, F., & Dearing, B. E. (1979). *Exploratory data analysis*. Beverly Hills, CA: Sage.
- Haslam, N. (1997). Evidence that male sexual orientation is a matter of degree. *Journal of Personality and Social Psychology, 73*, 862-870.
- Haslam, N. (1999). Taxometric and related methods in relationships research. *Personal Relationships, 6*, 519-534.
- Haslam, N., & Cleland, C. (2002). Taxometric analysis of fuzzy categories: A Monte Carlo study. *Psychological Reports, 90*, 401-404.
- Haslam, N., & Kim, H. (2002). Categories and continua: A review of taxometric research. *Genetic, Social, and General Psychology Monographs, 128*, 271-320.
- Korfine, L., & Lenzenweger, M. F. (1995). The taxonicity of schizotypy: A replication. *Journal of Abnormal Psychology, 104*, 26-31.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Lenzenweger, M. F. (1999). Deeper into the schizotypy taxon: On the robust nature of maximum covariance analysis. *Journal of Abnormal Psychology, 108*, 182-187.
- Lenzenweger, M. F., & Korfine, L. (1992). Confirming the latent structure and base rate of schizotypy: A taxometric analysis. *Journal of Abnormal Psychology, 101*, 567-571.
- Lorr, M. (1994). Cluster analysis: Aims, methods, and problems. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 179-195). New York: Springer.
- McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering*. New York: Dekker.
- McLachlan, G. J., & Peel, D. (2001). *Finite mixture models*. New York: Wiley.
- Meehl, P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist, 17*, 827-838.
- Meehl, P. E. (1967). Theory testing in psychology and in physics: A methodological paradox. *Philosophy of Science, 34*, 103-115.
- Meehl, P. E. (1986). Diagnostic taxa as open concepts: Metatheoretical and statistical questions about reliability and construct validity in the grand strategy of nosological revision. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology: Toward the DSM-IV* (pp. 215-231). New York: Guilford.
- Meehl, P. E. (1990). Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. *Journal of Personality Disorders, 4*, 1-99.
- Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality, 60*, 117-174.
- Meehl, P. E. (1995a). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50*, 266-274.
- Meehl, P. E. (1995b). Extension of the MAXCOV-HITMAX taxometric procedure to situations of sizable nuisance covariance. In D. Lubinski & R. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 81-92). Palo Alto, CA: Consulting Psychologists Press.
- Meehl, P. E. (1999). Clarifications about taxometric method. *Applied and Preventive Psychology, 8*, 165-174.
- Meehl, P. E., & Golden, R. R. (1982). Taxometric methods. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127-181). New York: Wiley.
- Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports, 74*, 1059-1274.
- Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports, 78*, 1091-1227.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166.
- Milligan, G. W. (1996). Clustering validation: Results and applications for applied analyses. In P. Arabie, L. J. Hubert, & G. DeSoete (Eds.), *Clustering and classification* (pp. 341-375). River Edge, NJ: World Scientific.
- 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.
- Murphy, E. A. (1964). One cause? Many causes? The argument from the bimodal distribution. *Journal of Chronic Disease, 17*, 301-324.
- Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1-33). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Ruscio, A. M., & Ruscio, J. (2002). The latent structure of analogue depression: Should the BDI be used to classify groups? *Psychological Assessment, 14*, 135-145.
- Ruscio, A. M., Ruscio, J., & Keane, T. M. (2002). The latent structure of post-traumatic stress disorder: A taxometric investigation of reactions to extreme stress. *Journal of Abnormal Psychology, 111*, 290-301.
- Ruscio, J. (2000). Taxometric analysis with dichotomous indicators: The modified MAXCOV procedure and a case removal consistency test. *Psychological Reports, 87*, 929-939.
- Ruscio, J. (2003). *Program code to perform taxometric analyses and simulate comparison data in the R and S+ languages*. Elizabethtown, PA: Elizabethtown College, Department of Psychology. Retrieved July 21, 2003, from <http://www.etown.edu/psychology/faculty/ruscio.htm>
- Ruscio, J., & Ruscio, A. M. (2000). Informing the continuity controversy: A taxometric analysis of depression. *Journal of Abnormal Psychology, 109*, 473-487.
- Ruscio, J., & Ruscio, A. M. (2002). A structure-based approach to psychological assessment: Matching measurement models to latent structure. *Assessment, 9*, 5-17.
- Ruscio, J., & Ruscio, A. M. (2004). Clarifying boundary issues in psychopathology: The role of taxometrics in a comprehensive program of structural research. *Journal of Abnormal Psychology, 113*, 24-38.
- Ruscio, J., Ruscio, A. M., & Keane, T. M. (2004). Using taxometric analysis to distinguish a small latent taxon from a latent dimension with positively skewed indicators: The case of Involuntary Defeat Syndrome. *Journal of Abnormal Psychology, 113*, 145-154.
- Ruscio, J., Ruscio, A. M., & Meron, M. (2004). *Simulating comparison data to test the suitability of data for taxometric analysis*. Manuscript submitted for publication.
- Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin, 38*, 1409-1438.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Uebersax, J. S. (1999). Probit latent class analysis with dichotomous or ordered category measures: Conditional independence/dependence models. *Applied Psychological Measurement, 23*, 283-297.
- Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage.

- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236-244.
- Widiger, T. A. (2001). What can be learned from taxometric analyses? *Clinical Psychology: Science and Practice*, *8*, 528-533.
- Widiger, T. A., & Clark, L. A. (2000). Toward *DSM-V* and the classification of psychopathology. *Psychological Bulletin*, *126*, 946-963.