BEHAVIOR THERAPY 35, 403–447, 2004

A Conceptual and Methodological Checklist for Conducting a Taxometric Investigation

JOHN RUSCIO

Elizabethtown College

AYELET MERON RUSCIO

The Pennsylvania State University

The taxometric method is an increasingly popular statistical approach that tests whether the structure of a latent construct is categorical or continuous. This article presents the key conceptual and methodological issues that should be addressed in an informative taxometric investigation. We aim to help potential users of taxometrics determine: (a) when a taxometric analysis is scientifically justified, (b) whether their data are suitable for taxometric analysis, (c) whether they have properly implemented a sufficient variety of taxometric procedures, (d) whether they have appropriately presented and interpreted the obtained results, and (e) whether they have adequately articulated the implications of their structural solution. Annotated program code and empirical examples are provided to illustrate how taxometric analysis is applied in practice.

The taxometric method pioneered by Paul Meehl (1973, 1995, 1999) and developed with several of his colleagues (e.g., Golden & Meehl, 1979; Grove & Meehl, 1993; Meehl & Golden, 1982; Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998) is one of a growing number of statistical approaches available for studying the nature of latent variables. Taxometric procedures explore the relations among a set of manifest indicator variables to determine whether the structure underlying them is *dimensional* (consisting of a single latent continuum) or *taxonic* (consisting of two discrete latent classes traditionally referred to as the *taxon* and *complement*). Like any other statistical approach, taxometrics addresses a particular kind of research question, works best with certain data parameters, and poses a number of critical decision points at each stage of analysis. The present article seeks to provide investigators who are new to the taxometric method with an introduction to these basic features. By the end of the article, we hope that readers will be able to: (a) determine

Address correspondence to John Ruscio, Elizabethtown College, One Alpha Drive, Elizabethtown, PA 17022; e-mail: rusciojp@etown.edu.

005-7894/04/0403–0447$1.00/0

whether taxometrics can be fruitfully applied to their own data and research questions, (b) identify the conceptual and methodological issues that should be considered in conducting an informative taxometric investigation, and (c) apply the program code shown in the Appendix to begin performing taxometric analyses.

This article is intended to stimulate research on latent structure through the sound application of taxometrics and presumes no prior familiarity with taxometric theory or practice. It begins with a list of five questions that we believe should be carefully considered and explicitly addressed in any taxometric investigation. It then proceeds to empirical illustrations of taxometric analysis in several research data sets. Due to space limitations, we focus on providing a broad introduction to the relevant issues using a brief, checklist-like format. For an elaboration of the points raised here as well as additional illustrations, we refer interested readers to an expanded, book-length treatment that we are currently preparing (J. Ruscio, Haslam, & Ruscio, 2004).

Question 1: Is a Taxometric Analysis Scientifically Justified?

The first question confronting the potential user of taxometrics is whether this method is an appropriate statistical tool for the basic and applied scientific issues that he or she wishes to address. Before considering the types of research questions that can be investigated using taxometrics, we describe the nature of the structural distinction that the taxometric method is designed to make and the types of evidence that it provides.

The Classification Problem

One of the fundamental challenges in any scientific discipline is to classify the objects of study. In psychological research, there is often heated debate about whether particular constructs are most appropriately classified using a categorical (taxonic) or a continuous (dimensional) framework. Because we wish to avoid applying a spurious typology to latent dimensional variables or representing latent taxa by artificial dimensions, it is important to empirically assess the latent structure of each construct of interest. Unfortunately, the process of distinguishing taxonic from dimensional latent structure is not as straightforward as it may seem (Grayson, 1987; Murphy, 1964). This is because the structure of a set of observed variables may or may not match that of the latent construct underlying these variables. For example, even valid indicators of two latent groups may be distributed unimodally due to measurement error. Figure 1 depicts the within-group and full-sample distributions of three indicators of biological sex. As we will later see, taxometric analyses of these indicators readily detect the two taxa (males and females) underlying the distributions. However, the substantial overlap of these manifest distributions, especially when exacerbated by other factors (e.g., within-group correlations among observed variables), makes it difficult to detect the existence of taxa using manifest-level data. To further complicate the picture,

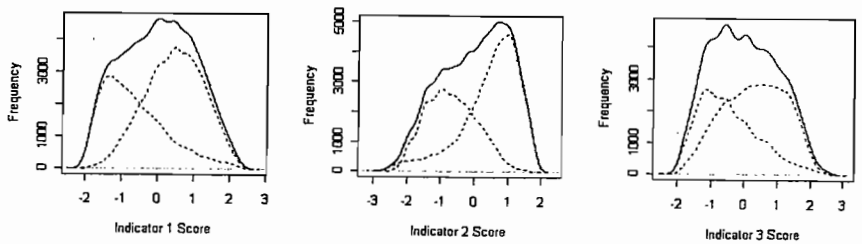

FIG. 1. Frequency distributions for three factor-score indicators of biological sex derived from the 24 items on the Masculinity-Femininity scale of the Minnesota Multiphasic Personality Inventory that most validly distinguish women (n = 7,994) from men (n = 5,586) in the Hathaway Data Bank. Solid lines represent the full sample; dotted lines represent the taxon (women, who score higher) and complement (men, who score lower) groups.

a latent dimensional construct can give rise to bimodally distributed indicators through such factors as threshold effects, sampling error (particularly in small samples), selective sampling from the extremes of a continuum, or observer bias. Thus, observed unimodality or bimodality of test scores, interview responses, and other manifest measures of a construct cannot be assumed to reflect the latent structure of that construct.

Making a Case for the Appropriateness of Taxometrics

The taxometric method was specifically designed to distinguish between taxonic and dimensional structure at the latent level. Because alternative procedures such as cluster analysis, mixture modeling, and latent class analysis are prone to false-positive identification of latent classes (see J. Ruscio & Ruscio, 2004), the goal from the earliest development of taxometric procedures was to safeguard against this possibility. There are two ways in which this defense is achieved. First, the taxometric method focuses on one putative taxonic boundary at a time using multiple indicators that have been carefully chosen to test for this boundary. Rather than submitting a large assortment of variables to a single analysis with the hope of simultaneously disentangling several potential boundaries between multiple latent classes, the taxometric method forces researchers to adopt a more focused approach tailored to the investigation of each latent class (J. Ruscio & Ruscio, in press). Second, the taxometric method includes a wide range of procedures and consistency tests that provide nonredundant evidence of latent structure. Investigators are expected to perform multiple procedures and tests within each taxometric study, and confidence in the uncovered structural solution accumulates to the extent that the results agree with one another.

Meehl (1973, 1995) often described the taxometric method as a heuristic search technique for identifying latent taxa, and others have conceptualized taxonic structure as an alternative hypothesis to be provisionally accepted if the null hypothesis of dimensional structure can be rejected (e.g., Beauchaine,

2003). According to this position, taxonic results provide evidence in favor of taxonic latent structure, but dimensional results are — like all "null" results — inherently ambiguous: They leave open the possibility that existing taxa were not detected because the data were unsuitable for analysis (e.g., the sample was too small, the indicators were insufficiently valid, and so forth). Thinking along these lines, some researchers prefer not to judge nontaxonic results as evidence of dimensional structure, believing that the alternative explanation of unsuitable data cannot be ruled out. In contrast, we believe that neither taxonic nor dimensional structure should be regarded as the null hypothesis in taxometric analysis, and that the taxometric method is capable of providing evidence in support of *either* structural solution. We take this position because recent advances in simulation methodology allow investigators to rule out the alternative explanation of unsuitable data in an appropriately conservative manner. As will later be described in greater detail, we advocate an approach that requires data to pass a "suitability test" before it is submitted to taxometric analysis (J. Ruscio, Ruscio, & Meron, 2004). If the data parameters — in combination with the planned analysis — are shown to be capable of differentiating taxonic from dimensional structure, researchers can confidently draw a dimensional inference if dimensional results are obtained.

Although the taxometric method has been demonstrated to accurately and powerfully distinguish taxonic from dimensional latent structure (e.g., Meehl & Yonce, 1994, 1996), it is important to emphasize that this is the only distinction that the method was intended to make. That is, taxometric results can be used to determine whether two groups or a single dimension better accounts for the observed relations among the indicators. More complex structural models (e.g., a hierarchical arrangement of types and subtypes, a taxon with meaningful dimensional variation) cannot be tested in a single taxometric analysis and will require additional applications of taxometric procedures and/or the use of complementary statistical tools to be fully resolved (see J. Ruscio & Ruscio, 2004, in press).

Making a Case for the Value of a Taxometric Investigation

We have argued that taxometrics may be fruitfully employed to distinguish taxonic from dimensional latent structure, and researchers seeking to make this distinction have a good case for performing taxometric analyses. However, noting the appropriateness of taxometrics for the structural question at hand is only part of the rationale that should be provided in a taxometric investigation. To fully justify such an investigation, researchers also need to articulate the scientific implications of taxonic versus dimensional structure for this particular construct and make a persuasive case that the outcome of the investigation will be of value to the field. As has been articulated elsewhere (e.g., Haslam, 1997; Meehl, 1992; Meehl & Golden, 1982; A. M. Ruscio, Borkovec, & Ruscio, 2001; J. Ruscio & Ruscio, 2002, 2004), structural understanding of a construct may enhance its theoretical conceptualization, shed light on its etiological origin, facilitate its accurate classification and

diagnosis, promote its valid and powerful measurement, specify the most appropriate research designs for its investigation, and help inform its management in clinical and public health settings. Until now, the classification value of taxometrics has been most heavily emphasized, with some mental health professionals suggesting that the method can help to "do diagnostics right" by basing nosological systems on empirically derived latent structure (Joiner & Schmidt, 2002; Meehl, 1995). The result is that the majority of taxometric studies performed to date have focused on the structure of diagnostic entities appearing in the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*; American Psychiatric Association, 1994) with the goal of informing the decades-old debate in this area. However, taxometric analysis has also been effectively applied to a number of other psychological and behavioral constructs (Haslam & Kim, 2002) and has the potential to valuably elucidate the nature of variables beyond the *DSM* disorders.

Whatever the reasons underlying a particular taxometric investigation, it is essential that they be clearly enumerated and grounded in an appropriate theoretical and empirical context. The explosion in taxometric studies in recent years, many of which have appeared in prestigious journals (Haslam & Kim, 2002), raises the possibility that some researchers will use taxometrics simply because they perceive the method to be fashionable or view studies that use it as attractive to editors and reviewers. Therefore, it is important to clearly articulate the scientific rationale for performing a taxometric analysis to avoid the impression that one is merely employing the "flavor of the month" statistical technique.

Question 2: Are the Data Suitable for Taxometric Analysis?

An informative taxometric investigation requires an appropriate sample of data in which to search for the hypothesized taxonic boundary. We begin by describing the sample and indicator parameters that are relevant to taxometrics, then present a technique that researchers can use to empirically evaluate the suitability of their data for taxometric analysis.

Sample Considerations

Sample size. On the basis of considerable Monte Carlo research, Meehl (1995) recommended a minimum sample size of $N = 300$ for taxometric investigations. Although taxometric procedures have distinguished latent structures in smaller samples, this has generally occurred only when the data possessed otherwise ideal characteristics (e.g., equal-sized groups that were separated by a large amount on indicators that were uncorrelated within groups). Because actual research data are unlikely to possess such desirable parameters, Meehl's recommendation appears to be a reasonable rule of thumb.

Taxon representation. In addition to having a large sample, it is important that this sample contain a sufficient number of putative taxon members to permit their detection. Monte Carlo studies have generally shown that taxo-

metric procedures can reveal taxonic structure with a taxon base rate as low as .10 (and, conversely, as high as .90). However, there are several reasons to treat this information with caution. First, Monte Carlo studies have seldom evaluated the performance of taxometric procedures with taxon base rates smaller than .10, so the lower limit of what the procedures can detect is presently unknown. Second, there is evidence that the absolute number of taxon members in the sample may be at least as important as the proportion of the sample that they comprise. We have found that taxometric procedures continue to detect a taxon of a constant size even when extremely large numbers of complement members are added to the sample, causing the taxon base rate to fall well below .10 (J. Ruscio & Ruscio, 2004). Thus, we use the term "small taxon" rather than "low base rate taxon" to underscore the importance of the absolute size of the putative group *as well as* its base rate in a given sample. Third, the sensitivity of taxometric procedures to the existence of a small taxon depends on many other characteristics of the data. Especially important is the validity with which the indicators separate the groups, but also relevant are the number of indicators, the magnitude of within-group correlations (or "nuisance covariance"), and the degree of indicator skew. It may therefore be misleading to set a single acceptable base rate threshold without regard for other characteristics of the data. Until further Monte Carlo research is conducted to explore this issue, researchers should take care to use the .10/.90 guide with caution, and — where possible — to collect data from a population in which the putative taxon base rate is closer to the ideal of .50.

Population sampled. The nature of the population from which cases are sampled will influence the likely size of the putative taxon and the adequacy with which the full range of functioning is represented. Samples drawn from relevant clinical populations will often provide greater representation of a putative psychopathology taxon (as well as more intermediate or "subthreshold" cases) than will samples drawn from community or student populations. For example, whereas 25% or more of the individuals served by a clinic specializing in the treatment of mood and anxiety disorders may receive a diagnosis of major depressive disorder (MDD), the base rate of MDD among unselected, nonclinical samples may be a small fraction of that amount. Because a far larger community or analogue sample will be required to obtain the same number of MDD-diagnosed cases available in a well-chosen clinical sample, it may be more feasible to conduct a high-power taxometric investigation in an appropriate clinical sample.

However, there are also potential disadvantages to conducting taxometric investigations with clinical samples. First, such samples may contain too *many* cases belonging to the putative taxon. For instance, a clinic that specializes in the treatment of social phobia may have a client roster consisting almost entirely of individuals with this diagnosis, thereby including too few members of the complement to distinguish it from the putative social phobia taxon. Similarly, if the demand for services exceeds the capacity of a clinic or hospital, clinical staff may be able to provide services only to the individuals exhibiting the most severe levels of distress or impairment, which may eliminate most or all members of the putative complement and artificially constrain the range of functioning within the sample to the point where it is no longer suitable for taxometric analysis. Conversely, not all clinical samples will include sufficiently high rates of all forms of psychopathology. For example, some conditions may be too rare to be powerfully studied in a general outpatient setting, requiring data to be collected in an inpatient facility or a specialty clinic to obtain a sufficiently large putative taxon for analysis. Finally, research estimating the base rate of a taxon in the general community or evaluating the latent structure of clinically relevant — but relatively prevalent — phenomena may be most appropriately undertaken in epidemiological samples (see Kessler, 2002).

Problematic Sampling Practices

Admixing samples. Three sampling approaches are worthy of special note because they can undermine the results of taxometric analysis. One practice that has been used with some frequency is to combine patient and nonpatient subsamples into a single sample for analysis. The problem with this is that the omission of cases experiencing intermediate or subclinical levels of functioning, along with the introduction of other systematic differences that distinguish clinical samples from normal controls, can bias results toward a taxonic solution. This false taxon, or "pseudotaxon" (Grove, 1991a), reflects the artificial admixture of two different populations rather than the true structure of the underlying construct.

Dividing samples into subsamples. A second problematic practice consists of splitting the available sample into subsamples for analysis to permit replication of the taxometric results. Although replication is, in principle, a highly desirable goal, analyses performed in subsamples drawn from a single population only examine the influence of sampling error without addressing the more important question of whether results generalize to other populations or to different conceptualizations and measurements of the target construct. Also, because taxometric procedures require large samples for analysis, we suggest that the best use of a given data set is to retain all cases for a single series of maximally powerful taxometric analyses. Splitting a sample into subsamples may be especially problematic in research involving a small hypothesized taxon, as it reduces the number of putative taxon members available for each analysis and lowers the odds of correctly detecting the taxon in any analysis.

Trimming the complement. A third questionable sampling practice involves discarding likely complement members to increase the base rate of the putative taxon in the sample. The appeal of this technique probably stems from the common assumption that it is the taxon base rate — not the absolute number of taxon members — that is critical to taxon detection by taxometric procedures. Earlier we argued that the number of taxon members may be at least as important as their base rate within the sample, and since the removal

of likely complement members does not increase the number of taxon members, this practice may be of little use to improve taxon detection. Indeed, eliminating cases with the lowest indicator scores (who are most likely to be belong to the complement) will not influence the regions of taxometric graphs that are critical to interpretation and will only diminish statistical power by unnecessarily reducing the size of the sample.

An alternative approach — eliminating cases at random from the putative complement group — is even more problematic than eliminating the lowest-scoring cases. This approach uses a fallible criterion, such as diagnostic status, to randomly discard some of the complement members (in an effort to increase the taxon base rate) or a sufficient number of complement members to equate the sizes of the two hypothesized groups. Unfortunately, this strategy can produce pseudotaxonic results. To illustrate the distorting effect of this sampling approach, consider the case of a researcher who posits that 100 out of a sample of 1,000 individuals are taxon members and that complement members should be dropped to equate the base rates. The left panel of Figure 2 shows the distribution of one of four indicators of a latent dimension in a sample of 1,000 cases, with the highest-scoring 100 cases and lowest-scoring 900 cases on this dimension designated by dotted lines. The right panel shows the distribution of this indicator in a subsample formed by randomly dropping all but 100 of the putative complement members (i.e., removing 800 of the 900 lowest-scoring cases). With the random removal of putative complement members, the indicator distribution becomes bimodal; the dotted lines reflect artificial groups created by this sampling technique. Subsequent taxometric analysis of the four indicators in the modified sample yielded pseudotaxonic results for this dimensional construct, suggesting that — like other statistical procedures — taxometric analyses cannot distinguish this type

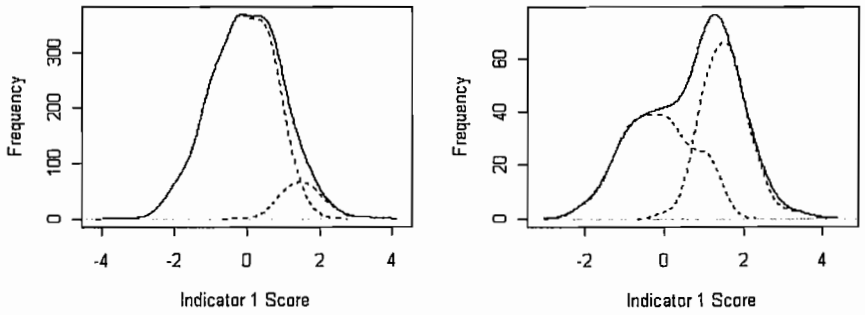

FIG. 2. Frequency distributions for one of four indicators of a latent dimension. The left graph shows the distribution of scores in the full sample of 1,000 cases, and the right graph shows the distribution of scores in a subsample of 200 cases formed by retaining the 100 highest-scoring cases plus a random selection of 100 of the 900 lower-scoring cases. Solid lines represent the full sample; dotted lines represent the putative taxon (higher-scoring) and complement (lower-scoring) groups.

of artificial bimodality from the "real thing" caused by taxonic latent structure. Although the removal of 800 cases is a particularly drastic step, the removal of fewer cases poses the same problem to a lesser degree. There is no apparent justification for introducing this potential structural confound.

In sum, we urge taxometricians to draw their sample from a single population that contains the full range of presentation of the target construct and provides a sufficient representation of the putative taxon and complement groups. We further recommend that researchers use all available cases in each analysis and studiously avoid sample selection and construction techniques that are known to increase the odds of pseudotaxonicity through sampling artifacts.

Selecting and Constructing Indicators

Content coverage. The use of any statistical tool that uses manifest indicator variables to infer latent structure requires careful consideration of how well the indicators represent the target construct (Widiger, 2001). This problem can be conceptualized in terms of content and discriminant validity (Cronbach & Meehl, 1955). For a taxometric analysis to yield informative results, one needs assurance that the indicators assess all relevant facets of the target construct *and* that they do not inadvertently assess some other construct. For example, a taxometric investigation of MDD whose indicators exclusively assess somatic symptoms may raise questions about content validity. However, even if the indicators assess each of the primary somatic, cognitive, and affective features believed to characterize MDD, they must also be shown not to converge on one or more other mood disorders (e.g., dysthymia, bipolar disorder), anxiety disorders (e.g., GAD), or psychotic disorders (e.g., schizoaffective disorder) whose features overlap with those of MDD. It is important to remember that the nature of the uncovered taxon or dimension is determined primarily by evaluating the constellation of indicators that were used to reveal it. Thus, careful selection of indicators is essential for meaningful results.

Using the available items in a data set, indicators can be constructed according to one or more accepted theoretical conceptualizations of the latent construct, according to empirical criteria (as when the observed relations among items in correlational or factor analysis form the basis for item composites), or according to a blend of theoretical and empirical considerations. So long as each of these indicator sets provides good content coverage of the underlying construct, confidence in a structural solution is increased when their results converge.

Validity and nuisance covariance. In addition to adequately representing the target construct, the indicators chosen for taxometric analysis must possess sufficiently high validity and be correlated at an acceptably low level within the hypothesized groups. Of course, without an infallible criterion to classify cases into the putative taxon and complement, it can be challenging to estimate these parameters and evaluate their adequacy for taxometric analysis. Nonetheless, there are ways to gauge the likely validity and within-

group correlations of a candidate indicator set. For example, one can assign cases to groups based on a criterion measure such as diagnostic status or a conventional threshold on one or more well-validated assessment instruments. Alternatively, one can estimate the likely base rate of taxon members in the sample, then assign the highest-scoring cases on all available indicators to the taxon and the remaining cases to the complement (the so-called "base-rate classification method" evaluated in J. Ruscio, Haslam, & Ruscio, 2004). Once cases are classified, it is a simple matter to correlate the indicators within groups as well as to estimate indicator validity. In the taxometric literature, indicator validity is usually expressed as the mean difference between the taxon and complement, standardized by the pooled within-groups variance, the metric known more familiarly as Cohen's *d*:

$$d = \frac{M_t - M_c}{\sqrt{\frac{(SD_t^2)(n_t - 1) + (SD_c^2)(n_c - 1)}{N - 2}}} \quad (1)$$

Meehl (1995) suggested that suitable indicators should separate the taxon and complement with $d \geq 1.25$ and be correlated within groups at $r \leq .30$. While these values provide useful rules of thumb, we caution researchers to evaluate indicator suitability not by the adequacy of each parameter in isolation, but by their *joint* sufficiency. For example, whereas four indicators with an average validity of $d = 1.25$ may detect a taxon with base rate $P = .50$ in a sample of 600 cases when within-group correlations are close to 0, greater validity will be required if there are fewer indicators, if the taxon is substantially smaller (or larger), if the sample is smaller, or if there are nontrivial correlations within groups.

Distributional properties. Indicator distributions also influence the suitability of data for analysis. Although taxometric analyses make no assumptions regarding normality or continuity, positively skewed indicators tend to yield rising curves in many taxometric procedures regardless of latent structure, which can make it difficult to distinguish a small taxon from a latent dimension (A. M. Ruscio & Ruscio, 2002; J. Ruscio, Ruscio, & Keane, 2004). Likewise, discrete indicators (i.e., those possessing dichotomous or other small response scales) may require special accommodations in taxometric analysis. For example, it may be advantageous to aggregate discrete items assessing the same facet of the target construct into composite indicators whose distributions better approximate continuity. Alternatively, some taxometric procedures can be modified to accommodate discrete indicators by forming composites as needed to run the analysis (Gangestad & Snyder, 1985; J. Ruscio, 2000).

Number of indicators. Finally, the number of indicators included in a taxometric analysis has implications for analytic flexibility and statistical power. Taxometric procedures can be performed with as few as two indicators, but the range of analyses that can be conducted — and the statistical power of each — increases as the number of indicators increases. It is important to emphasize, however, that we recommend against simply submitting as many indicators as possible to taxometric analyses. Although indicators must be positively correlated in the full sample, each should be relatively independent of the others, assessing a unique facet of the target construct to keep within-group correlations low. It may be the rare psychological construct that can be represented by more than a half-dozen indicators without introducing substantial redundancy. Unfortunately, some taxometric investigators have taken the approach of first identifying a desired number of indicators and then assigning items at random to composite indicators, a strategy that virtually guarantees high indicator redundancy and, in turn, high within-group correlations. We suggest instead that researchers begin with a list of theoretically relevant facets of the target construct, then form composite indicators by joining together those variables that most validly assess each facet. The empirical suitability of this candidate indicator set can be evaluated (as described below) and refined as necessary.

In conclusion, the indicators submitted to taxometric analysis must sensitively and specifically represent each facet of the target construct and distinguish the putative latent classes with sufficient validity and tolerably low within-group correlations. Indicator selection or construction should be guided by relevant theory, with consideration for empirical properties (e.g., validity, distributions, cross-indicator redundancy) that may influence the suitability of the indicators for taxometric analysis.

Empirically Evaluating Data Suitability

As noted in the preceding section, evaluating the suitability of individual data parameters overlooks the potential importance of their joint influence on taxometric analysis. One favorable parameter may (or may not) compensate for a questionable value on another parameter, different taxometric procedures and consistency tests are unlikely to be equally effective for a given set of parameters, and it can be exceptionally difficult to judge the suitability of a planned analysis for a particular set of research data. Monte Carlo studies provide only limited guidance in this regard, as their unavoidably idealized parameters and fixed implementation approach cannot cover all of the possible permutations found in specific studies. Although interpolations from the parameters of Monte Carlo studies would be feasible, the problem faced by researchers considering a taxometric analysis is often one of extrapolation, as the full range of potentially influential factors has never been systematically varied.

For these reasons, we recommend that investigators simultaneously evaluate all characteristics of their indicators in the context of the particular analyses planned for the research data. Rather than looking to large-scale simulation studies for detailed guidance, researchers can perform their own simulations in ways that are custom-tailored to their particular study. By generating tax-

onic and dimensional data sets that reproduce the parameters of the research data, then submitting these simulated data sets to each analysis intended for the research data, investigators can judge whether the research data are suitable for these analyses. If an analysis yields noticeably different results for the simulated taxonic and dimensional data, this suggests that the parameters of the research data (which were reproduced in the simulated data sets) are suitable for this analysis. If the simulated taxonic and dimensional data do not produce distinguishable results, this suggests that the indicators need to be refined (e.g., select variables that are more valid, combine redundant variables to reduce within-group correlations), that the taxometric procedure needs to be performed in a more effective way, or — if these improvements fail to produce distinguishable results — that the available data are simply unsuitable for this taxometric procedure.

An iterative simulation technique. We have developed an iterative technique for simulating taxonic and dimensional comparison data that reproduces the distributional and correlational properties of a specified set of research data (J. Ruscio, Ruscio, et al., 2004). To simulate dimensional data sets, the program begins by reading the sample size, the number of indicators, and the distribution of scores on each indicator in the research data. It then reproduces the observed indicator correlation matrix by applying shared loadings on a common latent factor to vectors of random normal deviates used to represent each indicator. Next, it pastes the observed score distributions onto each simulated indicator to replace the normally distributed, continuous scores with those from the research data. Because this alteration in the simulated indicator distributions typically reduces the indicator correlations, the program checks to see how closely these correlations now match those in the research data. The discrepancy is used to update the target correlation matrix, and the program launches a new iteration by reproducing this target matrix through new factor loadings, again pasting the observed score distributions onto the simulated indicators. The program continues until 10 successive iterations fail to improve the accuracy with which the target correlation matrix is reproduced. Thus, the simulated dimensional comparison data will match each indicator's score distribution and reproduce the indicator correlation matrix as accurately as sampling error and indicator variability allows. Naturally, correlations will be reproduced more accurately with large samples and with indicators that vary across a wide range of values.

To simulate taxonic data sets, the program requires a classification variable that indicates whether each case in the research data has been assigned to the putative taxon or complement. This classification variable can be created by considering diagnostic status, by applying a theory-based algorithm, or by simply assigning the highest-scoring cases to the taxon and the remainder to the complement following a prespecified taxon base rate. Provided with the research indicators and a fallible criterion variable denoting putative group membership, the program reproduces the indicator distributions and correlations observed within each group by submitting each subsample (i.e., the taxon and the complement) to the dimensional simulation algorithm described above, then merges the results. In this way, the technique reproduces not only within-group distributions and correlations, but full-sample parameters as well.

The resulting simulated data sets will be indistinguishable in their observed indicator distributions and correlations; they will systematically differ only in their latent structure. Subsequent taxometric analysis of these simulated data will reveal which analyses are likely to provide informative structural tests of the research data. In general, when working with very large samples (e.g., many thousands of cases), a single set each of simulated taxonic and dimensional comparison data may suffice, whereas investigators working with smaller samples may wish to simulate multiple sets of taxonic and dimensional data to determine whether sampling error has a substantial effect on the taxometric results. Likewise, researchers may wish to simulate multiple sets of taxonic comparison data using different classification criteria to ensure that the suitability test was not passed (or failed) owing only to an overly optimistic (or pessimistic) estimated rate of taxon membership in the simulated data. Such applications of this simulation technique are likely to provide a more informative and empirically rigorous evaluation of the unique parameter configuration of a research data set than an evaluation guided solely by Monte Carlo findings.

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

A cornerstone of the taxometric method is its reliance on consistency across nonredundant analytic techniques, rather than on significance testing, to establish confidence in a given structural solution. In this section, we present a conceptual overview of a number of taxometric procedures and consistency tests that can be used to amass evidence about latent structure. As different approaches will be appropriate for different studies, no one template can be created to guide all taxometric investigations. Instead, we suggest that researchers consider a wide range of possible procedures and consistency tests, but submit their data only to those analyses in which simulated taxonic and dimensional comparison data pass the suitability test. In what follows, we briefly review the major taxometric procedures and consistency tests, note the extent to which each has (or has not) been supported by Monte Carlo studies, and offer suggestions for their appropriate selection and implementation. The taxometric programs that we have written can be used to perform all of the procedures and tests described below.

Taxometric Procedures

Although a number of taxometric procedures have been developed, a core subset has received the most empirical attention in Monte Carlo studies and has been applied most frequently in taxometric investigations (Haslam & Kim, 2002). We will focus on three procedures that nicely complement one

another, providing nonredundant evidence and facilitating a substantial number of consistency tests. Each of these procedures yields distinct graphical results for taxonic and dimensional latent structures, and structural inferences are based primarily on inspection of these curves. Additional information yielded by the procedures and by ancillary consistency tests provide further clues to latent structure.

MAMBAC. The MAMBAC procedure (Mean Above Minus Below a Cut; Meehl & Yonce, 1994) requires just two indicators and is based on the search for a cutting score that would optimally distinguish taxon and complement members if these groups in fact exist. The technique is simple: Assign one indicator to the role of "input," which forms the x axis of the MAMBAC graph, and the other indicator to the role of "output." Then, calculate mean differences on the output between cases scoring above and below each of several cutting scores along the input. These mean differences are plotted as the y values for each cutting score. Because the difference between latent classes will be greatest at the optimal cutting point and lower elsewhere, taxonic structure produces a peaked MAMBAC curve. For dimensional structure, the absence of an optimal cutting point produces a concave MAMBAC curve.

Although this procedure requires only two indicators, it can be performed using all available data by removing one variable at a time to serve as the output and summing the remaining variables to serve as a composite input indicator. The use of composite input indicators accommodates variables that vary across too few values to serve as input indicators alone. For example, whereas dichotomous items cannot adequately rank-order cases along the x axis, the sum of many dichotomous items can. On the other hand, when indicators possess sufficient variation to serve as input indicators, the traditional use of input-output indicator pairs yields more MAMBAC curves (k variables produce $k[k - 1]$ curves) than does the use of composite input indicators (which yields k curves). Obtaining more curves affords more opportunities to check the consistency of results. This benefit must be weighed against the potentially greater statistical power of composite input indicators, which are likely to provide a more valid rank-ordering of cases along the x axis.

When performing MAMBAC, researchers must decide how cuts will be made along the input indicator. Cuts can be placed at fixed *SD* units (e.g., locate cuts at every .25 *SD* units along the input, which is usually standardized for this purpose), at intact scale values (e.g., locate cuts at particular values of the input indicator), between each case in the data set, or at a fixed number of equally-spaced locations between cases. Whereas cuts made at scale values or *SD* units yield relatively few points on a MAMBAC curve and may make it more difficult to interpret, cuts made between every case in the data set yield an enormous number of points that may be excessive. A compromise position is to locate a fixed number of cuts at equally-spaced intervals between cases, with about 50 cuts (beginning and ending 25 cases from either extreme to avoid undue sampling error in the calculation of mean differences) often working well to delineate the shape of the curve. Because this approach may place cuts between equal-scoring cases, it can be useful to perform internal replications to remove this obfuscating influence on curve shape. This is done by randomly resorting tied cases and rerunning the MAMBAC analysis. The mean differences are averaged across these replications to plot the final MAMBAC curve. When replications are necessary to make a curve interpretable, we have found that as few as 5 or 10 such replications are usually sufficient to smooth the shape of the curve.

MAXCOV/MAXEIG. The MAXCOV (MAXimum COVariance; Meehl, 1973; Meehl & Yonce, 1996) and MAXEIG (MAXimum EIGenvalue; Waller & Meehl, 1998) procedures both require at least three indicators and operate in a similar way, causing us to group them here. Both procedures examine the association among two or more "output" indicators within subsamples of cases ordered along an "input" indicator that forms the x axis for the graph. In MAXCOV, the covariance is calculated between two output indicators and plotted as the y value for each subsample of cases; in MAXEIG, the first (largest) eigenvalue of the covariance matrix of two or more output indicators (with variances on the diagonal replaced by zeros so that only covariances remain) is calculated and plotted as the y value for each subsample. Taxonic structure causes a peak to emerge in the subsample containing an equal mixture of taxon and complement members, with the association among indicators tapering off in adjacent subsamples. Dimensional structure produces a nonpeaked curve.

MAXCOV can be performed using indicators in all possible input-output-output triplets (which yields $k[k - 1][k - 2]/2$ curves) or by removing two variables at a time to serve as output indicators and summing the remaining variables to serve as a composite input indicator (which yields $k[k - 1]/2$ curves). As with MAMBAC, the latter technique may increase statistical power and can accommodate the use of variables with restricted response scales (Gangestad & Snyder, 1985; J. Ruscio, 2000). MAXEIG can be performed by using indicators in all possible input-output-output triplets, by using the composite input indicator technique, or by removing one variable at a time to serve as the input indicator and using all remaining variables as output indicators (yielding k curves). While MAXCOV may help to isolate the relative influence of individual indicators in the early stages of indicator selection and construction, we recommend using the multivariate MAXEIG procedure for the final analyses (J. Ruscio, 2004).

To implement MAXCOV/MAXEIG, the researcher must also decide how to divide cases into ordered subsamples along the input indicator. This can be done using nonoverlapping *intervals* or overlapping *windows*. Intervals can be constructed using *SD* units (e.g., divide the sample every .25 *SD* along the input indicator, which is usually standardized for this purpose), intact scale values (e.g., divide the sample according to particular scores or ranges of scores on the input indicator), or fixed-size subsamples (e.g., divide the sample into deciles). Windows are constructed as fixed-size subsamples that overlap (conventionally by 90%) with adjacent subsamples. We recommend that

windows be used when performing either MAXCOV or MAXEIG, as this technique provides more data points for the resulting curve — and thus facilitates interpretation of curve shape — with no increase in sampling error. For example, a sample of 1,000 cases can be divided into decile intervals to create 10 subsamples containing cases 1–100, 101–200, 201–300, . . ., 901–1,000, or the same sample can be divided into 91 windows with 90% overlap containing cases 1–100, 11–110, 21–120, . . ., 901–1,000. Alternatively, if one was satisfied with just 10 points on the curve, the use of windows would permit larger subsamples to be formed (and sampling error to be reduced) by allowing these subsamples to overlap, as in cases 1–526, 54–579, 106–631, . . ., 475–1,000. Or, an intermediate number of windows could be selected such that each subsample contains more than 100 cases *and* more than 10 data points appear on the resulting curve. In general, Waller and Meehl (1998) described the relationship between the sample size of each window (n_w), the number of windows (W), and the proportion of overlap between adjacent windows (O) as:

$$n_w = \frac{N}{W \times (1 - O) + O} \tag{2}$$

In MAXCOV/MAXEIG, just as in MAMBAC, internal replications can offset the distorting effect of drawing subsample divisions between cases possessing equal scores on the input indicator.

L-Mode. L-Mode (Latent Mode; Waller & Meehl, 1998) requires a minimum of three indicators. It works by examining the distribution of scores on a single factor — estimated in a factor analysis of all available indicators — to determine whether multiple modes are evident in this distribution. For taxonic data, two modes should emerge, whereas for dimensional data, one mode should emerge. There are no significant implementation decisions to be made when performing L-Mode, which uses all available indicators in a single analysis and produces one graph depicting the distribution of factor scores, with lower and upper modes highlighted. By convention, L-Mode curves are automatically smoothed to minimize any "lumpiness" due to sampling error that could interfere with mode location.

Consistency Tests

In addition to the three taxometric procedures described above, a number of additional tests can be conducted to distinguish taxonic from dimensional structure. We review three broad types of consistency tests: those that involve performing taxometric procedures multiple times in multiple ways, those that evaluate latent parameters and classified cases, and those that examine the fit of taxonic and dimensional structural models to the data. We have attempted to provide a comprehensive list of the consistency tests that have been performed in taxometric investigations. However, it is important to note that these tests vary considerably in quality and utility, and that while some can be

endorsed for all taxometric investigations, others may be appropriate only in some instances, and still others may seldom provide informative results and are included here primarily for cautionary purposes.

Taxometric procedures as checks for one another. Perhaps the most basic rule for any taxometric investigation is that multiple taxometric procedures should be performed. Because MAXCOV and MAXEIG operate in a similar way, they provide highly redundant evidence. Thus, we suggest including only one of these in any given study, with preference given to MAXEIG. MAMBAC should serve as a complementary source of evidence in any study, as should L-Mode when a sufficient number of indicators is available for the meaningful calculation of factor scores.

Repeated application of each taxometric procedure. Although L-Mode yields a single graph, MAMBAC and MAXCOV/MAXEIG can each be performed a number of times by using the available variables in different input-output indicator configurations. Even if averaged curves are presented to conserve space in a manuscript, researchers can inspect the full panel of curves generated by each procedure to evaluate the consistency of results.

Analyzing multiple sets of indicators in multiple populations. In addition to performing multiple procedures using multiple configurations of a single indicator set, one can construct and analyze several different indicator sets within a given sample of data. Conversely, similar or different indicators can be analyzed in samples drawn from different populations as a further test of consistency. Greater confidence can be held in a structural inference based on convergent results across multiple measures of the target construct in multiple populations.

Inchworm consistency test. When a MAXCOV/MAXEIG analysis is performed, an existing taxon that is very small (or very large) may produce a cusp at the upper (or lower) end of the curve rather than a clearly defined peak. Such a cusp can be especially difficult to distinguish from dimensional results when indicators are skewed (J. Ruscio et al., 2004). Fortunately, the "inchworm consistency test" (Waller & Meehl, 1998) can help to clarify matters. Using overlapping windows, researchers can systematically increase the number of windows over a series of analyses to determine whether a cusp near the upper end of the curve becomes a better-defined peak. The premise is that this cusp may suggest the presence of a taxon so small that taxon members are outnumbered by complement members even in the uppermost windows. By increasing the number of windows, the sample size within each window decreases, and the number of taxon members should eventually equal and then surpass the number of complement members in the uppermost windows. By contrast, when latent structure is dimensional, a cusp due solely to indicator skew should remain in the uppermost window even when the number of windows increases. Although a similar effect could be produced by reducing the amount of overlap between windows, we recommend against this practice because it unnecessarily reduces the number of data points on the MAXCOV/MAXEIG curve. Given its considerable utility in distinguishing

taxonic from dimensional structure under the adverse conditions of extreme taxon base rates and indicator skew, we believe that the inchworm consistency test should be required in any taxometric investigation that involves a small taxon.

Estimates of the taxon base rate. Each taxometric curve can be used to estimate a number of latent parameters, most notably the taxon base rate. This base rate is estimated from the relative heights of the endpoints of the curve in MAMBAC (Meehl & Yonce, 1994) and from the location of the maximal association (in covariances or eigenvalues) among the output indicators in MAXCOV/MAXEIG (Meehl & Yonce, 1996; J. Ruscio, 2004). In L-Mode, the locations of the lower and upper modes each yield an estimate of the taxon base rate, as does an empirical classification of cases (Waller & Meehl, 1998). The resulting base rate estimates can be compared within and between procedures. To the extent that they diverge, this suggests that no single group of cases was consistently detected by the differing procedures. In contrast, a high level of consistency among base rate estimates is suggestive of taxonic structure. However, it should be noted that there are other circumstances—such as when skewed indicators tilt taxometric curves in consistent ways (J. Ruscio et al., 2004)—wherein even dimensional data will produce highly concordant base rate estimates. The interpretation of base rate estimates will be discussed at greater length in the empirical illustrations appearing later in this article.

Although a number of additional latent parameters (e.g., indicator validity) can be estimated from taxometric results, research has not yet determined whether such estimates can be used to validly distinguish latent structures. Thus, it is presently unknown whether coherence among such estimates provides support for an inference of taxonic structure, nor how large a discrepancy between these estimates is required to support an inference of dimensional structure.

Case removal consistency test. Although earlier we advised against removing likely complement members to increase the taxon base rate for analysis, a related technique can be useful as a consistency test. Specifically, one can remove a targeted subset of cases that constitute either likely taxon members or likely complement members, rerun the taxometric analyses, and assess whether the direction and magnitude of the resulting change in estimates of the taxon base rate are consistent with what one would expect if a taxon exists (Meehl & Yonce, 1994; J. Ruscio, 2000). For example, if an initial series of taxometric analyses in a sample of 1,000 cases yields an averaged taxon base rate estimate of .20, rerunning the analyses following the removal of the lowest-scoring 25% of cases (almost all of whom are likely complement members) should produce a taxon base rate estimate that approaches $(.20 \times 1000)/750 = .27$ if latent structure is taxonic. In contrast, dimensional data often yield a base rate estimate that is inconsistent with the expected direction of change or that changes minimally in the expected direction.

T-tests between cases classified into the taxon and complement. A few taxometric reports have compared the means of putative taxon and complement members on variables expected to be associated with the taxon. Unfortunately, these kinds of comparisons do not pose a useful consistency test of either latent structure. This is because substantial mean differences would be expected on any variable that is related to the target construct, regardless of whether the classified groups correspond to true latent classes or to artificial groupings created along a latent dimension. When the dependent variables in these analyses are the very indicators that were employed in taxometric analyses, researchers *must* find substantial mean differences or else conclude that the indicators were not sufficiently valid for taxometrics. While the presence or absence of mean differences on external variables may be relevant to the construct validity of the target construct, they have little to no bearing on its latent structure.

Distribution of Bayesian probabilities of taxon membership. The results of a MAXCOV/MAXEIG analysis can be used to estimate not only the taxon base rate, but also the valid and false positive rates achieved by the optimal cutting score on each indicator (Meehl, 1995; J. Ruscio, 2004). Plugging this information into Bayes' Theorem yields estimates of the probability of taxon membership for each case in the sample (Meehl, 1995; Meehl & Golden, 1982), and plotting the distribution of Bayesian probabilities may provide evidence pertinent to latent structure. If the probabilities cluster around the values of 0 and 1, the resulting U-shaped distribution is suggestive of taxonic structure. On the other hand, a more broadly dispersed distribution of probabilities is suggestive of dimensional structure.

Although this consistency test has been employed in many taxometric studies (Haslam & Kim, 2002), we advise caution in the interpretation of U-shaped distributions. This is because it is possible for dimensional data to produce a U-shaped distribution of probabilities. A requirement of taxometric analysis is that the indicators be substantially correlated in the full sample, as all valid indicators of a construct will be positively associated due either to the mixture of latent classes or to shared loadings on a latent dimension. Thus, any individual who scores above the optimal cutting score on one indicator will tend to score above threshold on other indicators as well, with the reverse being true for those scoring below the threshold. For this reason, Bayesian probabilities of taxon membership can tend toward 0 or 1 even in the presence of dimensional structure (J. Ruscio & Ruscio, in press), making U-shaped distributions a poor marker of taxonic structure. However, the *absence* of a U-shaped distribution may prove to be a relatively useful marker of dimensional structure.

Goodness of Fit Index (GFI). Using estimates of the taxon base rate, of each indicator's validity, and of each indicator's variance among cases assigned to the taxon and complement, one can reproduce the indicator variance-covariance matrix and compare its fit to the observed matrix. Waller and Meehl (1998) introduced the GFI, familiar to users of structural equation modeling, as an

index suitable for use in taxometric research. Unfortunately, Monte Carlo studies do not support the use of the GFI as a taxometric consistency test (Cleland, Rothschild, & Haslam, 2000; Haslam & Cleland, 2002). These studies found the GFI to poorly differentiate taxonic and dimensional structure, and neither the suggested cutoff of .90 or any other value was generally appropriate for discriminating these structures even when taxonic and dimensional data did yield different GFI values.

Using simulated comparison data to evaluate results obtained for the research data. The limitations of the GFI can be summarized by making reference to two overarching challenges. First, the GFI provides a single measure of fit to a taxonic model, with no measure of fit to a dimensional model provided for comparison. Hence, one cannot determine which structural model achieves a superior fit. Second, attempts to identify a single GFI cutoff value that is generally indicative of taxonic structure does not allow researchers to take into account the unique parameters of a particular data set that may artificially inflate or deflate the GFI, independent of the actual fit of a taxonic model. Fortunately, investigators can avoid both of these problems by examining — and even quantifying — the extent to which the taxometric results yielded by the research data are similar to those yielded by the simulated taxonic and dimensional comparison data.

This process begins by using the iterative technique described previously to simulate one or more sets of taxonic and dimensional comparison data. Next, these data are submitted to the same analyses that were performed on the research data. To the extent that the results of the research data more closely resemble those of one type of simulated data than the other, confidence in that structural solution is increased. Thus, simulated comparison data matching the parameters of the research data can serve not only as an essential suitability test, but also as a valuable interpretive benchmark against which to evaluate the research results.

This technique has proven especially helpful in studies that have attempted to distinguish a small taxon from positively skewed indicators of a latent dimension (A. M. Ruscio & Ruscio, 2002; J. Ruscio et al., 2004), as it holds indicator skew constant across data sets and identifies any remaining differences in curve shapes and parameter estimates arising from latent structure. The approach may also be able to redeem consistency tests that have performed poorly in isolation but that, when applied to both taxonic and dimensional simulated data and compared to the research data, may yield more conclusive results. For example, although there may not be one single GFI value that validly distinguishes taxonic from dimensional structure under varying data conditions, the GFIs yielded by simulated taxonic and dimensional comparison data within a given study may be quite distinct. If the GFI yielded by the research data closely approximates one (but not the other) of these values, a defensible structural conclusion may be reached without the need for a broadly applicable threshold.

Quantifying fit to curves of known structure. It is also possible to quantify the fit of taxometric curves yielded by the simulated comparison data sets to those yielded by the research data. For example, after conducting a series of MAXCOV/MAXEIG analyses and obtaining an averaged curve for the research data, researchers can perform two additional series of analyses to obtain averaged curves for the simulated taxonic and dimensional data. Then, in addition to visually gauging whether the research curve is more consistent with the simulated taxonic or dimensional curve, the similarity of these curves can be quantified. This is done by calculating the root mean square residual (RMSR) of the y values on the averaged curves of the research data and the simulated data. This index is calculated once to evaluate the fit of the averaged curve for the simulated taxonic data and once to evaluate the fit of the averaged curve for the simulated dimensional data. The Fit_{RMSR} index is calculated as follows:

$$Fit_{RMSR} = \sqrt{\frac{\sum (y_{res.data} - y_{sim.data})^2}{N}}, \tag{3}$$

where $y_{res.data}$ refers to a data point on the averaged curve for the research data, $y_{sim.data}$ refers to the corresponding data point on the averaged curve for simulated taxonic or dimensional data, and N refers to the number of points on each curve. *Lower* values of Fit_{RMSR} reflect better fit, with perfect fit represented by a value of 0. To the extent that Fit_{RMSR} differs across the simulated taxonic and dimensional data, evidence accumulates in favor of the structure that yielded the superior fit.

If multiple sets of taxonic and dimensional comparison data are generated, one can calculate Fit_{RMSR} for each and then compare the average fit of the simulated taxonic data sets to the research data relative to the average fit of the simulated dimensional data sets. This is done by calculating the mean and standard deviation of the Fit_{RMSR} values for the taxonic data sets (M_t and SD_t) and the dimensional data sets (M_d and SD_d), then calculating a fit index based on Cohen's d as follows:

$$Fit_d = \frac{M_t - M_d}{\sqrt{\frac{SD_t^2 + SD_d^2}{2}}} \tag{4}$$

Because lower values of Fit_{RMSR} are indicative of taxonic structure, subtracting the average fit of the curves for dimensional data from those for taxonic data should yield negative values of Fit_d for taxonic structure and positive values for dimensional structure. Thus, the sign of the Fit_d suggests a structural conclusion and its magnitude suggests the strength of the evidence for this conclusion.

When we have performed the Fit_d consistency test in the past, we have tended to use 10 sets each of simulated taxonic and dimensional comparison data. However, further study is needed to determine whether larger numbers of data sets yield more accurate results that offset the additional computational demand, as well as to establish positive and negative threshold values for inferring dimensional and taxonic structure, respectively. The potential utility of the Fit_{RMSR} and Fit_d indices will be explored in our empirical illustrations, though more rigorous Monte Carlo evaluations of these indices are also needed.

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

There are few hard-and-fast rules concerning the presentation of taxometric results. Rather than espousing a one-size-fits-all approach to the reporting of taxometric analyses, this section will highlight a few issues that warrant careful consideration and offer a few tentative suggestions for communicating the wealth of information yielded by a taxometric investigation in an economical way without compromising its informativeness.

Graphing Considerations

Summarizing graphical results. Taxometric studies inevitably produce a large number of graphs. We believe that it is essential to provide readers with sufficient information about how these graphs were generated, along with a sufficient representation of the graphs themselves, to permit an independent evaluation of their shapes. To conserve space, it is often helpful to present a single averaged curve or a small number of representative curves in lieu of an entire panel of curves generated by a given taxometric procedure. However, because averaging curves has the potential to distort results, we recommend that researchers interpret the full panel of curves, carefully inspect averaged curves to ensure their representativeness of the full panel, and make the full panel available to interested readers upon request.

Smoothing and scaling graphs. While early taxometric studies relied heavily on curve smoothing techniques (e.g., variants on the running medians procedure, Tukey, 1977; locally weighted scatterplot smoothing, Cleveland, 1979), this practice has been controversial because it can artificially flatten genuine taxonic peaks. If smoothing is performed, it may be best to present both "raw" and smoothed data points on the same graph. A related consideration specific to the graphing of MAXCOV/MAXEIG results concerns the scaling of the *y* axis: Too narrow a range of values may give fluctuations caused by sampling error the appearance of a peaked curve, whereas too wide a range of values may artificially flatten the curve. Provided that interpretations are based on full panels of curves, this should not be a significant problem, as a consistent curve shape should emerge in taxonic data even if a poor choice of scaling is made. The program that we have written to perform taxometric analyses holds the *y* axis constant across each triplet of research, simulated taxonic, and simulated dimensional curves to facilitate their accurate comparison. In addition, the curves for multiple sets of simulated comparison data, along with the average of these curves, are presented on the same graph so that one can see the extent to which sampling error is influential on curve shape for a given analysis and sample size.

Interpretation Considerations

In addition to the choices that must be made in communicating graphical results, it is essential that this graphical output and its accompanying quantitative information be interpreted correctly. In the foregoing discussions of data suitability, taxometric procedures, and consistency tests, we have stressed the value of using simulated comparison data to help guide interpretations, a point that we reiterate here.

Regardless of the structural inference that is drawn, the construct to which this inference pertains must be carefully identified by scrutinizing the indicators that were submitted to analysis. Too often, researchers either assume it to be self-evident that their indicators represent the target construct or provide insufficient detail about the nature of the construct that they intended to examine. However, in any study that uses manifest indicators to draw inferences about latent constructs, a principled argument is required to establish construct validity. In particular, researchers must establish that their indicators converge on the desired construct and discriminate it from other constructs whose structure they did not intend to test. Follow-up analyses examining associations between the taxon (or dimension) and external variables may help to more firmly establish its construct validity (Watson, 2003).

Question 5: Are Implications of the Findings Clearly Articulated?

Just as we argued for the importance of beginning a taxometric study with a compelling scientific rationale for its conduct, we encourage researchers to return in their discussions to the basic and applied scientific issues that motivated their study, drawing conclusions in light of the obtained structural results. Simply stating that the construct appears to be taxonic or dimensional is not sufficient. Instead, researchers should clearly articulate what they regard as the theoretical, empirical, and practical implications of these results. Even better would be to use the taxometric results as a launching pad for additional theoretical or empirical analysis, such as further delineation of remaining structural complexities underlying the construct, examination of its association with other variables at the latent level, and evaluation of competing causal models that may explain its revealed structure (J. Ruscio, Haslam, et al., 2003; J. Ruscio & Ruscio, 2004).

Empirical Illustrations

Having identified the particular research questions that taxometrics can address and outlined the major conceptual and methodological issues that an

informative taxometric investigation should consider, we turn now to a series of empirical illustrations to demonstrate how taxometric procedures and consistency tests may be selected, performed, and interpreted using real data. Though our presentation is far more concise than a typical taxometric report (to make room for illustrations of the method in several data sets), we highlight the major steps involved in each series of analyses so that prospective users of the method can gain a sense for how it is applied in practice. An annotated listing of all command lines that were used to perform these analyses appears in Appendix A, and sample output for one analysis appears in Appendix B.

Because the purpose of these examples is to illustrate how taxometric analyses are performed, we concentrate our discussion on the methodological considerations outlined in Question 3 of the foregoing checklist. Due to space limitations, we focus on three taxometric procedures — MAMBAC, MAXEIG, and L-Mode — and on consistency tests whose results can be concisely summarized. In addition, we employed the same three consistency tests in each example to underscore their varying performance under different data conditions. These included the base rate consistency test, selected because of its widespread popularity and use; the new curve-fit indices Fit_{RMSR} and Fit_d, included to demonstrate their potential utility; and the GFI, selected to show how even a test that may yield ambiguous results in isolation can be informative when applied in a comparative context using simulated data.

Biological Sex

For our first example, we performed taxometric analyses using fallible indicators of a construct with known structure: biological sex. We chose this data set to demonstrate the ability of taxometrics to correctly reveal a known taxon and to accurately estimate its base rate in the sample. This example also illustrates how taxometrics can be performed using very different kinds of indicators drawn from the same set of research data.

Sample and indicators. The data for this example were drawn from the Hathaway Data Bank, a large data set comprising all available Minnesota Multiphasic Personality Inventories (MMPI) completed at the University of Minnesota Hospitals between 1940 and 1976. The present sample included one valid MMPI for each patient (see J. Ruscio & Ruscio, 2000, for a description of how valid MMPIs were identified). Within this sample, the 55 dichotomous items of the Masculinity-Femininity (*Mf*) scale were used to construct two indicator sets of biological sex (see J. Ruscio, Ruscio, et al., 2004). The first indicator set (MMPI#1, N = 14,049) contained the 8 *Mf* items that most validly distinguished females (the taxon) from males (the complement) while also being minimally correlated within groups; these groups were formed using the criterion of reported sex. We retained 8 of these dichotomous items because studies have found this number to produce interpretable taxometric curves (e.g., Gangestad & Snyder, 1985; J. Ruscio, 2000). The second indicator set (MMPI#2, N = 13,580) was constructed by submitting the 24 most valid *Mf* items to a factor analysis constrained to a

3-factor solution. Factor scores were calculated on each of the three resulting indicators for each case possessing complete data. Thus, both indicator sets were constructed according to purely empirical considerations, given the empirical criterion-keying approach underlying the MMPI. The large sample size, moderate taxon base rate (.59), and unselected nature of the sample were quite appropriate for taxometric analysis.

Table 1 summarizes the distributions and estimated validities of each indicator from the two indicator sets, and Table 2 summarizes their correlations. Both indicator sets appear to possess dubious validity. This is especially true of MMPI#1, in which only 1 of the 8 indicators achieved a validity greater than the conventional threshold of d = 1.25 *SD*. We note this concern to underscore the value of evaluating data suitability using parallel analyses of simulated taxonic and dimensional comparison data. As will be seen shortly, these indicator sets performed admirably in taxometric analysis even though they might have been rejected outright based on conventional rules of thumb.

Taxometric analysis of MMPI#1. Because the MMPI#1 indicator set consisted solely of dichotomous items, MAMBAC and MAXEIG were performed using composite input indicators, and L-Mode was not performed at all.

TABLE 1
DISTRIBUTIONS AND ESTIMATED VALIDITIES FOR EACH INDICATOR SET

	Sex, Dichotomous Items[a]							
	1	2	3	4	5	6	7	8
Taxon *M*	.81	.68	.43	.54	.79	.36	.48	.57
Comp. *M*	.40	.13	.16	.20	.37	.11	.19	.24
Validity (*d*)	.96	1.33	.62	.75	.97	.59	.64	.70

	Sex, Factors[b]			PTSD[c]				GAD[d]		
	1	2	3	1	2	3	4	AB	C	E
Skew	.02	−.29	.17	−.42	−.55	−.18	.14	.56	.98	.69
Kurtosis	−.89	−.87	−.87	.10	.10	−.27	−.59	−.63	−.21	−.02
Taxon *M*	.40	.47	.31	.45	.46	.45	.44	1.88	1.95	1.83
Taxon *SD*	.80	.88	.93	.72	.69	.75	.84	.50	.63	.63
Comp. *M*	−.64	−.68	−.44	−.97	−.98	−.99	−.95	−.12	.02	−.15
Comp. *SD*	.90	.76	.93	.79	.82	.69	.58	.92	.92	.87
Validity (*d*)	1.29	1.38	.80	1.91	1.97	1.97	1.80	2.24	2.14	2.34

[a] Taxon = 8,282 women (59%); complement = 5,767 men (41%); N = 14,049.
[b] Taxon = 7,994 women (59%); complement = 5,586 men (41%); N = 13,580.
[c] Post-traumatic Stress Disorder (N = 1,063); taxon n = 723 (68%), complement n = 340 (32%); indicator 1 = reexperiencing and situational avoidance; indicator 2 = withdrawal and emotional numbing; indicator 3 = arousal and lack of control; indicator 4 = self-persecution (guilt and suicidality).
[d] Generalized Anxiety Disorder (N = 4,824); taxon n = 376 (8%), complement n = 4,448 (92%); indicators represent the corresponding *DSM-IV* diagnostic criteria.

TABLE 2
SUMMARY OF CORRELATIONS FOR EACH INDICATOR SET

Indicator Set	Full Sample		Taxon		Complement	
	M	*SD*	*M*	*SD*	*M*	*SD*
Sex, Dichotomous Items[a]	.20	.07	.03	.08	.21	.14
Sex, Factors[b]	.27	.02	−.02	.06	.14	.05
PTSD[c]	.61	.03	.36	.06	.19	.08
GAD[d]	.63	.08	.18	.12	.51	.10

[a] Taxon = 8,282 women (59%); complement = 5,767 men (41%); N = 14,049.
[b] Taxon = 7,994 women (59%); complement = 5,586 men (41%); N = 13,580.
[c] Post-traumatic Stress Disorder (N = 1,063); taxon n = 723 (68%), complement n = 340 (32%).
[d] Generalized Anxiety Disorder (N = 4,824); taxon n = 376 (8%), complement n = 4,448 (92%).

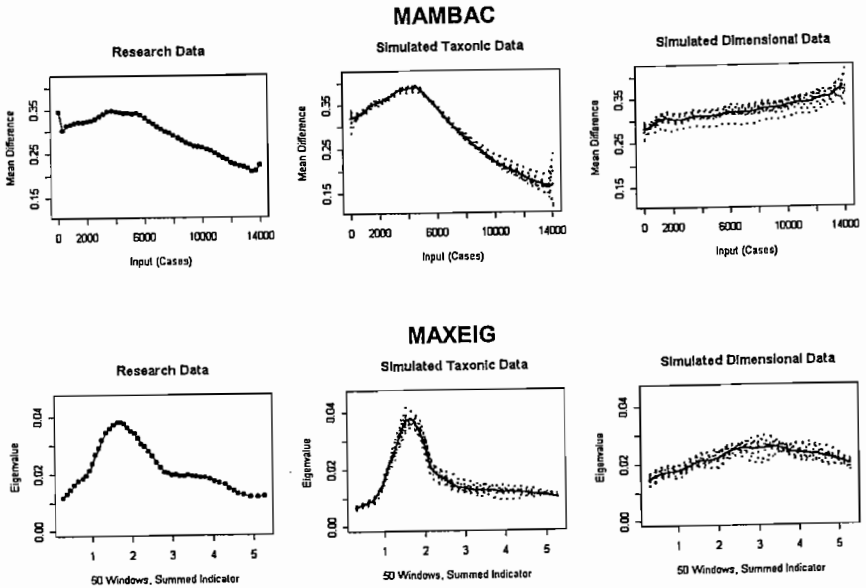

FIG. 3. Analyses of the eight dichotomous indicators chosen from items on scale 5 (Masculinity-Femininity) of the Minnesota Multiphasic Personality Inventory in the Hathaway Data Bank (N = 14,049). To simulated taxonic comparison data, the taxon consisted of 8,282 women (59%) and the complement consisted of the remaining 5,767 men. To conserve space, only the averaged curves are presented for each series of 8 MAMBAC and 28 MAXEIG analyses, all of which were performed using composite input indicators. For the simulated comparison data, the solid line represents the average of the 10 individual curves (dotted lines).

MAMBAC was conducted with 50 equally-spaced cuts along the input indicator (beginning and ending 25 cases from both extremes), whereas MAXEIG was conducted using 50 windows that overlapped by 90%; these analytic approaches were applied in all subsequent analyses. To generate taxonic comparison data, the infallible criterion was provided to the simulation program. As can be seen in Figure 3, the averaged MAMBAC and MAXEIG curves for the research data were virtually identical to those for the simulated taxonic data and quite distinct from those for the simulated dimensional data. Hence, the suitability test was passed, and the taxometric curves were easily interpreted as taxonic.

Next, the taxon base rate was estimated from each taxometric analysis (see Table 3), and these values were inspected for consistency. Examination of the standard deviations within and between procedures revealed few differences across simulated taxonic and dimensional data sets, suggesting that the base rate consistency test was not particularly informative for this indicator set. Two observations are worthy of note, however. First, taxon base rate estimates derived from the research data were highly accurate: The averaged estimate of .61 was close to the true base rate of .59. Second, MAMBAC base

TABLE 3
SUMMARY OF TAXON BASE RATE ESTIMATES

Indicator Set	*N*	Research Data		Sim. Tax. Data		Sim. Dim. Data	
		M	*SD*	*M*	*SD*	*M*	*SD*
Sex, Dichotomous Items							
MAMBAC	8	.62	.15	.78	.21	.43	.07
MAXEIG	28	.59	.13	.65	.10	.43	.15
M (*SD*)		.61 (.02)		.72 (.09)		.43 (.00)	
Sex, Factors							
MAMBAC	6	.32	.50	.58	.46	.48	.09
MAXEIG	3	.65	.05	.75	.06	.34	.13
L-Mode	3	.61	.07	.63	.05	.49	.49
M (*SD*)		.53 (.18)		.65 (.09)		.44 (.08)	
PTSD							
MAMBAC	4	.52	.05	.52	.04	.51	.04
MAXEIG	4	.73	.14	.64	.02	.71	.03
L-Mode	3	.45	.44	.61	.08	.51	.50
M (*SD*)		.57 (.15)		.59 (.06)		.58 (.12)	
GAD							
MAMBAC	3	.33	.06	.32	.02	.33	.02
MAXEIG[a]	3	.26	.02	.18	.02	.24	.06
L-Mode	3	.35	.02	.29	.04	.64	.33
M (*SD*)		.31 (.05)		.26 (.07)		.40 (.21)	

Notes. PTSD = Posttraumatic Stress Disorder; GAD = Generalized Anxiety Disorder.
[a] These estimates were derived from MAXEIG analyses performed using 100 windows.

TABLE 4
SUMMARY OF GOODNESS OF FIT (GFI) VALUES

Indicator Set	Research Data	Taxonic Data	Dimensional Data
Sex, Dichotomous Items			
MAMBAC	1.00	1.00	1.00
MAXEIG	1.00	1.00	1.00
Sex, Factors			
MAMBAC	1.00	.99	1.00
MAXEIG	1.00	1.00	.98
L-Mode	.99	.99	1.00
PTSD			
MAMBAC	.86	.90	.84
MAXEIG	.79	.91	.79
L-Mode	.85	.92	.84
GAD			
MAMBAC	.96	.95	.95
MAXEIG[a]	.80	.93	.87
L-Mode	.94	.92	.82

Notes. PTSD = Posttraumatic Stress Disorder; GAD = Generalized Anxiety Disorder.
[a] GFI is reported for MAXEIG analyses performed using 100 windows.

rate estimates for the simulated dimensional data were quite consistent with one another (SD = .07), contrary to the expectation of widely diverging estimates for dimensional structure.

As a second consistency test, GFI values were calculated from estimates of indicator validity and within-group variance obtained by classifying cases using each taxometric procedure. For MAMBAC, cases were classified using the averaged base rate estimate: the cases scoring highest on the sum of all indicators were assigned to the taxon at a proportion equal to the base rate, with the remaining cases assigned to the complement. For MAXEIG, cases were classified using Bayes' Theorem. (Had L-Mode been performed with this indicator set, cases could have been classified using the profile similarity algorithm described in Waller & Meehl, 1998.) As can be seen in Table 4, the GFI values failed to differentiate taxonic from dimensional structure, with all three data sets yielding perfect fit. Parallel computation of the GFI in the simulated data sets was especially valuable in this instance; without this comparison, it would have been tempting to interpret the GFI of 1.00 for the research data as strongly supportive of a taxonic conclusion. Indeed, the fact that seemingly perfect fit was achieved for simulated dimensional as well as taxonic data argues against the use of any fixed threshold for interpreting the GFI.

Finally, both of the new curve-fit indices (Fit_{RMSR} and Fit_d) confirmed what was plainly evident in visual inspection of the curves: The MAMBAC and MAXEIG curves yielded by the simulated taxonic data more closely fit those of the research data than did the curves yielded by the simulated dimensional

TABLE 5
SUMMARY OF CURVE-FIT VALUES FOR MAMBAC AND MAXEIG ANALYSES

	Fit_{RMSR}		
Indicator Set	Taxonic	Dimensional	Fit_d
Sex, Dichotomous Items			
MAMBAC	**.034**	.072	−7.62
MAXEIG	**.006**	.009	−4.44
Sex, Factors			
MAMBAC	**.104**	.227	−9.67
MAXEIG	**.041**	.086	−9.15
PTSD			
MAMBAC	.143	**.030**	7.46
MAXEIG	.139	**.041**	4.76
GAD			
MAMBAC	.090	**.075**	1.61
MAXEIG — 50 windows	.046	**.019**	7.12
MAXEIG — 100 windows	.048	**.026**	3.15

Notes. PTSD = Posttraumatic Stress Disorder; GAD = Generalized Anxiety Disorder. For each comparison of Fit_{RMSR} values across the simulated taxonic and dimensional comparison data, the value representing better fit appears in bold print.

data (see Table 5). Thus, whereas the GFI was unable to differentiate taxonic and dimensional structure, the curve-fit indices made this distinction well.

Taxometric analysis of MMPI#2. The MMPI#2 (factors) indicator set consisted of three continuously distributed indicators, enabling us to perform MAMBAC, MAXEIG, and L-Mode analyses. Given the continuous distributions of the indicators and favorable data parameters (e.g., very large N, moderate taxon base rate), MAMBAC was performed with individual (rather than composite) indicators to generate more curves for consistency testing. Given these reasons as well as the availability of only three indicators, MAXEIG was performed with individual indicators in all possible input-output-output configurations. Finally, a single L-Mode analysis was performed with all three indicators.

Results of these analyses are displayed in Figure 4. The suitability test was clearly passed for MAMBAC and MAXEIG, whereas the L-Mode curves for the two simulated data sets were a bit more difficult to differentiate. When analyses were performed on the research data, MAMBAC and MAXEIG results were highly similar to those yielded by the simulated taxonic data and very different from those yielded by simulated dimensional data. Interestingly, the L-Mode results for the research data were clearly bimodal, providing strong evidence for taxonicity even though the initial suitability test was only marginally passed.

Taxon base rate estimates generated by MAXEIG and L-Mode were consistent and accurate for both the research data and simulated taxonic data, and

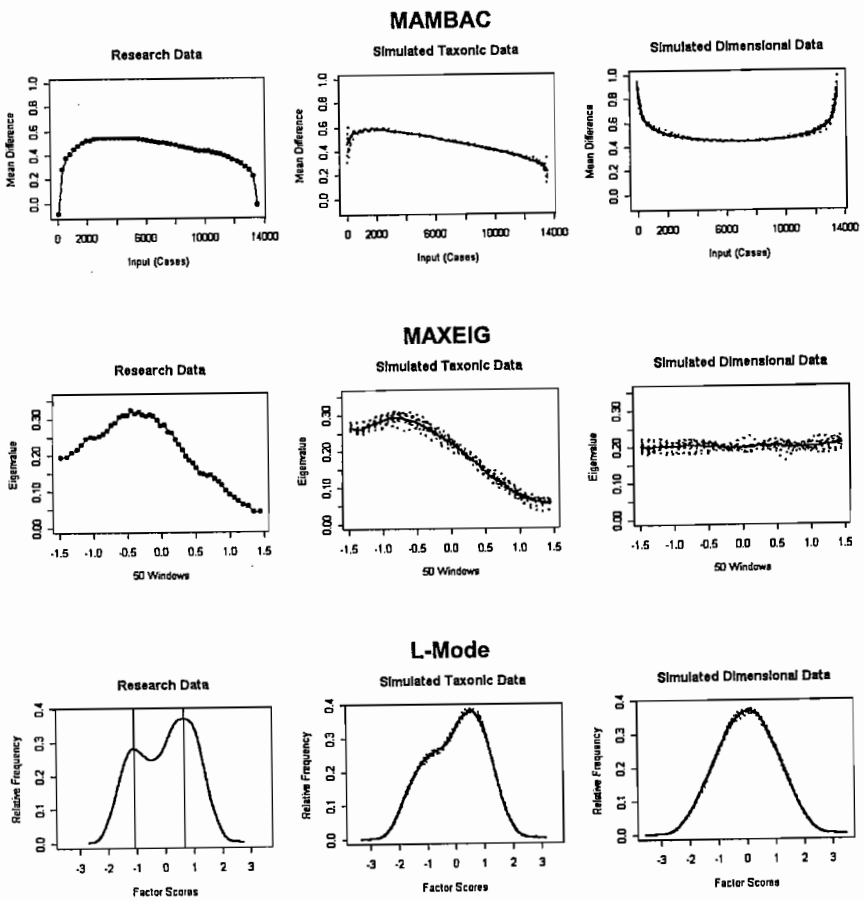

FIG. 4. Analyses of the three composite indicators constructed through a factor analysis of the 24 most valid items on scale 5 (Masculinity-Femininity) of the Minnesota Multiphasic Personality Inventory in the Hathaway Data Bank (N = 13,580). To simulated taxonic comparison data, the taxon consisted of 7,994 women (59%) and the complement consisted of the remaining 5,586 men. To conserve space, only the averaged curves are presented for each series of 6 MAMBAC and 3 MAXEIG analyses. For the simulated comparison data, the solid line represents the average of the 10 individual curves (dotted lines).

less consistent for the simulated dimensional data (see Table 3). In contrast, base rate estimates yielded by MAMBAC were once again markedly discrepant for the research data and the simulated taxonic data, yet consistent for the simulated dimensional data. This underscored the value of interpreting base rate estimates within the context of a comparison with both taxonic and dimensional simulated data. GFI values again failed to discriminate between taxonic and dimensional structures (see Table 4), whereas the curve-fit indices were clearly consistent with a taxonic solution (see Table 5).

Posttraumatic Stress Disorder (PTSD)

Our second example examined the latent structure of a construct — PTSD — whose structure was unknown. We chose this data set to illustrate the performance of taxometrics with data whose parameters are generally well-suited for this analytic approach. We also present this example to demonstrate the ability of taxometric analysis, when used with simulated comparison data, to support a dimensional structural solution.

Sample and indicators. The data set for this example included 1,063 male combat veterans who completed an outpatient psychological assessment at the Boston Veterans Administration Medical Center between 1985 and 2000 (see A. M. Ruscio, Ruscio, & Keane, 2002, for further details). Veterans reported a wide range of PTSD symptom severity, and 68% qualified for a *DSM-IV* diagnosis of PTSD by the symptom-calibrated scoring rule of the Clinician-Administered PTSD Scale (Weathers, Ruscio, & Keane, 1999), making this a highly appropriate sample for evaluating the latent structure of PTSD.

Indicators for taxometric analysis were drawn from the Mississippi Scale for Combat-Related PTSD (Keane, Caddell, & Taylor, 1988), a reliable and valid measure of the diagnostic criteria and associated symptoms of combat-related PTSD. The 35 items of the measure, each rated on a 5-point Likert scale, were summed to create four composite indicators, each corresponding to one of the four factors of PTSD uncovered in prior confirmatory factor analyses of the measure. Thus, this indicator set was constructed through a blend of theoretical and empirical approaches, with attention to the relevant facets of the latent construct as well as the empirical relations among the items in previous research. Distributions and estimated validities of the four indicators appear in Table 1 and inter-indicator correlations appear in Table 2. Indicator validity appeared to be quite strong, though within-group correlations were somewhat high among cases assigned to the putative taxon.

Taxometric analysis. To generate taxonic comparison data, we assigned the 68% of cases scoring highest on the sum of all indicators into the taxon and the remaining cases into the complement and provided this fallible criterion to the simulation program; we constructed a criterion variable in this way because PTSD diagnoses were missing for some cases and we wished to perform analyses using the full sample of data. MAMBAC was performed using composite input indicators to enhance reliability and power, whereas MAXEIG (which already includes all indicators in each analysis to enhance power) was performed using each indicator as the input and all other indicators as the output. L-Mode was performed in the usual manner with all four indicators.

The MAMBAC, MAXEIG, and L-Mode results for all data sets are shown in Figure 5. Each procedure yielded markedly different curve shapes for the taxonic versus dimensional comparison data, indicating that the PTSD indicators were suitable for analysis with all three procedures. Hence, while the estimated within-group correlations initially seemed to be potentially problematic, suitability testing based on the simultaneous consideration of all data parameters was satisfactorily passed. Subsequent analysis of the research data yielded results that were far more consistent with those of the simulated

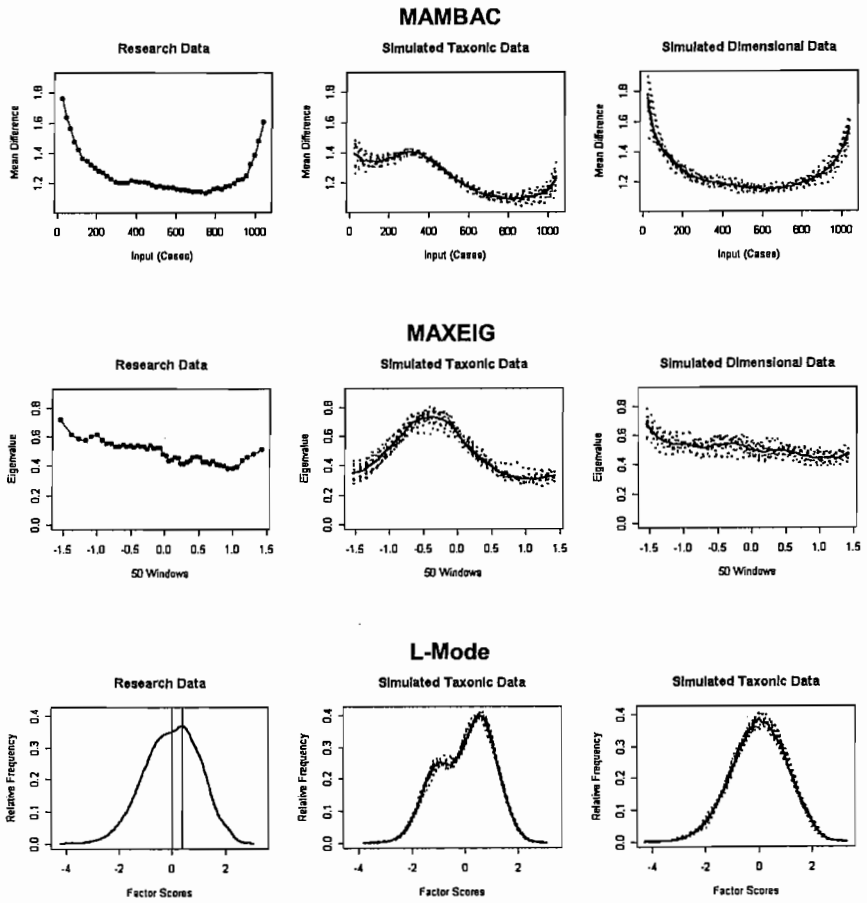

FIG. 5. Analyses of the four composite indicators constructed from items on the Mississippi scale in the Posttraumatic Stress Disorder data set (N = 1,063). Taxonic comparison data were simulated by assigning the 68% of cases with the highest total scores on the 4 indicators to the taxon and the remaining 32% of cases to the complement. To conserve space, only the averaged curves are presented for each series of 4 MAMBAC (performed using composite input indicators) and 4 MAXEIG analyses. For the simulated comparison data, the solid line represents the average of the 10 individual curves (dotted lines).

dimensional data than with those of the simulated taxonic data, providing support for the dimensionality of PTSD.

Across all three taxometric procedures, taxon base rate estimates yielded by the research data and the simulated dimensional data were similarly inconsistent, and estimates by both were less consistent than those yielded by the simulated taxonic data (see Table 3). Within procedures, however, these estimates were not always as informative. As it did for both MMPI indicator sets, MAMBAC once again generated consistent base rate estimates for the

research, dimensional, and taxonic data sets. Although these illustrations are limited in scope, they are consistent with our general observation that MAMBAC often produces coherent base rate estimates even for dimensional data, suggesting that such a finding provides only weak evidence of taxonicity. In the present example, MAXEIG also yielded consistent base rate estimates for both of the simulated data sets. Only the estimates generated by L-Mode provided a useful consistency test in this data set.

In contrast to the previous example, the GFI nicely distinguished the taxonic and dimensional simulated data, signifying its suitability as a consistency test for the PTSD indicator set. Subsequent computation of GFI values for the research data revealed them to closely approximate those for the simulated dimensional data, and to differ from those for the simulated taxonic data, across all three taxometric procedures (see Table 4). Finally, the curve-fit indices provided additional support for dimensional structure (see Table 5).

In sum, all procedures and consistency tests passed the suitability test and pointed toward the dimensional structure of PTSD, a conclusion that was further supported by a broader range of taxometric analyses performed on additional indicator sets in this sample (see A. M. Ruscio et al., 2002). Because the parallel analysis of simulated data helped to rule out the alternative explanation that nontaxonic findings were due to unsuitable data parameters, we could more confidently reach a dimensional conclusion.

Generalized Anxiety Disorder (GAD)

Our final example evaluated the latent structure of GAD, another construct with unknown latent structure. This illustration demonstrates the performance of taxometrics with data whose suitability for analysis is questionable, ultimately resulting in ambiguous results and suspended judgment about latent structure.

Sample and indicators. The sample consisted of 4,824 unselected undergraduate students enrolled at a large, northeastern university. Students reported a range of anxiety symptom severity, with 8% of the sample qualifying for a *DSM-IV* diagnosis of GAD by the Generalized Anxiety Disorder Questionnaire (GAD-Q-IV; Newman et al., 2002). Indicators were constructed from the items of two questionnaires: the GAD-Q-IV, a self-report diagnostic measure assessing each symptom of GAD, and the Penn State Worry Questionnaire (Meyer, Miller, Metzger, & Borkovec, 1990), a measure of the trait worry characteristic of individuals with GAD. Initially, four composite indicators were constructed by summing items assessing *DSM-IV* criteria A, B, C, and E. However, because the Criterion A (excessive, pervasive, chronic worry) and Criterion B (uncontrollable worry) indicators were so highly correlated (r > .80 in the full sample and within the GAD and non-GAD groups), they were combined into a single indicator, yielding three composite indicators that provided good content coverage of the construct while endeavoring to attain adequate data parameters for analysis. Descriptive data for the GAD indicator set appear in Tables 1 and 2. Although the estimated indicator validities were very high, so were the correlations within the puta-

tive complement. Combined with a moderate amount of indicator skew and a small putative taxon, the suitability of these data for taxometrics appeared uncertain.

Taxometric analysis. To generate taxonic comparison data, GAD-Q-IV diagnostic status was provided to the simulation program as a fallible criterion. To maximize statistical power under these challenging data conditions, MAMBAC analyses were performed using composite input indicators. However, because only three indicators were available, there was no choice but to perform MAXEIG with indicators in the three possible input-output-output configurations. One series of MAXEIG analyses was performed with 50 windows, whereas a second series was performed with 100 windows to implement the inchworm consistency test. L-Mode was conducted in the usual manner with all three indicators.

Results for all of these analyses are shown in Figure 6. MAXEIG results based on 50 windows were only slightly different for the simulated taxonic and dimensional data, but this difference grew clearer with 100 windows, thereby passing the suitability test. L-Mode results also appeared to pass the suitability test, though perhaps more marginally. However, the MAMBAC results were virtually indistinguishable across taxonic and dimensional structures, arguably failing the suitability test. Under these circumstances, it would not be advisable to perform MAMBAC on the research data; the research results are provided in Figure 6 only for illustrative purposes, and this MAMBAC curve is indeed very difficult to interpret. The MAXEIG curves of both the research data and the simulated taxonic data evidenced a slight peak with 50 windows that grew more pronounced with 100 windows, whereas the curve for the simulated dimensional data did not peak in either analysis. Finally, the L-Mode research curve exhibited a hint of an upper mode that was more similar to the simulated taxonic curve than to the simulated dimensional curve, but that was less pronounced and located in a slightly different position along the x axis.

Quantitative indices also yielded mixed results. Estimates of the taxon base rate (see Table 3) yielded by MAMBAC and MAXEIG analyses were low and consistent for all three data sets. This finding was interesting for two reasons. First, the fact that taxonic data simulated using a base rate of .08 yielded base rate estimates as high as .32 (MAMBAC) and .18 (MAXEIG) suggests that these procedures can substantially overestimate small base rates. This upward bias may be attributable to indicator skew, within-group correlations, or other factors. Second, the fact that each procedure yielded consistent base rate estimates for the dimensional data — and that consistency between the estimates of these two procedures was greatest for the dimensional data — once again advises caution in drawing a taxonic inference from coherent base rate estimates. As was noted earlier, skewed indicators can easily produce such coherence regardless of latent structure. Notably, although the MAMBAC and MAXEIG base rate estimates did not lend support to either a taxonic or a dimensional inference, the L-Mode estimates provided support for taxonic

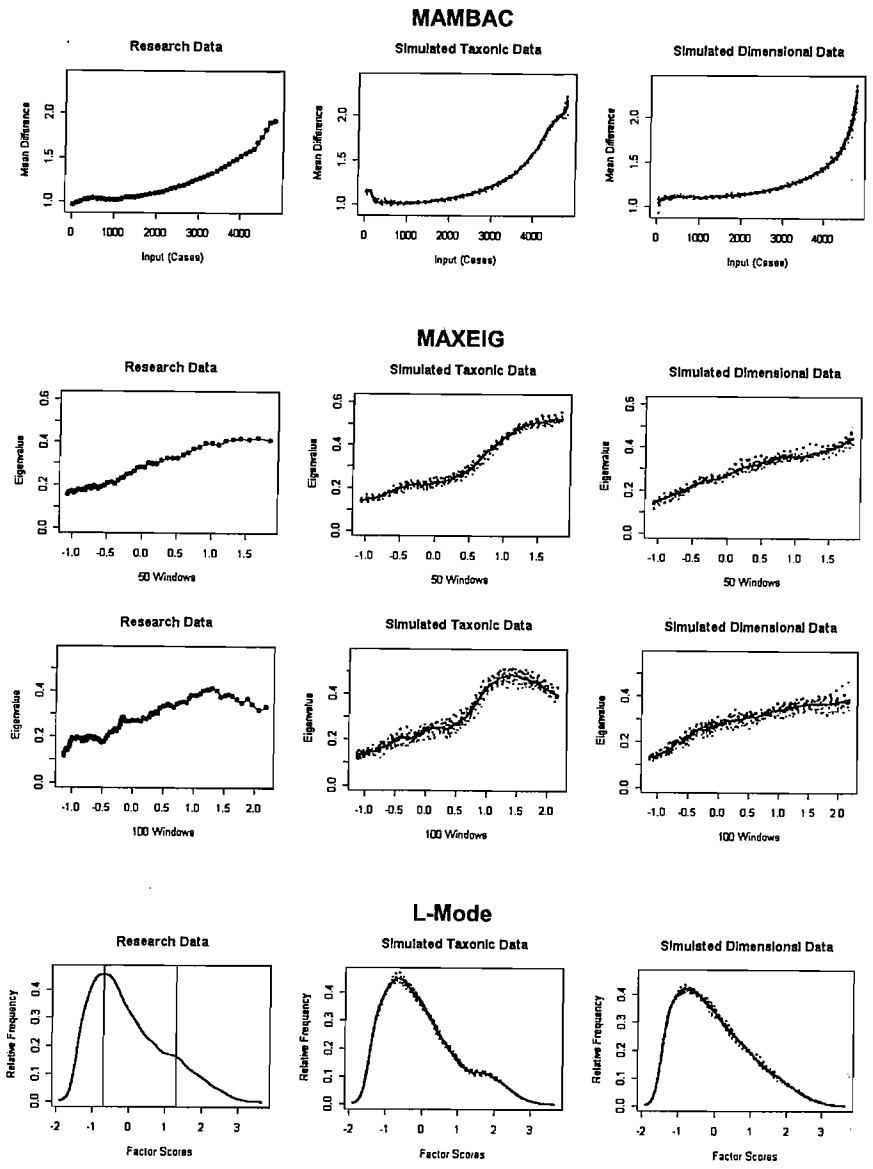

FIG. 6. Analyses of the three composite indicators constructed according to diagnostic criteria for GAD (N = 4,824). Taxonic comparison data were simulated by assigning the 376 cases (8%) who met criteria for GAD to the taxon and the remaining 4,448 cases (92%) to the complement. To conserve space, only the averaged curves are presented for each series of 4 MAMBAC and 6 MAXEIG analyses, all of which were performed using composite input indicators. MAXEIG was conducted first using 50 windows, and then using 100 windows. For the simulated comparison data, the solid line represents the average of the 10 individual curves (dotted lines).

structure, with the research data and the simulated taxonic data evidencing similar levels of consistency that were much higher than those evident for the simulated dimensional data.

Whereas the GFI values derived from the MAMBAC analyses were uniformly high and therefore uninformative, clear differences emerged for the GFI values derived from the MAXEIG and L-Mode analyses, with the pattern of results suggesting a dimensional inference for MAXEIG and a taxonic inference for L-Mode (see Table 4). Both of the curve-fit indices favored a dimensional inference for both procedures, though the magnitude of the Fit_d values for the MAMBAC analysis and the MAXEIG analysis with 100 windows were not as large as in the previous analyses.

In sum, taxometric results for the GAD indicators were inconclusive. The suitability tests were barely passed by some procedures and consistency tests, and not passed by others. Those tests that did pass the suitability test did not yield strong or consistent results, suggesting that neither a taxonic nor a dimensional conclusion is warranted. For a structural conclusion to be possible, follow-up taxometric analyses are needed using other indicator sets that are drawn from additional measures of GAD, constructed in different ways, and analyzed in different samples of data — ideally, samples with stronger representation of putative taxon members. It is important to emphasize that had the research data been analyzed without the parallel analysis of simulated comparison data, these results may have been interpreted in support of a small GAD taxon. However, because comparisons with simulated data revealed inconsistent support for taxonic structure, we would argue that no conclusion is justified at this time and that judgment should be withheld pending further research.

Conclusions

This overview was designed to introduce the major considerations involved in conducting a taxometric investigation and to stimulate interested readers to apply the taxometric method to constructs in their areas of specialization. Rigorous taxometric investigations have much to offer the science of clinical psychology, and careful attention to the checklist of conceptual and methodological issues presented here should maximize the informativeness of such research. In addition, as we emphasized throughout the paper, the informativness of taxometric research is particularly likely to be enhanced when thoughtful analyses are accompanied by the simulation and parallel analysis of taxonic and dimensional comparison data. We hope that our empirical illustrations provide a sense for how taxometric analysis may be implemented in practice, and hence promote its use by new investigators. Finally, we offer up our taxometrics program code — available with detailed documentation at http://www.etown.edu/psychology/faculty/ruscio.htm — with the hope that it will facilitate the sound application of the taxometric method to valuable new questions about latent structure.

Appendix A
Annotated Command Lines for all Taxometric Analyses

Biological Sex Analyses—Dichotomous Indicators

The data set is called "MMPI1," and it contains the 8 dichotomous items selected to serve as indicators plus a 9th column denoting each participant's actual sex (1 = male, 2 = female). Because the data were dichotomous, composite indicators had to be used for MAMBAC and MAXEIG, and L-Mode was not conducted. Prior to analysis, taxonic and dimensional comparison data were simulated.

```
MMPI1t <- SimTax(MMPI1)
MMPI1d <- SimDim(MMPI1[,1:8])
```

1. Perform MAMBAC with composite input indicators. Request parallel analyses of simulated comparison data to generate estimates of curve fit. To simulate taxonic data, use the supplied criterion variable. Note that the 9th column will not be submitted to MAMBAC analyses, it will only be used to simulated taxonic comparison data.

```
MAMBAC(MMPI1,Sim.Data=T,Supplied.Class=T,N.Samples=10,Ind.Comp=T)
```

2. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel MAMBAC analyses:

```
MAMBAC(MMPI1t[,1:8],Ind.Comp=T)
MAMBAC(MMPI1d[,1:8],Ind.Comp=T)
```

3. Perform MAXEIG with composite indicators and curve fitting via parallel analyses of simulated comparison data.

```
MAXEIG(MMPI1,Sim.Data=T,Supplied.Class=T,N.Samples=10,Ind.Comp=T)
```

4. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel MAXEIG analyses:

```
MAXEIG(MMPI1t[,1:8],Ind.Comp=T)
MAXEIG(MMPI1d[,1:8],Ind.Comp=T)
```

Biological Sex Analyses—Factor Score Indicators

The data set is called "MMPI2," and it contains a patient ID number (an arbitrary code added to the Hathaway Data Bank to denote multiple MMPIs completed by the same patient but that contains no identifying information), the 3 composite indicators generated through a factor analysis of the 24 most valid *Mf* items, and a 5th column denoting each participant's actual sex (1 = male, 2 = female). Before performing any analyses, taxonic and dimensional comparison data were simulated.

```
MMPI2t <- SimTax(MMPI2[,2:5])
MMPI2d <- SimDim(MMPI2[,2:4])
```

1. Perform MAMBAC with curve fitting via parallel analyses of simulated comparison data. To simulate taxonic data, use the supplied classification codes.

```
MAMBAC(MMPI2[,2:5],Sim.Data=T,Supplied.Class=T,N.Samples=10)
```

2. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel MAMBAC analyses:

```
MAMBAC(MMPI2t[,1:3])
MAMBAC(MMPI2d[,1:3])
```

3. Perform MAXEIG with curve fitting via parallel analyses of simulated comparison data.

```
MAXEIG(MMPI2[,2:5],Sim.Data=T,Supplied.Class=T,N.Samples=10)
```

4. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel MAXEIG analyses:

```
MAXEIG(MMPI2t[,1:3])
MAXEIG(MMPI2d[,1:3])
```

5. Perform L-Mode with parallel analyses of simulated comparison data; note that curve fitting is not performed for L-Mode. Note that because the height of the visually apparent lower mode was less than the height of the curve at $x = 0$, the location of the lower mode was set manually.

```
LMode(MMPI2[,2:4],Mode.L=-1,N.Samples=10)
```

6. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel L-Mode analyses, locating visually apparent modes manually as needed:

```
LMode(MMPI2t[,1:3],Mode.L=-1.2)
LMode(MMPI2d[,1:3])
```

PTSD Analyses

The data set is called "Miss," and it contains just the 4 composite indicators. Before performing any analyses, taxonic and dimensional comparison data were simulated. Whereas the simulation of dimensional comparison data required just a single command (a), the simulation of taxonic comparison data required the creation of a criterion variable. This was done in several steps: (b) an additional column representing the total score on the 4 indicators was created, (c) the data were sorted by total scores, (d) a vector of classification codes (1 = complement, 2 = taxon) was created, and (e) the classification codes were appended to the data set. Finally, the taxonic comparison data were simulated (f).


```
(a) Missd <- SimDim(Miss[,1:4])
(b) Miss <-cbind(Miss,(Miss[,1]+Miss[,2]+Miss[,3]+Miss[,4]))
(c) Miss <- Miss[sort.list(Miss[,5]),]
(d) Class <- c(rep(1,340),rep(2,723))
(e) Miss <- cbind(Miss,Class)
(f) Misst <- SimTax(Miss[,c(1:4,6)])
```

1. Perform MAMBAC with composite input indicators and curve fitting via parallel analyses of simulated comparison data. To simulate taxonic data, assign the 68% of cases scoring the highest on the sum of the 4 indicators to the taxon and the remaining 32% of cases to the complement.

```
MAMBAC(Miss[,1:4],Sim.Data=T,Supplied.P=.68,N.Samples=10,Ind.Comp=T)
```

2. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel MAMBAC analyses:

```
MAMBAC(Misst[,1:4],Ind.Comp=T)
MAMBAC(Missd[,1:4],Ind.Comp=T)
```

3. Perform MAXEIG with curve fitting via parallel analyses of simulated comparison data. For the estimation of latent parameters, assign cases using the base-rate classification method. This was done because in an initial analysis, several of the curves reached their maximum eigenvalue in the first window, yielding nonsensical estimates of the valid and false positive rates for 3 of the 4 indicators (1.00 in each case). Thus, the default classification method (Bayes' Theorem) would have been inappropriate.

```
MAXEIG(Miss[,1:4],Sim.Data=T,Supplied.P=.68,N.Samples=10,Classify=1)
```

4. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel MAXEIG analyses:

```
MAXEIG(Misst[,1:4],Classify=1)
MAXEIG(Missd[,1:4],Classify=1)
```

5. Perform L-Mode with parallel analyses of simulated comparison data.

```
LMode(Miss[,1:4],Sim.Data=T,Supplied.P=.68,N.Samples=10)
```

6. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel L-Mode analyses, locating visually apparent modes manually if necessary:

```
LMode(Misst[,1:4],Mode.L=-1)
LMode(Missd[,1:4])
```

GAD Analyses

The data set is called "GAD," and it contains an arbitrary participant ID number; 4 composite indicators created according to DSM-IV criteria A, B,

C, and E; an additional composite indicator created by collapsing across criteria A and B; and GAD diagnostic status (1 = complement or non-GAD, 2 = taxon or GAD). Thus, columns 4 – 6 comprise the 3 composite indicators representing criteria C, E, and A/B. Before performing any analyses, taxonic and dimensional comparison data were simulated.

```
GADt <- SimTax(GAD[,4:7])
GADd <- SimDim(GAD[,4:6])
```

1. Perform MAMBAC with composite input indicators and curve fitting via parallel analyses of simulated comparison data. To simulate taxonic data, use the supplied classification codes.

```
MAMBAC(GAD[,4:7],Sim.Data=T,Supplied.Class=T,N.Samples=10,Ind.Comp=T)
```

2. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel MAMBAC analyses:

```
MAMBAC(GADt[,1:3],Ind.Comp=T)
MAMBAC(GADd[,1:3],Ind.Comp=T)
```

3. Perform MAXEIG with curve fitting via parallel analyses of simulated comparison data. Analyses were repeated with twice as many windows (100, rather than the default of 50) to implement the inchworm consistency test.

```
MAXEIG(GAD[,4:7],Sim.Data=T,Supplied.Class=T,N.Samples=10)
MAXEIG(GAD[,4:7],Sim.Data=T,Supplied.Class=T,Windows=100,N.Samples=10)
```

4. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel MAXEIG analyses:

```
MAXEIG(GADt[,1:3],Windows=100)
MAXEIG(GADd[,1:3],Windows=100)
```

5. Perform L-Mode with parallel analyses of simulated comparison data.

```
LMode(GAD[,4:7],Sim.Data=T,Supplied.Class=T,N.Samples=10,Mode.R=1.3)
```

6. To perform full analyses of the simulated comparison data and to generate estimates of latent parameters, submit these data sets to parallel L-Mode analyses, locating visually apparent modes manually if necessary:

```
LMode(GADt[,1:3],Mode.R=1.5)
LMode(GADd[,1:3])
```


Appendix B
Sample Output: MAXEIG Analysis of MMPI#2

```
> MAXEIG(MMPI2[,2:5],Sim.Data=T,Supplied.Class=T,N.Samples=10)
SUMMARY OF MAXEIG ANALYTIC SPECIFICATIONS

Sample size: 13580
Number of indicator variables: 3
Calculation method: Eigenvalues
Replications: 1
Inchworm consistency test: No
Subsamples: 50 windows with 0.9 overlap
n per window at 50 windows: 2302
Indicators: Each variable serves once as input, with all other variables as outputs
Total number of curves: 3
Y values smoothed for graphing and estimation: No
Base rate estimation: Adapted general covariance mixture theorem
Classification of cases: Bayes' Theorem

SUMMARY OF MAXEIG PARAMETER ESTIMATES

Estimated hitmax values and taxon base rates for each curve:

        Hitmax     P
Curve 1 -0.430 0.629
Curve 2 -0.259 0.612
Curve 3 -1.065 0.698

Summary of base rate estimates across curves:
   M = 0.646
  SD = 0.045

Estimated VP, FP values at each indicator's hitmax cut:

              VP    FP
Indicator 1 0.879 0.341
Indicator 2 0.861 0.292
Indicator 3 0.952 0.781

Base rate estimate for averaged curve = 0.626

Estimated latent group M, SD, validity on each indicator:

             Taxon M  Taxon SD    Comp M    Comp SD  Validity (Raw) Validity (SD)
Indicator 1    0.368     0.853    -0.957     0.671           1.326          1.643
Indicator 2    0.352     0.879    -0.933     0.642           1.284          1.565
Indicator 3    0.270     0.929    -0.698     0.823           0.968          1.075

Summary of estimated indicator validities (in SD units):
   M = 1.428
  SD = 0.308

Indicator correlations:

Full Sample ( N = 13580 ):
           Indicator 1 Indicator 2 Indicator 3
Indicator 1       1.000       0.298       0.251
Indicator 2       0.298       1.000       0.264
Indicator 3       0.251       0.264       1.000
Taxon ( n = 9829 ):
           Indicator 1 Indicator 2 Indicator 3
Indicator 1       1.000      -0.087       0.074
Indicator 2      -0.087       1.000       0.079
Indicator 3       0.074       0.079       1.000
Complement ( n = 3751 ):
           Indicator 1 Indicator 2 Indicator 3
Indicator 1       1.000       0.048      -0.313
Indicator 2       0.048       1.000      -0.213
Indicator 3      -0.313      -0.213       1.000
```

```
Summary of indicator correlations:
                 M      SD
Full Sample   0.271  0.024
Taxon         0.022  0.094
Complement   -0.159  0.186

Average indicator validity (in SD units) = 1.308
  * Estimated using the Meehl & Yonce (1996, p. 1146) formula.

Goodness of Fit Index (GFI) = 1

SUMMARY OF SIMULATED DATA CURVE FITTING
              Taxon: N = 7994 , Largest residual r = 0 , RMSR r = 0
         Complement: N = 5586 , Largest residual r = 0 , RMSR r = 0
    Taxonic data set: Largest residual r = 0 , RMSR r = 0

              Taxon: N = 7994 , Largest residual r = 0 , RMSR r = 0
         Complement: N = 5586 , Largest residual r = 0 , RMSR r = 0
    Taxonic data set: Largest residual r = 0 , RMSR r = 0

              Taxon: N = 7994 , Largest residual r = 0 , RMSR r = 0
         Complement: N = 5586 , Largest residual r = 0 , RMSR r = 0
    Taxonic data set: Largest residual r = 0 , RMSR r = 0

              Taxon: N = 7994 , Largest residual r = 0 , RMSR r = 0
         Complement: N = 5586 , Largest residual r = 0 , RMSR r = 0
    Taxonic data set: Largest residual r = 0 , RMSR r = 0

              Taxon: N = 7994 , Largest residual r = 0 , RMSR r = 0
         Complement: N = 5586 , Largest residual r = 0 , RMSR r = 0
    Taxonic data set: Largest residual r = 0 , RMSR r = 0

              Taxon: N = 7994 , Largest residual r = 0 , RMSR r = 0
         Complement: N = 5586 , Largest residual r = 0 , RMSR r = 0
    Taxonic data set: Largest residual r = 0 , RMSR r = 0

              Taxon: N = 7994 , Largest residual r = 0 , RMSR r = 0
         Complement: N = 5586 , Largest residual r = 0 , RMSR r = 0
    Taxonic data set: Largest residual r = 0 , RMSR r = 0

              Taxon: N = 7994 , Largest residual r = 0 , RMSR r = 0
         Complement: N = 5586 , Largest residual r = 0 , RMSR r = 0
    Taxonic data set: Largest residual r = 0 , RMSR r = 0

              Taxon: N = 7994 , Largest residual r = 0 , RMSR r = 0
         Complement: N = 5586 , Largest residual r = 0 , RMSR r = 0
    Taxonic data set: Largest residual r = 0 , RMSR r = 0

              Taxon: N = 7994 , Largest residual r = 0 , RMSR r = 0
         Complement: N = 5586 , Largest residual r = 0 , RMSR r = 0
    Taxonic data set: Largest residual r = 0 , RMSR r = 0
Dimensional data set: N = 13580 , Largest residual r = 0 , RMSR r = 0
Dimensional data set: N = 13580 , Largest residual r = 0 , RMSR r = 0
Dimensional data set: N = 13580 , Largest residual r = 0 , RMSR r = 0
Dimensional data set: N = 13580 , Largest residual r = 0 , RMSR r = 0
Dimensional data set: N = 13580 , Largest residual r = 0 , RMSR r = 0
Dimensional data set: N = 13580 , Largest residual r = 0 , RMSR r = 0
Dimensional data set: N = 13580 , Largest residual r = 0 , RMSR r = 0
Dimensional data set: N = 13580 , Largest residual r = 0 , RMSR r = 0
Dimensional data set: N = 13580 , Largest residual r = 0 , RMSR r = 0
Dimensional data set: N = 13580 , Largest residual r = 0 , RMSR r = 0
Fit(RMSR) of averaged curves:
Simulated taxonic data: 0.041
Simulated dimensional data: 0.086
Fit(d) of 10 simulated curves for each structure: -9.145
  * Negative values are more supportive of taxonic structure.
```


References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Beauchaine, T. P. (2003). Taxometrics and developmental psychopathology. *Development and Psychopathology, 15*, 501–527.

Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometric Bulletin, 2*, 47–53.

Cleland, C., Rothschild, L., & Haslam, N. (2000). Detecting latent taxa: Monte Carlo comparison of taxometric, mixture and clustering methods. *Psychological Reports, 87*, 37–47.

Cleveland, W. S. (1979). Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*, 829–836.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249–253.

Cronbach, L. J., & Meehl, P. E. (1995). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 8*, 341–349.

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

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.

Grove, W. M. (1991a). 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. (1991b). When is a diagnosis worth making? A statistical comparison of two prediction strategies. *Psychological Reports, 68*, 3–17.

Grove, W. M., & Meehl, P. E. (1993). Simple regression-based procedures for taxometric investigations. *Psychological Reports, 73*, 707–737.

Haslam, N. (1997). Evidence that male sexual orientation is a matter of degree. *Journal of Personality and Social Psychology, 73*, 862–870.

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.

Joiner, T. E., Jr., & Schmidt, N. B. (2002). Taxometrics can "do diagnostics right" (and isn't quite as hard as you think). In L. E. Beutler & M. L. Malik (Eds.), *Rethinking the DSM: A psychological perspective* (pp. 107–120). Washington, DC: American Psychological Association.

Keane, T. M., Caddell, J. M., & Taylor, K. L. (1988). Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology, 56*, 85–90.

Kessler, R. S. (2002). Epidemiological perspectives for the development of future diagnostic systems. *Psychopathology, 35*, 158–161.

Klein, D. N., & Riso, L. P. (1993). Psychiatric disorders: Problems of boundaries and comorbidity. In C. G. Costello (Ed.), *Basic issues in psychopathology* (pp. 19–66). New York: The Guilford Press.

Lilienfeld, S. O., Waldman, I. D., & Israel, A. C. (1994). A critical examination of the use of the term "comorbidity" in psychopathology research. *Clinical Psychology: Science and Practice, 1*, 71–83.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113*, 181–190.

Meehl, P. E. (1973). MAXCOV-HITMAX: A taxonomic search method for loose genetic syndromes. In P. E. Meehl (Ed.), *Psychodiagnosis: Selected papers* (pp. 200–224). Minneapolis: University of Minnesota Press.

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. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50*, 266–274.

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.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28*, 487–495.

Murphy, E. A. (1964). One cause? Many causes? The argument from the bimodal distribution. *Journal of Chronic Disease, 17*, 301–324.

Newman, M. G., Zuellig, A. R., Kachin, K. E., Constantino, M. J., Przeworski, A., Erickson, T., & Cashman-McGrath, L. (2002). Preliminary reliability and validity of the Generalized Anxiety Disorder Questionnaire–IV: A revised self-report diagnostic measure of generalized anxiety disorder. *Behavior Therapy, 33*, 215–233.

Ruscio, A. M., Borkovec, T. D., & Ruscio, J. (2001). A taxometric analysis of the latent structure of worry. *Journal of Abnormal Psychology, 110*, 413–422.

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. (2004). *Estimating taxon base rates using the MAXCOV, MAXEIG, and MAMBAC taxometric procedures: A Monte Carlo study*. Manuscript in preparation.

Ruscio, J., Haslam, N., & Ruscio, A. M. (2004). *Introduction to the taxometric method: A practical guide*. Manuscript in preparation.

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. (in press). A nontechnical introduction to the taxometric method. *Understanding Statistics*.

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.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage.

Watson, D. (2003). Investigating the construct validity of the dissociative taxon: Stability analyses of normal and pathological dissociation. *Journal of Abnormal Psychology, 112*, 298–305.

Weathers, F. W., Ruscio, A. M., & Keane, T. M. (1999). Psychometric properties of nine scoring rules for the Clinician-Administered Posttraumatic Stress Disorder Scale. *Psychological Assessment, 11*, 124–133.

Widiger, T. A. (2001). What can be learned from taxometric analyses? *Clinical Psychology: Science and Practice, 8*, 528–533.

RECEIVED: June 17, 2003
ACCEPTED: August 6, 2003