

Differentiating Categories and Dimensions: Evaluating the Robustness of Taxometric Analyses

John Ruscio and Walter Kaczetow
The College of New Jersey

Interest in modeling the structure of latent variables is gaining momentum, and many simulation studies suggest that taxometric analysis can validly assess the relative fit of categorical and dimensional models. The generation and parallel analysis of categorical and dimensional comparison data sets reduces the subjectivity required to interpret results by providing an objective Comparison Curve Fit Index (CCFI). This study takes advantage of developments in the generation of comparison data to examine the robustness of taxometric analyses to unfavorable data conditions. Very large comparison data sets are treated as populations from which many samples are drawn randomly, placing the method on a firmer statistical foundation and increasing its run-time efficiency. The impressive accuracy of the CCFI was consistent with prior findings and robust across novel manipulations of asymmetry, tail weight, and heterogeneous variances. Analyses, an empirical illustration using Minnesota Multiphasic Personality Inventory (MMPI) hypochondriasis data, and discussion focus on the practical implications for differentiating categories and dimensions.

Do people vary along a continuum of hypochondriasis, or are some people hypochondriacs and others not? Do the categories of heterosexual and homosexual accurately model observed human sexual orientations, or does variation along one or more dimensions better fit the data? In many research domains, significant theoretical and practical implications follow from empirical tests of the relative fit of categorical and dimensional models of observed data (Meehl, 1992; Ruscio, Haslam, & Ruscio, 2006). Should causal theories account for the bifurcation of individuals into groups or for their variation along a continuum?

Correspondence concerning this article should be addressed to John Ruscio, Department of Psychology, The College of New Jersey, Ewing, NJ 08628. E-mail: ruscio@tcnj.edu

Should constructs be assessed using measures based on latent class or latent trait models? Should individuals be classified using a categorical or a dimensional system? Issues like these arise in many areas of basic and applied behavioral science (Ruscio & Ruscio, 2008), and they receive special attention in the study of abnormal behavior as new editions of the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* are developed (Frances et al., 1991; Widiger & Trull, 2007). Meehl and colleagues (e.g., Meehl, 1995a; Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998) developed a family of taxometric data-analytic procedures to help researchers make the deceptively simple determination of whether the latent structure of a construct is best characterized as taxonic or dimensional.¹ Dimensional latent structure is modeled as a single latent class within which a set of variables covary due to shared loadings on one or more latent factors according to the common factor model (e.g., McDonald, 1999). Taxonic latent structure is modeled as two latent classes, referred to as taxon and complement, distinguished by a set of variables whose means differ across classes and that may covary within one or both classes due to shared loadings on one or more common latent factors.² Assessing the relative fit of these structural models for an empirical data set can be challenging. Simply examining observed score distributions can yield misleading conclusions for a number of reasons (e.g., score differences between classes may be too small for bimodality to emerge, chance variation may create the appearance of multiple modes; see Murphy, 1964). Even though variables that validly differentiate two latent classes should correlate in a sample containing members of both classes, the same is true of variables that load onto one or more latent dimensions (Bartholomew, 1987).

Rather than examining variables' distributions or covariances in a full sample of data, some taxometric procedures compute functions within ordered subsamples of data. For example, the MAXEIG (MAXimum EIGenvalue; Waller & Meehl, 1998) procedure plots the first eigenvalue of a covariance matrix (the usual variance-covariance matrix with 0s placed along the diagonal to leave only covariances) for a set of variables within each of a series of ordered subsamples created using another variable. If two classes are mixed in the full sample, this would be expected to yield larger eigenvalues within subsamples containing a

¹Meehl and his colleagues sometimes described the taxometric method as a tool to test the relative fit of two competing structural models (e.g., Meehl, 2004) but at other times described it as a tool to detect taxonic structure (e.g., Meehl, 1999). We acknowledge that there is some inconsistency in how this has been treated in the literature on taxometrics and discuss the issue in a section on inferential frameworks.

²We prefer to allow nonzero within-group correlations in the model representing categorical latent structure. This departs somewhat from the conventional approach in which, as a useful simplifying approximation, Meehl and colleagues usually assumed within-group correlations to be zero. Meehl (1995b) discussed ways to deal with the complexity introduced by relaxing this assumption, and some simulation studies of the performance of taxometric procedures have examined their robustness to nonzero within-group correlations (e.g., Ruscio, Ruscio, & Meron, 2007).

mixture of taxon and complement members than in subsamples more homogeneous with respect to class membership. Thus, Waller and Meehl describe the prototypical MAXEIG curve for taxonic data as peaked; for dimensional data, they suggest that no peak is expected.

Meehl's first technical reports on the taxometric method were written in the 1960s, and research interest has gained momentum in recent years. Ruscio et al. (2006) identified 57 published studies that used the method to examine constructs in research areas such as abnormal, personality, developmental, and social psychology. As of July 20, 2008, a more comprehensive bibliography of methodological and applied taxometric research at www.taxometricmethod.com identified more than 100 published studies spanning a broader range of topics. Among the 200+ total entries in this bibliography, more than two thirds were published from 2000 onward. Taxometric analyses are performed with increasing frequency in an expanding range of research areas. To our knowledge, the analysis of empirical data on hypochondriasis that we present toward the end of this article constitutes the first taxometric examination of this construct's latent structure. Following the approach taken in our simulation study and using its findings to help guide the interpretation of results, we analyze a large sample of responses to the 33 items on the Hypochondriasis scale of the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) to test the relative fit of models representing categorical and dimensional latent structures.

THE INTERPRETATION OF TAXOMETRIC RESULTS

Although the popularity of the taxometric method has increased, concerns about the method have been raised. One concern involves the subjective interpretation of taxometric curves produced by procedures such as MAXEIG. Prototypical curve shapes for idealized taxonic and dimensional data are available, but taxometric results for empirical data may or may not resemble them and difficult subjective judgments may be required. This issue has been addressed through the parallel analysis of taxonic and dimensional comparison data. Ruscio et al. (2007) described this approach, provided tools for implementing it using a multivariate bootstrap technique,³ and showed that the objectively calculated

³The use of the term "bootstrap" throughout the present article refers to the resampling strategy introduced by Efron (1979) to study the performance of a statistical procedure using data-based estimates of the relevant population parameters. Cronbach and Meehl (1955) used the term to describe the process by which one could develop measures more valid than the criterion variables that were originally used to validate them, and Meehl (1995a) has described his taxometric method in a similar way. Thus, the parallel analysis of comparison data in a taxometric study involves both types of bootstrapping. For excellent discussions of univariate and multivariate bootstrapping and studies that illustrate the importance of choosing the most appropriate technique to attain a particular goal, see Beasley et al. (2007) and Lee and Rodgers (1998).

Comparison Curve Fit Index (CCFI) can distinguish categorical and dimensional data with an impressive degree of validity; these results were replicated and extended in a study by Ruscio (2007). Beach, Amir, and Bau (2005) suggested that this approach introduced bias in the reproduction of data characteristics and failed to identify small taxa in a series of artificial data sets that they created, but a reanalysis by Ruscio and Marcus (2007) showed that neither of these claims was supported. Using comparison data alleviates concerns about subjective judgments by providing an objective index as an interpretive aid.

THE INFERENTIAL FRAMEWORK FOR TAXOMETRICS

Another concern strikes at more fundamental issues: What is the goal of a taxometric analysis, and what conclusions can be drawn from its results? The taxometric literature has not been sufficiently clear on this point, and reasonable people can disagree about what is implied. For example, the subtitle of Waller and Meehl's (1998) book on taxometrics, "Distinguishing Types from Continua," suggests that the goal is to test the relative fit of competing structural models. On the other hand, the final chapter presents "guidelines for corroborating taxonic models" (pp. 92–93), which suggests that the goal is to detect a taxon. Careful reading of this literature will provide many instances of apparent support for what can be called the "competing-models" and the "taxon-detection" inferential frameworks. Both Ruscio et al. (2006) and Ruscio (2007) discussed this issue, and we provide a brief overview to explain why we adopt the competing-models framework in our work.

Maraun and colleagues (e.g., Maraun & Slaney, 2005; Maraun, Slaney, & Goddyn, 2003) have outlined a methodological approach that adopts the taxon-detection inferential framework. They distinguish between the target of detection (taxonic latent structure), tools of detection (taxometric procedures, such as MAXEIG), and assumptions (conditions necessary for the tool to function properly). Their approach has been analytic, beginning with a definition of taxonic latent structure and then deriving expected results (e.g., curve shapes) for various taxometric procedures given specific assumptions have been met. This is analogous to null hypothesis significance testing and entails similar dilemmas in its logic and application (e.g., Cohen, 1994; Nickerson, 2000). For example, if tests of a model fail to meet criteria to support an inference of taxonic latent structure, there seems to be no way to determine whether a model of dimensional structure provides better fit or the data are inadequate for distinguishing between these structural possibilities. This is analogous to the problem of low statistical power. We believe that reconceptualizing taxometrics from a model comparison perspective, akin to that of Maxwell and Delaney (2004), affords more nuanced conclusions that address more directly the questions motivating researchers.

Whether in the behavioral, physical, or other sciences, the goal of a model is to represent something real and simplify it in a useful way. A model is not intended to capture the full complexity of its subject matter, and sufficiently sensitive tests are almost certain to reveal its imperfections. A model should be evaluated based on its utility for a specified purpose, and this implicitly involves comparisons with what can be achieved either without the model or with rival models. For example, by using Newton's laws one can estimate the velocities of vehicles involved in crashes from their masses and the distances they traveled after impact. Without this mathematical model, estimates would be much less accurate and important issues (e.g., traffic violations, insurance liability) more difficult to resolve. Einstein's theory of relativity provides a rival model that could be used to obtain more accurate estimates, but the added complexity is unwarranted because any differences between the two models' estimates would be far less than the margin of error associated with fallible measurements of mass and distance.

Empirically testing the latent structure of a psychological construct likewise involves the search for a model that serves its purpose better than available alternatives. Many theoretical and applied issues hinge on the determination of whether a construct is represented better as taxonic or dimensional, and one can use the taxometric method to test the relative fit of these two competing models of latent structure. Evidence might suggest that a model of categorical structure fits the data better, that a model of dimensional data fits the data better, or that neither of these models fits much better than the other (Meehl, 2004). Thus, one can reach an affirmative conclusion supporting categorical structure, an affirmative conclusion supporting dimensional structure, or an ambiguous conclusion that is analogous to identifying inadequate statistical power to differentiate between the models representing the structures of interest. Though we prefer the competing-models framework, it remains debatable whether this is incompatible with the taxon-detection framework and we do not believe it is necessary to take a strong position on this issue. Any real or apparent differences between these inferential frameworks may have little bearing on the utility of the approach to performing taxometric analyses accompanied by parallel analyses of categorical and dimensional comparison data that is the focus of the present research.

ROBUSTNESS TO UNFAVORABLE DATA CONDITIONS

We characterized the models of taxonic and dimensional structures as similar in allowing dimensional variation within classes, with the key distinction being whether a two-class or one-class model better fits the data. To test the fit of these models to empirical data, the unique characteristics of the data can be

reproduced using a dimensional model (with dimensionality equal to a value estimated from the data), and then the same can be done using a taxonic model (with dimensionality equal to that estimated within each of the putative groups). Then, the relative fit of these two models is compared by evaluating whether the results for the empirical data are more similar to those for comparison data generated using the dimensional or the taxonic model (Ruscio et al., 2007). If one model fits appreciably better than the other, one concludes that the structure represented by that model underlies the data.

Whereas Maraun and colleagues, operating under the taxon-detection framework, follow the lead of Meehl and colleagues by placing great emphasis on analytic derivations of expected curve shapes (e.g., flat, peaked, double-peaked), our approach to implementing taxometrics under the competing-models framework does not require such expectations. Rather, we interpret results by making comparisons to data known to be taxonic or dimensional because they were generated using these structural models. Whatever curve shapes may emerge, the interpretation is based on relative similarities and differences between the results for empirical and comparison data, as quantified using the CCFI (Ruscio et al., 2007). Both taxonic and dimensional comparison data reproduce key properties of the empirical data, such as indicators' distributions and correlations, thereby holding these properties constant across data used to test the relative fit of competing models. Nonetheless, it is important to evaluate the utility of taxometric procedures under a wide range of data conditions. It is easy to show that a procedure functions well under idealized conditions, but its robustness to unfavorable conditions is crucial.

In many simulation studies (e.g., Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998), taxonic data have been created such that variables are normally distributed within classes. These exclusively normal distributions may have yielded taxometric results more favorable than one would obtain under a wider range of conditions (Maraun & Slaney, 2005; Maraun et al., 2003). Micceri (1989) found that normal distributions are the exception rather than the rule, and Ruscio, Ruscio, and Keane (2004) demonstrated that positively skewed variables can make it difficult to distinguish taxometric results for dimensional data from those for a small taxon. Some simulation studies have included nonnormal data (e.g., Cleland & Haslam, 1996; Haslam & Cleland, 1996; Ruscio, 2007; Ruscio et al., 2007), and taxometric procedures exhibited varying degrees of robustness to skew. At present, nothing is known about the relative influence of asymmetry and tail weight as two ways in which variables' distributions can deviate from normality. Maraun and colleagues also distinguished between conditions of normality with equal variances and normality with unequal variances, deriving the expected curve shapes for taxonic data under each condition. We examine the performance of taxometric procedures across a range of variance ratios.

THE PRESENT STUDY

This study was designed to serve two purposes. First, it evaluates the robustness of taxometric analyses to unfavorable data conditions. In addition to manipulating factors whose influence on taxometric results has been studied—sample size, number of variables, taxon base rate, separation between classes, and within-group correlations—additional factors are manipulated to assess the influence of unequal variances and to tease apart the influence of asymmetry and tail weight as two aspects of nonnormal distributions. Data conditions include a wide range of deviations from idealized values.

The second purpose of the study is to introduce and appraise some procedural refinements to the generation and parallel analysis of comparison data. The technique presented by Ruscio et al. (2007) uses bootstrap methods (Efron & Tibshirani, 1993) and an iterative algorithm to reproduce the characteristics of an empirical data set, first using a dimensional structural model and then using a taxonic structural model. Several rigorous simulation studies demonstrated that the programs to generate comparison data reproduce data characteristics with good precision and negligible, if any, bias. However, one limitation of the way that the approach has been implemented is that some data characteristics will not vary from sample to sample or will vary less than would be expected by normal sampling error. For example, the taxon base rate in each sample of taxonic comparison data is identical to that in the empirical data. Also, the observed correlation matrices (in the full sample for dimensional comparison data, within groups for taxonic comparison data) are reproduced as precisely as the iterative algorithm can achieve for each sample of comparison data; this yields less variability in correlations than would be expected with random sampling from an appropriate population. Advances in computing power and improvements in the efficiency of the computer code (see Ruscio & Kaczetow, 2008) make it possible to generate much larger samples of comparison data, sufficiently large that they can be treated as populations. Thus, rather than generating a series of B bootstrap samples of size N , where N is the sample size of the empirical data, one can generate a very large sample, treat this as a population, and draw B samples of size N from it.

This implementation of the approach offers three significant advantages, both conceptual and practical. First, generating a very large sample enables the algorithm to reproduce data characteristics with excellent precision. One can treat this as a population with the characteristics of the empirical data. Second, drawing samples at random from a population introduces normal sampling error with respect to all data characteristics, including the taxon base rate and indicator correlations. This places the method on a firmer statistical foundation. Third, whereas the iterative algorithm is computationally intensive, drawing samples from a population is not. Generating one population using each struc-

tural model—taxonic and dimensional—and then drawing B samples can run much more quickly than generating a series of B samples using each structural model. Whereas the previous implementation executed the iterative algorithm $2B$ times, prohibiting the routine use of large values of B , the new implementation only executes the iterative algorithm twice, regardless of B . Previous simulation studies of the generation and parallel analysis of comparison data have used $B = 10$ (Ruscio, 2007; Ruscio & Marcus, 2007; Ruscio et al., 2007). We compare results obtained with $B = 10, 25, \text{ or } 50$ samples of comparison data.

METHOD

Design and Data Generation

A total of 25,000 taxonic and dimensional data sets (12,500 for each structure) were generated using a Monte Carlo design in which data parameters were independently randomly sampled. The data conditions were similar to those of Ruscio (2007, Study 2) and Ruscio et al. (2007, Study 3) but extended to serve the first purpose of this study—examining robustness to unfavorable data conditions. For taxonic target data,⁴ random values were drawn for the following parameters: sample size ($N = 300$ to 1,000), number of variables (which are referred to as *indicators* in the taxometric literature; $k = 3$ to 8), taxon base rate ($P = .10$ to $.50$), indicator validity (standardized mean difference of $d = 1.25$ to 2.00), within-group correlation ($r = .00$ to $.30$), asymmetry ($g = .00$ to $.30$), tail weight ($h = .00$ to $.15$), and taxon:complement variance ratio ($VR = .25$ to 4.00).

Values of $N, k, P, d, r, g,$ and h were drawn from uniform distributions (continuous for all but k , which was discrete) spanning the ranges listed earlier. The value of VR was determined by drawing a random value X from a uniform, continuous distribution ranging from 1 to 4; with probability $.50$, $VR = X$, and with probability $.50$, $VR = 1/X$. The values of g and h were used to generate data from a g -and- h distribution (Hoaglin, 1985, p. 486). The magnitude of g controls the asymmetry relative to a normal distribution (in which $g = 0$), and the magnitude of h controls the tail weight relative to a normal distribution (in which $h = 0$). Because positive values of g and h were used, conditions of positive skew and heavy tail weight were studied. For the g -and- h populations

⁴Throughout this article, the term “target data” refers to one of the 25,000 samples generated in the simulation study and “empirical data” refers to samples of research data in an actual taxometric study. “Comparison data” refers to taxonic or dimensional data generated for parallel analysis to help interpret results for target data (in the simulation study) or empirical data (in an actual taxometric study).

used in this study, the smallest skew (γ_1) and kurtosis (γ_2) values were $\gamma_1 = 0$, $\gamma_2 = 0$ for $g = 0$, $h = 0$ and the largest values were $\gamma_1 = 2.60$, $\gamma_2 = 38.89$ for $g = .30$ and $h = .15$; other pairings of g and h correspond to γ_1 and γ_2 values within this range.⁵

To generate a taxonic data set, the iterative technique of Ruscio and Kacetow (2008) was used to sample N cases from a g -and- h distribution with $\mu = 0$, $\sigma = 1$, and a correlation matrix in which all indicators correlated r with one another. Next, a proportion P of cases was randomly selected and identified as taxon members, with the remainder identified as members of the complement class. The variance ratio was achieved by multiplying scores in the taxon by X (when taxon variance exceeded complement variance) or $1/X$ (when complement variance exceeded taxon variance). Then, the desired standardized mean difference d was achieved by adding a constant to taxon members' scores.

For dimensional data, values of N through VR were drawn in the same way. However, because P , d , and r do not correspond to parameters of the dimensional (common factor) model, they were combined to yield an expected indicator correlation using the following formula (Meehl & Yonce, 1994): $r_{xy} = (P[1 - P]d^2 + r)/(P[1 - P]d^2 + 1)$. The iterative algorithm of Ruscio and Kacetow (2008) was used to sample N cases from a g -and- h distribution with $\mu = 0$, $\sigma = 1$, and a correlation matrix in which all indicators correlated r_{xy} with one another. The data generation program was checked extensively.

Data Analysis

MAXEIG was performed on each data set using 50 windows that overlapped 90% with their neighbors. Whereas researchers are encouraged to select an appropriate number of windows based on the characteristics of their data (see Ruscio et al., 2006), we held the number of windows constant to provide a conservative test; the optimal number of windows may be more or less than 50 for many data conditions. A series of MAXEIG analyses was performed for each target data set by using each variable once as the input indicator, with all other variables serving as output indicators (Waller & Meehl, 1998). This yielded k MAXEIG curves, each of which was used to calculate an estimate of the taxon base rate using the General Covariance Mixture Theorem (Ruscio et al., 2006, Appendix C).

For each target data set, $B = 50$ samples of taxonic comparison data and $B = 50$ samples of dimensional comparison data were generated. To generate taxonic comparison data, the base-rate classification technique was used to assign

⁵A figure showing a series of population distributions with g -and- h values spanning the ranges used in this study is available on request.

cases to groups because prior research suggests that this approach yields CCFI values with good validity for taxonic and dimensional target data (see Ruscio, 2007; Ruscio & Marcus, 2007; Ruscio et al., 2007). Specifically, the M estimate of the taxon base rate from the MAXEIG analyses of the target data was used to classify cases into the putative taxon (higher scoring) and complement (lower scoring) groups by rank-ordering cases according to their indicator total scores, then applying a threshold corresponding to the proportion of the sample estimated to be taxon members. Then, MAXEIG was performed for each sample of comparison data.

Curves were averaged for each series of MAXEIG analyses, which yielded one curve for target data and B curves for each type of comparison data. Curves for taxonic and dimensional comparison data were averaged for the first 10 samples of comparison data, for the first 25 samples, and for all 50 samples. At each level of B , the CCFI was calculated as shown in Ruscio et al. (2007). CCFI values can range from 0 to 1, with lower values indicative of dimensional structure and higher values indicative of taxonic structure. Because a CCFI value of .50 corresponds to equivalent fit for both structures, this was used as an a priori threshold: data yielding CCFI values $> .50$ were classified as taxonic and data yielding CCFI values $< .50$ were classified as dimensional (in this study, all CCFI $\neq .50$). Additional analyses were performed to evaluate the accuracy achieved by applying a sample-optimized threshold rather than .50.

RESULTS

Number of Bootstrap Samples of Comparison Data

The first series of analyses tested for differences in the performance of the CCFI across levels of B . Across all 25,000 data sets, pairwise two-tailed z tests for differences in proportions revealed no statistically significant differences in the proportion of target data sets correctly classified as taxonic or dimensional across the three levels of B (10, 25, and 50), each $|z| = 1.54$, $p \geq .123$; accuracy rates were 92.4% at $B = 10$ and 93.0% at $B = 50$. The same was found across the 12,500 taxonic data sets, each $|z| \leq .72$, $p \geq .470$. For the 12,500 dimensional data sets, however, increasing B from 10 to 50 yielded an improvement that was statistically significant, $z = 2.92$, $p = .004$. This improvement was rather modest, 94.9% at $B = 10$ and 95.7% at $B = 50$. Considered as a whole, one might interpret these findings as evidence that performance reached a plateau by $B = 10$, or one might reason that there is nothing to lose by increasing B to 50, which helps a bit in correctly classifying dimensional data. In all subsequent analyses, only the results for $B = 50$ are used because nearly identical patterns emerged for $B = 10$ or 25.

Overall Classification Accuracy

The second series of analyses examined classification accuracy using all 25,000 data sets in a variety of ways. The difference between 90.2% accuracy for taxonic data and 95.7% accuracy for dimensional data was statistically significant, $z = 16.80$, $p < .001$. Of course, researchers are not in the position of knowing the structure of an empirical data set and asking whether their data analytic procedure correctly identified this structure. Rather, they know the results of their analysis and would like to know how much confidence to place in an inference of structure based on these results. In the present context, this corresponds to an inference of taxonic or dimensional structure given a CCFI value. Because there were more errors for taxonic than dimensional data,⁶ this means that more of the CCFI values below .50 would lead to mistaken conclusions than the CCFI values above .50. Specifically, accuracy was 95.4% for CCFI $> .50$ and 90.7% for CCFI $< .50$. When one conditions on data structures, results were more accurate for dimensional than taxonic data. When one conditions on CCFI values, results were more accurate for CCFIs $> .50$ than CCFIs $< .50$.

The accuracy rates reported previously were all calculated using the a priori threshold of .50. A sample-optimized threshold can be located at the point where the distributions of CCFI values for taxonic and dimensional data intersect; see Figure 1, left panel. This threshold, which was located at CCFI = .483, maximizes the percentage of the 25,000 target data sets correctly classified. The accuracy of the optimal threshold (93.2%) was not statistically significantly different from that of the a priori threshold (93.0%), $z = 1.14$, $p = .254$. A receiver operating characteristic analysis was conducted to examine the performance of the CCFI independent of the threshold used to classify data as taxonic or dimensional. The results are shown in the middle panel of Figure 1. The area under the curve (.980) shows that the CCFI differentiated the taxonic and dimensional data sets with an impressive degree of accuracy. One way to interpret this value is that a randomly selected taxonic data set had a 98.0% chance of yielding a higher CCFI value than a randomly chosen dimensional data set.

The right panel of Figure 1 displays results in a format that might be especially helpful for researchers. This graph plots the probability that the CCFI classified correctly within each of a series of ordered subsamples created by cutting at CCFI values .025 apart from one another. The dotted lines show that the probability of correct classification remains above .90 for the intervals centered outside an intermediate range of .40 to .60. If one were to treat the

⁶When we refer to CCFI values as being correct or incorrect, accurate or inaccurate, this is shorthand for the accuracy with which they identify latent structure as taxonic or dimensional and not a statement about the precision or bias with which sample CCFI values estimate population values.

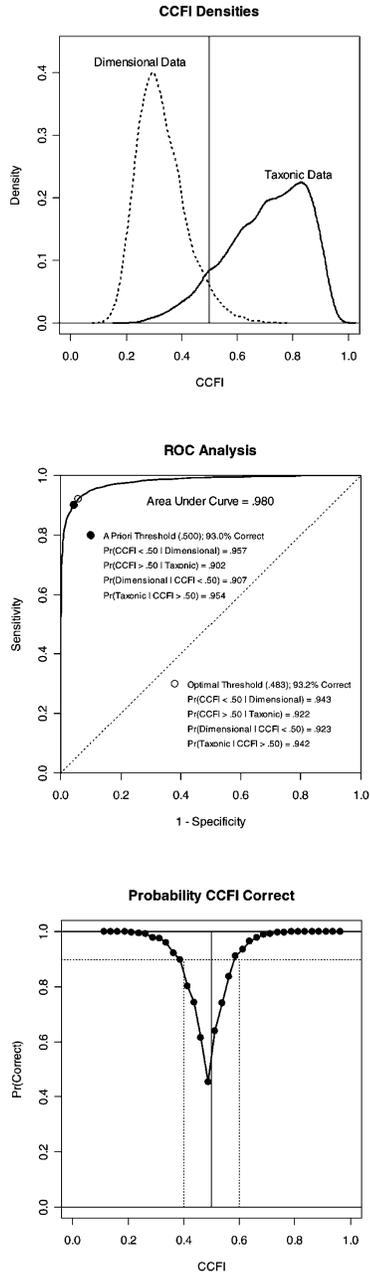


FIGURE 1 Summaries of results for 25,000 data sets with $B = 50$ bootstrap samples of comparison data of each structure. CCFI = Comparison Curve Fit Index; ROC = receiver operating characteristic. In the top and bottom panels, solid vertical lines show the a prior threshold of CCFI = .50; in the bottom panel, dashed lines (vertical and horizontal) show that the probability of correctly identifying structure remained above .90 when the CCFI was outside an intermediate range of .40 to .60. In the middle panel, the diagonal line represents chance-level classification accuracy.

19.0% of CCFIs between .40 and .60 as ambiguous and withhold judgment, the accuracy rate for the 81.0% of samples with CCFI values outside this intermediate range was 97.9%. This interpretive approach is most consistent with the rationale for generating and analyzing comparison data, which is to determine first whether the data appear capable of distinguishing between taxonic and dimensional structural models and, if so, which model provides better fit. CCFI values near .50 represent cases in which the data do not yield an informative test between these competing structural models and withholding judgment would be prudent.

Classification Accuracy for Each Design Factor

The final series of analyses examined the accuracy with which the CCFI classified data sets as a function of each design factor. For these analyses, a CCFI threshold of .50 was applied to all data sets. Performance was evaluated separately for taxonic and dimensional data, in part because different design factors are pertinent to each (i.e., P , d , r , and VR do not denote parameters in the dimensional structural model). All results are shown in Figure 2, with those for taxonic data plotted using solid lines and those for dimensional data plotted using dashed lines.

For dimensional data, accuracy fluctuated only modestly across data conditions (N , k , r_{xy} , g , and h). The poorest accuracy levels were observed for samples with few cases or few indicators; accuracy increased quickly beyond small values of N or k . Trends for the other factors were discernible but weak: accuracy remained above 92% at all levels of indicator correlation (r_{xy}), asymmetry (g), and tail weight (h). When samples with ambiguous CCFI values ($.40 \leq \text{CCFI} \leq .60$) were excluded, accuracy was above 98% for most data conditions.

For taxonic data sets, accuracy varied considerably across some data conditions. Three design factors exerted substantial influences: accuracy increased as the number of indicators (k) increased, indicator validity (d) increased, and within-group correlation (r) decreased. For the remaining five design factors—sample size (N), taxon base rate (P), asymmetry (g), tail weight (h), and variance ratio (VR)—comparatively modest effects were observed. Accuracy neither fell below 86% nor rose above 94% at any level of these factors. When ambiguous CCFI values ($.40 \leq \text{CCFI} \leq .60$) were excluded, accuracy was above 95% for most data conditions.

DISCUSSION

Ruscio et al.'s (2007) procedure for generating and analyzing simulated taxonic and dimensional comparison data to provide an interpretive aid in a taxo-

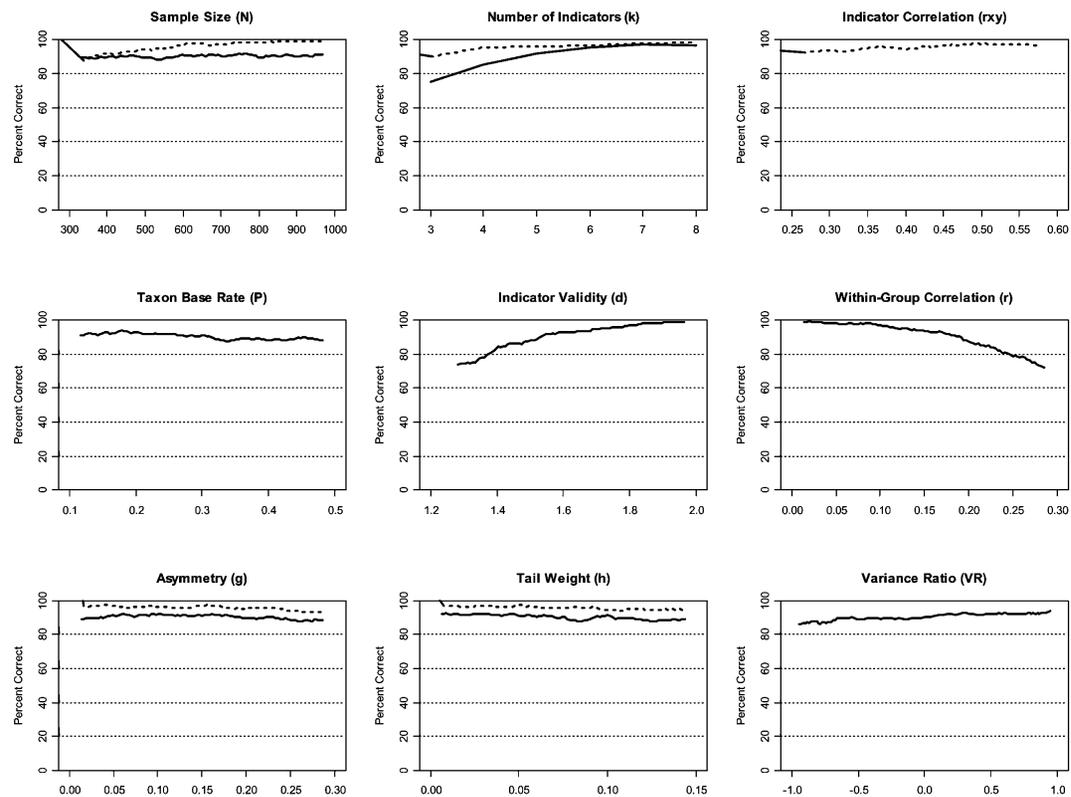


FIGURE 2 Percentage of data sets correctly classified. Solid lines show results for taxonic data and dashed lines show results for dimensional data. Each graph contains 100 data points representing subsamples of $n = 1,147$ data sets such that adjacent subsamples overlapped 90% with their neighbors; the exception is the graph for k , where nonoverlapping subsamples of approximately equal size were created for data sets with $k = 3, 4, 5, \dots, 8$ indicators. Note that the base-4 logarithm of VR values is plotted to preserve the symmetry inherent in $VRs < 1$ (larger variance in the complement) and $VRs > 1$ (larger variance in the taxon).

metric analysis eliminates the subjectivity of curve interpretation, implements a competing-models inferential framework, and achieves an impressive degree of validity in distinguishing taxonic and dimensional data. The present study capitalized on what has been achieved using comparison data and extended the investigation of the taxometric method in two ways.

First, taking advantage of improvements in the data generation algorithm provided by Ruscio and Kaczetow (2008) places the approach on a more solid statistical foundation because the updated programs introduce normal sampling error in the reproduction of key data characteristics and facilitates the routine use of multiple samples of comparison data because the updated programs run more efficiently. We found that performance improved very little when going beyond $B = 10$. These results might support an investigator's decision to use $B = 10$, but the increased efficiency of the programs to generate these data allows researchers to use larger values of B , even $B = 50$ or more, with ease.

Second, the robustness of taxometric analysis to unfavorable data conditions was examined by varying them systematically in new ways, including separate manipulations of asymmetry and tail weight as well as the taxon:complement variance ratio. Along with manipulations of many other potentially influential factors—sample size, number of indicators, taxon base rate, indicator validity, and within-group correlation—this afforded a far-reaching examination of robustness to unfavorable data conditions. For dimensional data, none of these factors substantially influenced the accuracy of conclusions reached on the basis of CCFI values; as shown in Figure 2, all effects were fairly small. For taxonic data, comparatively large effects were observed for just three factors: CCFI values were more accurate with more indicators, larger indicator validity, and smaller within-group correlations. These findings are consistent with those of Ruscio et al. (2007) and Ruscio (2007). More noteworthy, however, is that neither asymmetry, tail weight, nor variance ratios influenced the accuracy of CCFI values very much. Across the range of data conditions studied, taxometric analysis—or at least the MAXEIG procedure—appears to be robust to deviations from normal distributions and equal variances.

In an actual taxometric investigation, researchers may exert some control over the data conditions (Ruscio & Ruscio, 2004). The sampling plan—including the selection of a target population—will affect not only the sample size but also the putative taxon base rate (e.g., a mental disorder of interest might be more prevalent in a sample from a clinical than a community population). Choosing what to assess and how indicators are selected or constructed from available items will affect many other characteristics of the data, including asymmetry and tail weight of indicator distributions as well as between-group indicator validity and within-group indicator correlations. In short, the decisions made to design and execute a taxometric investigation will impact the data conditions,

and thoughtful choices are required to maximize the ability to test the relative fit of taxonic and dimensional models.

In addition to addressing our two primary goals, the accuracy of structural inferences drawn from CCFI values was examined in a variety of ways and presented in a form that should be especially helpful for researchers. For the particular implementation of the MAXEIG taxometric procedure in this study, the CCFI achieved greater accuracy for dimensional data (96%) than for taxonic data (90%). However, usually the research question is whether the data provide stronger evidence in favor of a taxonic or a dimensional structural model, in which case one needs to know the accuracy of structural inferences given CCFI values. This study found that CCFI values greater than .50 were more accurate (95%) than CCFI values less than .50 (91%). More generally, the right panel of Figure 1 plots accuracy as a function along the full continuum of CCFI values. To interpret the latter accuracy rates as posterior probabilities requires the assumption of equal priors. In the present study, the prior probabilities for taxonic and dimensional structure were equal by design: there were 12,500 target data sets of each structure. In actual taxometric investigations, we suspect that most investigators would be willing to entertain equal or approximately equal priors because to do otherwise would weaken the justification for performing a taxometric analysis. At best, a taxometric study can lead to a shift in the evidence favoring taxonic vs. dimensional structural models. If the evidence already strongly supports either model, this leaves little opportunity for a new study to have much of an impact. We expect most taxometric investigations to examine constructs for which approximately equal prior probabilities for taxonic and dimensional structure constitutes a reasonable starting point.

We recommend that researchers withhold judgment when the CCFI is close to .50. For example, to achieve predictive power of at least 90%, the present results suggest CCFI values for MAXEIG analyses must be less than about .40 or greater than about .60. For reasons discussed later, we hesitate to recommend these thresholds for the interpretation of CCFI values. Instead, we emphasize only that it is possible to obtain ambiguous taxometric results and that researchers should consider, on a case-by-case basis, whether it is advisable to draw any conclusions from their particular results.

In this study, CCFI values were calculated from the results of MAXEIG analyses of data that varied in certain ways. Other implementations of the MAXEIG procedure, or the use of different taxometric procedures, may have yielded different results. For example, we performed MAXEIG very conservatively by holding the number of overlapping windows constant at 50. In addition to performing MAXEIG differently, the CCFI can also be calculated from the results of other taxometric procedures. For example, Ruscio (2007) found that CCFI values calculated from MAMBAC (Mean Above Minus Below A Cut; Meehl & Yonce, 1994) analyses achieved greater classification accuracy than CCFI values

calculated from MAXEIG analyses of the same data sets. Future research might compare the validity of the CCFI across different implementations of different taxometric procedures and develop indices akin to the CCFI for other taxometric procedures or consistency tests. It would also be useful to go beyond our main-effects analysis to examine the performance of the CCFI across interacting design factors.

Though every simulation study includes a finite range of data conditions, we believe that our target data sets span a wide range that is broadly representative of the regions occupied by most empirical data sets submitted to taxometric analysis. Provided that one generates and analyzes taxonic and dimensional comparison data as an interpretive aid and obtains a CCFI value not too close to .50, the risk of misclassifying taxonic data as dimensional (or vice versa) appears reassuringly small. As discussed and illustrated elsewhere (e.g., Maraun & Slaney, 2005; Maraun et al., 2003; Ruscio et al., 2006), the same cannot be said of taxometric analyses performed in the absence of comparison data.

EMPIRICAL ILLUSTRATION

To demonstrate the use of comparison data in a taxometric analysis, we revisited the question that opened this article: Do people vary along a continuum of hypochondriasis, or are some people hypochondriacs and others not? Our data come from the Hathaway Data Bank, which contains approximately 34,000 sets of responses to the MMPI collected at University of Minnesota affiliated hospitals between 1940 and 1976.⁷ Following procedures described elsewhere (Ruscio & Ruscio, 2000), we screened for invalid response profiles and identified one record per patient. This left us with complete data on the 33 hypochondriasis items (Scale 1, *HS*) for 20,936 patients. Because all MMPI items are dichotomous (true/false), we formed composites by assigning the 33 items as evenly as possible to each of k summative indicators, with k ranging from 3 to 6 for separate analyses. (Every third item was assigned to each of three indicators, every fourth item to each of four indicators, etc.) MAXEIG analyses with increasing numbers of windows converged on taxon base-rate estimates between .25 and .30. We assigned the 5,792 cases with total scores of 17 or higher to the putative taxon, for a base rate of .277, and the remaining cases to the putative complement.

Using these group assignments, indicator validity and within-group correlations were estimated for each level of k . As k increased from 3 to 6, estimated indicator validity decreased from $d = 2.39$ to $d = 1.93$; as the same 33 items were spread across more indicators, this tended to weaken the validity of each

⁷We thank Leslie Yonce and the late Paul Meehl for granting access to the Hathaway Data Bank.

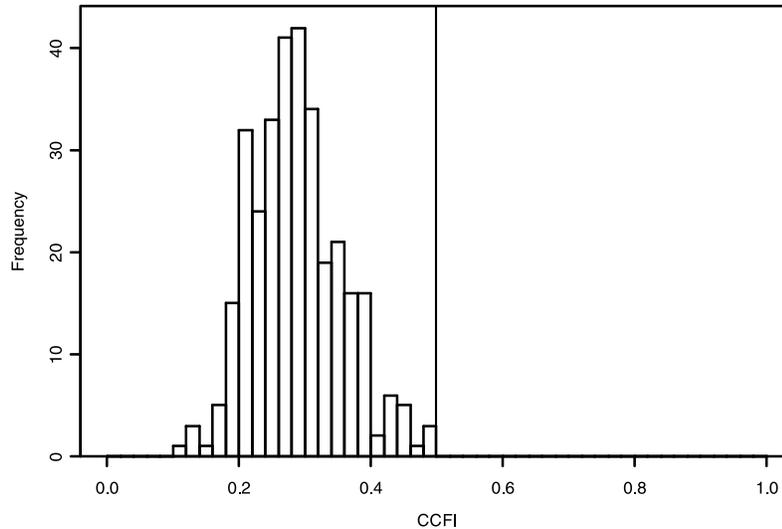


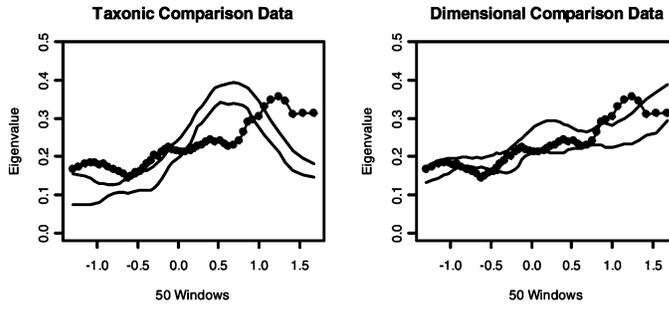
FIGURE 3 Distribution of all 320 Comparison Curve Fit Index (CCFI) values for analyses of hypochondriasis data plus illustrative maximum eigenvalue (MAXEIG) graphs for analyses of hypochondriasis data with $k = 3, 4, 5,$ and 6 indicators and $N = 650$. In the distribution of CCFIs, the vertical line at $.50$ shows the a priori threshold for distinguishing taxonic ($> .50$) and dimensional ($< .50$) data. Each MAXEIG graph shows the averaged curve for the empirical data superimposed above a band of values at ± 1 standard deviation from the mean value at each data point for parallel analyses of comparison data. (continued)

indicator. At the same time, increasing k from 3 to 6 reduced estimated within-group correlations from $r = .37$ to $r = .25$. Constructing indicators from available data is only one way in which investigators can exert control over data conditions; as demonstrated later, this choice can influence CCFI values, too.

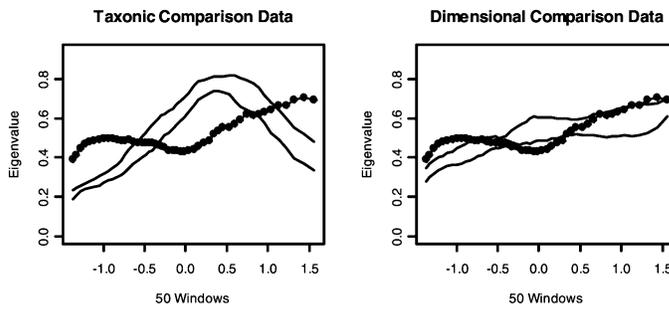
Performing MAXEIG using nearly 21,000 cases may provide clearer results than most researchers can expect, so we analyzed random samples with $N = 300, 400, 500, \dots,$ or 1,000 cases. Within each sample, we randomly assigned the 33 items as evenly as possible to $k = 3, 4, 5,$ or 6 composite indicators. For each of these $8(N) \times 4(k) = 32$ conditions, we analyzed 10 random samples of data using MAXEIG and calculated the CCFI.⁸ As shown in Figure 3, all 320 of these CCFI values were below $.50$ ($Mdn = .284, IQR = .240$ to $.331$),

⁸Whereas we randomly selected subsamples from the full empirical data set and assigned items to varying numbers of indicators to illustrate the utility of parallel analyses of comparison data, in actual studies investigators should carefully consider all issues that influence the data conditions. Ruscio et al. (2006) reviewed many pertinent concerns and offered suggestions based on the strengths and weaknesses of alternative approaches as well as the guidance offered by the extant methodological literature.

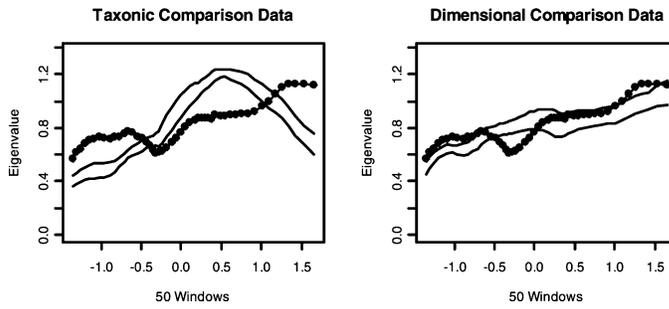
$k = 3, CCFI = .276$



$k = 4, CCFI = .279$



$k = 5, CCFI = .292$



$k = 6, CCFI = .280$

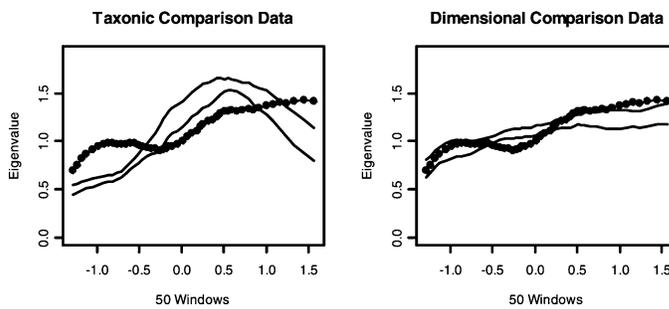


FIGURE 3 (Continued).

consistently providing evidence favoring a dimensional rather than a taxonic model. As k increased, the effect of decreases in estimated d values was more than offset by the effect of decreases in estimated r values. Analyses of the 160 samples with $k = 5$ or 6 yielded stronger CCFI values ($Mdn = .270$) than analyses of the 160 samples with $k = 3$ or 4 ($Mdn = .292$).

Figure 3 also contains illustrative MAXEIG graphs for analyses with $N = 650$ (the midpoint of the range from 300 to 1,000) and $k = 3, 4, 5,$ and 6; these representative graphs correspond to CCFIs close to the median of .284. These results conform poorly to the prototypical shapes expected for taxonic (peaked) or dimensional (flat) MAXEIG curves (Waller & Meehl, 1998). Instead, most fluctuate around a generally increasing line, which is consistent with prior findings for taxometric analyses of positively skewed data. Because the distribution of hypochondriasis total scores had a skew value of .48, most or all of the composite indicators probably were a bit positively skewed as well. This demonstrates the utility of parallel analyses of comparison data, as one need not make subjective judgments about which reasonable people might disagree (e.g., Is the graph for $k = 3$ peaked or not?). The strength of this approach lies in the fact that one need not apply a template to taxometric results. Instead, results for data of known taxonic or dimensional structure, but with comparable indicator distributions and correlations, serve as a guide to interpretation. Even if done visually, it is clear that each curve for the empirical data is more similar to that for the dimensional than the taxonic comparison data. The CCFI quantified this objectively and underscored the substantially stronger support for the model of dimensional structure.

The strong and increasing popularity of the taxometric method attests to its ability to address questions about the categorical or dimensional structure of constructs that interest many investigators. The present findings suggest that when implemented using parallel analyses of comparison data, the validity of conclusions drawn from taxometric results is robust across a wide range of data conditions that might be encountered in empirical studies.

REFERENCES

- Bartholomew, D. J. (1987). *Latent variable models and factor analysis*. New York: Oxford University Press.
- Beach, S. R. H., Amir, N., & Bau, J. J. (2005). Can sample-specific simulations help detect low base-rate taxonicity? *Psychological Assessment, 17*, 446–461.
- Beasley, W. H., DeShea, L., Toothaker, L. E., Mendoza, J. L., Bard, D. E., & Rodgers, J. L. (2007). Bootstrapping to test for nonzero population correlation coefficients using univariate sampling. *Psychological Methods, 12*, 414–433.
- Cleland, C., & Haslam, N. (1996). Robustness of taxometric analysis with skewed indicators: I. A Monte Carlo study of the MAMBAC procedure. *Psychological Reports, 79*, 243–248.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997–1003.

- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*, 1–26.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. San Francisco: Chapman & Hall.
- Frances, A. J., First, M. B., Widiger, T. A., Miele, G. M., Tilly, S. M., Davis, W. W., et al. (1991). An A to Z guide to DSM-IV conundrums. *Journal of Abnormal Psychology*, *100*, 407–412.
- Haslam, N., & Cleland, C. (1996). Robustness of taxometric analysis with skewed indicators: II. A Monte Carlo study of the MAXCOV procedure. *Psychological Reports*, *79*, 1035–1039.
- Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory* (Rev. ed.). Minneapolis: University of Minnesota Press.
- Hoaglin, D. C. (1985). Summarizing shape numerically: The *g*-and-*h* distributions. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Exploring data, tables, trends, and shapes* (pp. 461–511). New York: Wiley.
- Lee, W.-C., & Rodgers, J. L. (1998). Bootstrapping correlation coefficients using univariate and bivariate sampling. *Psychological Methods*, *3*, 91–103.
- Maraun, M. D., & Slaney, K. (2005). An analysis of Meehl's MAXCOV-HITMAX procedure for the case of continuous indicators. *Multivariate Behavioral Research*, *40*, 489–518.
- Maraun, M. D., Slaney, K., & Goddyn, L. (2003). An analysis of Meehl's MAXCOV-HITMAX procedure for the case of dichotomous indicators. *Multivariate Behavioral Research*, *38*, 81–112.
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Erlbaum.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences of kind. *Journal of Personality*, *60*, 117–174.
- Meehl, P. E. (1995a). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist*, *50*, 266–274.
- Meehl, P. E. (1995b). Extension of the MAXCOV-HITMAX taxometric procedure to situations of sizable nuisance covariance. In D. Lubinski & R. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 81–92). Palo Alto, CA: Consulting Psychologists Press.
- Meehl, P. E. (1999). Clarifications about taxometric method. *Applied and Preventive Psychology*, *8*, 165–174.
- Meehl, P. E. (2004). What's in a taxon? *Journal of Abnormal Psychology*, *113*, 39–43.
- Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*, *74*, 1059–1274.
- Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports*, *78*, 1091–1227.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166.
- Murphy, E. A. (1964). One cause? Many causes? The argument from the bimodal distribution. *Journal of Chronic Disease*, *17*, 301–324.
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*, 241–301.
- Ruscio, J. (2007). Taxometric analysis: An empirically-grounded approach to implementing the method. *Criminal Justice and Behavior*, *24*, 1588–1622.
- Ruscio, J., Haslam, N., & Ruscio, A. M. (2006). *Introduction to the taxometric method: A practical guide*. Mahwah, NJ: Erlbaum.

- Ruscio, J., & Kaczetow, W. (2008). Simulating multivariate nonnormal data using an iterative algorithm. *Multivariate Behavioral Research, 43*, 355–381.
- Ruscio, J., & Marcus, D. K. (2007). Detecting small taxa using simulated comparison data: A reanalysis of Beach, Amir, and Bau's (2005) data. *Psychological Assessment, 19*, 241–246.
- 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. (2004). A conceptual and methodological checklist for conducting a taxometric investigation. *Behavior Therapy, 35*, 403–447.
- Ruscio, J., & Ruscio, A. M. (2008). Categories and dimensions: Advancing psychological science through the study of latent structure. *Current Directions in Psychological Science, 17*, 203–207.
- 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. (2007). Applying the bootstrap to taxometric analysis: Generating empirical sampling distributions to help interpret results. *Multivariate Behavioral Research, 42*, 349–386.
- Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage.
- Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist, 62*, 71–83.