Taxometrics

John Ruscio
The College of New Jersey, U.S.A.

Taxometric analyses are performed to determine whether individuals differ from one another in a categorical or dimensional manner. Paul Meehl (1920–2003) and his colleagues developed a family of taxometric data analytic procedures that provide nonredundant evidence bearing on the latent structure of a target construct. For example, Meehl's taxometric method can be used to determine whether distinct groups of psychopaths and nonpsychopaths exist, or whether individuals differ along one or more psychopathic latent traits. The method addresses the latent structure of a target construct, not the manifest structure of a data set that depends critically on how one chooses to conceptualize and measure the construct. In many clinical, forensic, research, or administrative contexts, one might wish to classify individuals as psychopaths or nonpsychopaths. This goal can be met regardless of the latent structure of psychopathy. On the other hand, knowledge of a construct's latent structure can be advantageous for basic or applied science. Theories of psychopathy, assessment instruments, and diagnostic classification can be developed or refined using structural information. Distinguishing between categorical and dimensional structure also holds implications for the design of research and the statistical power of data analyses.

Because latent structure is important for so many reasons, the taxometric method was designed expressly to "carve nature at its joints." The fundamental goal is to establish where categorical boundaries can be drawn in a nonarbitrary manner and where individual differences are dimensional in nature. As with other latent variable modeling techniques, observed variables serve as indicators of a latent construct. A taxometric data analytic procedure examines the relationships among variables and provides evidence that can be used to assess the relative fit of two competing structural models. Individuals may differ along one or more dimensions (continua, traits, latent factors) or there may be a categorical separation between two groups (taxa, types, latent classes); in the latter case, variables may be correlated along one or more dimensions within either or both groups. In the taxometric literature, the groups are labeled the taxon (the higher-scoring group) and its complement.

Perhaps because Meehl was a clinical psychologist who developed his taxometric method to address research questions in clinical psychology, this remains the realm in which it has been applied most often. Nick Haslam and his colleagues have published a number of reviews of taxometric research. The most recent of these papers (Haslam, Holland, & Kuppens, 2012) not only documents the breadth of constructs that have been studied, the conclusions reached by the original investigators, and the ways in which the taxometric method has been implemented, but also incorporates quantitative analyses of factors related to structural conclusions. Haslam et al. (2012) identified 177 articles published in peer-reviewed journals that included taxometric analyses, and the large list of target constructs was classified into domains such as mood disorder, anxiety disorder, eating disorder, substance use, externalizing, personality disorder, normal personality, or other individual differences. Although the original authors reached a conclusion of categorical structure for 39% of the findings, Haslam et al. (2012) estimated that the true prevalence of categories was 14% after they statistically controlled for factors such as the recency and methodological strength of the studies. They concluded that "most latent variables of interest to psychiatrists

The Encyclopedia of Clinical Psychology, First Edition. Edited by Robin L. Cautin and Scott O. Lilienfeld.

DOI: 10.1002/9781118625392.wbecp176

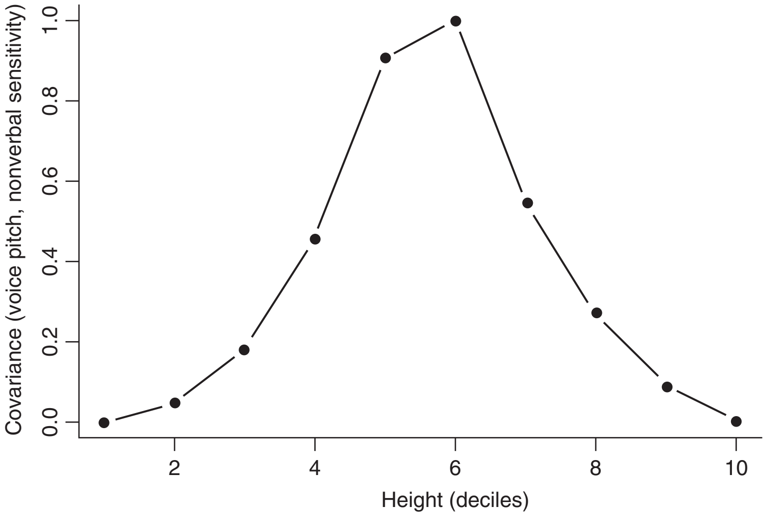

Figure 1 Illustrative MAXCOV analysis showing peaked curve for three indicator variables that distinguish men and women.

and personality and clinical psychologists are dimensional" (p. 903).

An Illustrative Procedure

The iconic taxometric procedure is MAXCOV (*max*imum *cov*ariance), which requires three indicator variables. Cases are sorted along one variable, which forms the *x* axis of the graph on which results will be plotted. Within a subsample of cases with the lowest scores on this variable (e.g., the lowest decile), the covariance between the two remaining variables is calculated. This covariance is plotted along the *y* axis of the graph. Next, a subsample of cases with slightly higher scores is formed (e.g., the second-lowest decile), and a new covariance value is calculated and plotted. This procedure is repeated until the subsample includes cases with the highest scores (e.g., the highest decile). The shape of the resulting MAXCOV curve provides a clue to the latent structure of the target construct. Specifically, the curve is expected to contain a distinct peak for categorical structure, but not for dimensional structure.

For example, consider the analysis of the three variables height, voice pitch, and nonverbal sensitivity in a sample that contains equal numbers of men and women. Despite considerable overlap on all three variables, there are systematic group differences: Men tend to be taller and women tend to score higher on the other two variables. Sorting cases by height and calculating the covariance between voice pitch and nonverbal sensitivity within each decile should yield a peaked MAXCOV curve (see Figure 1). The lowest deciles will contain mostly women, among whom there is no association between voice pitch and nonverbal sensitivity. Likewise, the highest deciles will contain mostly men, among whom there is no association between these variables. Within the middle decile, corresponding to the average height in the sample, there will be a mixture of men and women and a positive association between voice pitch and nonverbal sensitivity. The positive association occurs because women tend to score higher on both variables. A scatterplot would reveal two clouds of points, one toward the lower-left representing men's paired scores and one toward the upper-right representing women's paired scores, and therefore the covariance would be positive. In each decile near the middle, the covariance will still be positive. Its magnitude is a function of how evenly mixed the groups are, with the maximum covariance occurring for the subsample that contains the most nearly equal numbers of men and women. It is this mixture of groups that differ systematically on the indicator variables that gives rise to a pattern of results indicative of categorical structure. One need not have an infallible classification of cases into groups (i.e., men and women) to identify the fact that a categorical structure underlies these data. In contrast, in the absence of a categorical distinction there is no reason to expect a peak in the MAXCOV curve.

Consistency Testing

Many other quantitative methods exist to differentiate data with categorical and dimensional latent structures. What makes the taxometric method unique is its emphasis on consistency testing. Rather than performing a single analysis and drawing a conclusion based on a test of statistical significance or an index of absolute or relative model fit, taxometric analysis requires that multiple nonredundant sources of evidence be obtained. To the extent that the evidence consistently supports a categorical or a dimensional structural model, one can draw a conclusion with a degree of confidence corresponding to the strength of the evidence. In some cases, the data will afford a powerful test and the results will clearly differentiate between these two competing structural models. In other cases, the results will not be so clear, perhaps even highly ambiguous. As with any data analytic technique, conclusive results are not guaranteed. Although ambiguous results may be disappointing to the investigators, the fact that taxometric methodology itself provides an indication of when strong conclusions are unwarranted is an invaluable asset.

A handful of taxometric procedures have emerged as the most popular tools for obtaining evidence whose consistency can be checked. MAXCOV has been used most often. One variant of MAXCOV that can be performed using only two variables is MAXSLOPE (*max*imum *slope*), and another that can be performed with three or more variables is the multivariate generalization MAXEIG (*max*imum *eig*envalue). Because MAXCOV, MAXSLOPE, and MAXEIG operate in very similar ways, they would not be expected to yield nonredundant evidence suitable for consistency testing.

Two other taxometric procedures are more distinct and have often served as consistency tests in studies that include MAXCOV, MAXSLOPE, or MAXEIG. The MAMBAC (*m*ean *a*bove *m*inus *b*elow *a* *c*ut) procedure requires two variables, one of which is used to sort cases and serves as the *x* axis of a graph as in MAXCOV. A cutting score is located near the lower end on this variable, and the mean difference on the second variable is calculated for subsamples scoring above and below the cut. This mean difference is plotted along the *y* axis of a MAMBAC graph. The cutting score is then moved upward to form a new pair of subsamples and then calculate and plot their mean difference. This procedure is repeated until the cutting sore has moved near the upper end of the first variable's range. If the data are categorical, a peak in the MAMBAC curve is expected to emerge near the cutting score that best distinguishes the two groups from one another. For dimensional data, no peak is expected. The L-Mode (latent mode) procedure requires at least three variables. A factor analysis of all variables is performed, and scores are estimated for the first principal factor. For categorical data a bimodal distribution of factor scores is expected, whereas for dimensional data a unimodal distribution is expected.

Some of these procedures (e.g., MAXCOV and MAMBAC) can be run multiple times to obtain a panel of curves. For example, with three variables—*X*, *Y*, and *Z*—MAXCOV can be performed three times (once using each variable on the *x* axis of a new graph) and MAMBAC can be performed six times (each of the three variable pairs *XY*, *XZ*, and *YZ* is used twice, once with each variable on the *x* axis of a new graph). With even more variables, considerably more curves can be generated and examined for consistency. In addition to providing curves whose shapes can be inspected for clues to the target construct's latent structure, each curve can be used to calculate an estimate of the relative size of the higher-scoring group within the sample, or what is known as the taxon base rate. These estimates can be calculated even if the data are dimensional, but they are not expected to cohere around a single value unless there is in fact a taxon whose relative size is being estimated. In other words, it has been proposed that one can examine the variability of the taxon base rate estimates as a consistency test, with smaller values more indicative of categorical structure. Other parameters of a categorical structural model (e.g., the validity with which each variable distinguishes two groups) can also be estimated and used for consistency testing.

Although many consistency tests have been proposed, considerable research remains to be done to determine which are worth including in a taxometric analysis. Ruscio, Walters, Marcus, and Kaczetow (2010) outlined criteria that a candidate consistency test should meet to warrant inclusion. The candidate test should differentiate categorical and dimensional data. Regardless of how plausible the rationale, a test should be demonstrably valid. If it meets this criterion, the candidate test should provide incremental validity. The test should provide evidence that is not redundant with other results already being used. If it meets these criteria, there should be an objective decision rule for reaching conclusions on the basis of the candidate test. This prevents confirmation bias from compromising the utility of a test that requires a subjective judgment. Relatively few of the many proposed consistency tests have been examined rigorously to determine whether they meet any, let alone all, of these criteria. One approach to consistency testing that appears to meet these criteria will be discussed in a later section.

Inferential Framework

Beginning with many of the seminal research reports dating back to the 1960s (e.g., Meehl, 1965), there has been some ambiguity regarding the inferential framework for taxometrics. Historically, it was not uncommon for authors to describe the purpose of taxometric analysis as the detection of a taxon. Under a taxon-detection framework, if nonredundant sources of evidence consistently supported the existence of a taxon, one would reach a categorical conclusion. On the other hand, the failure to detect a taxon is analogous to the failure to reject a null hypothesis, and no firm conclusion can be drawn. From this perspective, dimensional structure is analogous to a null hypothesis that one cannot accept.

A conceptual problem with the taxon-detection framework is that if one adopts it the taxometric method is no longer a tool for distinguishing categorical and dimensional data because one can seldom, if ever, reach the latter conclusion with much confidence (as one can seldom, if ever, accept a null hypothesis with much confidence). Even those authors who appear to adopt the taxon-detection framework have not always been consistent in adhering to this perspective. At best, there is mixed evidence in the taxometric literature suggesting that authors implicitly endorse the taxon-detection framework; there is little explicit support for it.

When the results of taxometric procedures are interpreted in the conventional fashion described above (e.g., inspecting a MAXCOV curve for the presence of a peak indicative of categorical structure), it is in fact difficult to differentiate between results supportive of dimensional structure and merely ambiguous results that afford no firm conclusion. For example, even for a truly categorical construct the available data may be insufficiently valid for peaks to emerge in MAXCOV curves. This outcome is analogous to low statistical power in conventional significance testing. Until recently there was no compelling alternative to this inferential framework.

The introduction of methods involving the generation and parallel analysis of categorical and dimensional comparison data provided an alternative; these methods are described below. The careful use of comparison data affords a fruitful interpretive aid for taxometric results and, more importantly, enables the adoption of an inferential framework in which two competing structural models are put to a test. Under this competing-models framework, the goal is to determine whether the evidence provides better support for a categorical or a dimensional structural model. Thus, unlike under the taxon-detection framework, one can obtain evidence that supports a dimensional conclusion. Not only is the distinction between dimensional structure and ambiguous evidence formally acknowledged, but also the methodology affords a basis for differentiating between these possibilities. Since the use of comparison data was introduced, most taxometric reports have implicitly adopted the competing-models framework and many reports have done so explicitly.

Interpreting Taxometric Results

Taxometric procedures provide a wealth of output, including curves whose shapes can be inspected and parameter estimates whose coherence can be evaluated. Meehl's emphasis on consistency testing continues as a bedrock principle of taxometrics. However, the interpretation of taxometric results remained a fairly subjective task for a long time. Conventionally, findings were interpreted relative to the prototypical results expected for categorical and dimensional data, which were presented when a procedure was introduced into the literature. For example, one examined each MAXCOV curve and judged whether or not it appeared to be peaked. If all or most of the MAXCOV curves were judged to be peaked, this evidence was interpreted as supporting a categorical structural model.

From the beginning, the influences of some complicating factors were recognized. For example, the location of a peak within a MAXCOV curve depends on the taxon base rate in the sample of data. When the taxon and its complement are of approximately equal size, the peak is expected to emerge near the center of the curve. If the taxon is smaller than its complement, however, the peak is expected to shift toward the right along the curve. The reason is that one must reach higher-scoring subsamples of data to find the one in which the groups are mixed in approximately equal proportions. In a sample that contains 10% men and 90% women, the maximum covariance between voice pitch and nonverbal sensitivity will not occur in the middle decile on height, which will still contain far more women than men. The peak may occur in the ninth or even the tenth decile, wherever the number of men most closely matches the number of women within that subsample. Thus, it was understood that a peaked MAXCOV curve need not be symmetric. A peak could take the form of a cusp at the upper end of the curve, where a cusp refers to a maximum value that is approached through rising values on one side but not followed by falling values on the other because it occurs at the end of a curve.

The taxometric literature is replete with discussions of ways that various data conditions can influence the shape of curves for each data analytic procedure. Two particular monographs catalog and illustrate such influences especially extensively. Meehl and Yonce (1994, 1996) show prototypical MAMBAC and MAXCOV curves across a number of data conditions. They demonstrate not only the extent of random fluctuations that one should expect due to sampling error with several different sample sizes, but also the systematic variations in curve shapes produced by changing the levels of parameters such as the taxon base rate, within-group correlations, and between-group validities. Publications such as these provided a great service to investigators attempting to interpret their own taxometric results.

Even with these guides available, two substantial interpretational challenges remained. First, there was not always strong agreement on how well a particular empirical result matched any of the prototypes. Reaching a judgment still contained a considerable amount of subjectivity. In part, this challenge can be attributed to the fact that prototypes cannot be produced for all possible data conditions. Novel configurations would be expected with each new empirical application of the taxometric method. To the extent that an investigator understood the ways that data conditions are expected to influence taxometric results, he or she could reach an informed judgment about how to interpret empirical results that do not match any of the prototypes appearing in the literature.

A second challenge is that the prototypes were based on idealized data conditions that may not be approximated well in many empirical samples, particularly those studied by clinical psychologists. For example, the data used to generate prototypical taxometric results for categorical data were normally distributed within groups, and groups possessed equal variances. These are common parametric assumptions, and an eminently reasonable starting point for developing an understanding of taxometric results across data conditions. Because empirical data often deviate substantially from normality and/or equal variances (Micceri, 1989; Ruscio & Roche, 2012), though, one might expect empirical results to deviate from the prototypical taxometric curves based on idealized data.

One particularly vexing problem involves the analysis of skewed data. Asymmetric score distributions on important variables are common in clinical research. On the one hand, even if data are distributed normally and with equal variances across two groups, the full-sample score distribution would be skewed if the groups are of unequal size. If psychopathy were categorical, for example, scores for a small number of psychopaths would elongate the upper tail of an indicator variable's distribution when combined with the scores for a large number of nonpsychopaths. On the other hand, there is no reason that data need be normally distributed even in the absence of group mixture. As Micceri (1989) documented, even measures expressly designed to yield normal score distributions in specific populations often provide skewed data. The results for categorical data with a small taxon can be very difficult to distinguish from those for dimensional data with skewed indicator variables (Ruscio, Ruscio, & Keane, 2004). In both of these data conditions, for example, MAXCOV curves can rise to an upper-end cusp. A small taxon shifts the peak toward the upper end for reasons described above. Skewed indicator variables can have a similar effect through a differential restriction of range across subsamples. As one moves from lower-scoring to higher-scoring subsamples, the range of values increases for the two variables whose covariance is being calculated. The end result can be a rise in covariances that mimics the expectation for categorical data with a low taxon base rate. Such a cusped curve shape is therefore interpretationally ambiguous. Because these data conditions occur fairly frequently in clinical research, it could be challenging to justify a conclusion by making reference to prototypical taxometric results obtained under idealized data conditions.

Comparison Data

As computing power increased, a solution to these interpretational challenges became possible that would not have been feasible when the methodology was developed originally. Rather than relying on an extensive, but still limited, set of prototypes, one can obtain reference curves for categorical and dimensional data tailored to the data conditions in one's empirical sample. The keys to this approach are to generate both categorical and dimensional data sets that share the important features of a unique sample of empirical data and to submit these comparison data to the same taxometric analyses as the empirical data. There was strong precedent for this approach in more mainstream statistical applications, namely the resampling strategies by which empirical data are treated as an unbiased sample from the relevant population and the parallel analysis technique used to determine the number of factors to retain in an exploratory factor analysis.

Consider first that the usual parametric assumptions of normality and equal variances were introduced when testing statistical significance of mean differences across groups to simplify an otherwise overwhelming calculation problem (Rodgers, 1999). Sophisticated tests that did not require these assumptions were available, but it was not feasible to do the necessary resampling of observations and repetitive calculations for all but the smallest of data sets. Deriving tables of critical values for different degrees of freedom, one-tailed vs. two-tailed tests, and commonly used α levels made it feasible for investigators to perform statistical tests for a much wider range of applications. Many decades later, advances in computing power made it possible to do the resampling and calculation required for a cleaner test rather than imposing parametric assumptions.

Ruscio, Ruscio, and Meron (2007) adapted a common resampling technique known as the bootstrap for use in taxometric analysis. They introduced a method for generating artificial comparison data that reproduced important characteristics of an empirical sample, including its sample size, the number of variables, and these variables' marginal distributions and correlations. The iterative algorithm that carries out this work is grounded in standard bootstrap methods. It can be applied to a full sample of empirical data to reproduce its characteristics in an artificial data set with purely dimensional structure. It can also be applied to subsamples of the empirical data provisionally assigned to groups in order to reproduce these data characteristics within groups. By merging the artificial data generated for these subsamples, an artificial data set with categorical structure is obtained. Rather than relying on prototypical taxometric results based on idealized data, one could generate both categorical and dimensional comparison data that approximate the important characteristics of a unique sample of empirical data.

Ruscio and Kaczetow (2008) later improved the programming efficiency of the data-generation algorithm and showed how it could be used to create very large, but still finite, populations of artificial comparison data from which random samples could be drawn for analysis. By submitting each of these to the same taxometric analyses as the empirical data, the results would provide a helpful interpretive aid. This method is analogous to submitting artificial data to parallel analysis to determine how many factors to retain in an exploratory factor analysis. Results for the empirical data can be compared to those for the comparison data rather than applying a more idealized, hypothetical standard. In the case of taxometrics, one can determine whether the empirical results are more similar to those for categorical or dimensional comparison data. It is also possible to identify results that are interpretationally ambiguous. The results for the comparison data may not differ across structures, which suggests that the empirical data are not up to the task of differentiating between these structural models when analyzed in this way, or the empirical results may resemble—or fail to resemble—those of both kinds of comparison data to a similar extent. In the case of ambiguous results, no conclusion is warranted. The ability to differentiate between results that support a dimensional structural model and ambiguous results enables the adoption of the competing-models inferential framework.

In addition to providing an aid to the subjective interpretation of taxometric results, the generation and parallel analysis of comparison data affords an opportunity to quantify the fit of empirical results to those obtained for each structural model. Ruscio et al. (2007) introduced the comparison curve fit index (CCFI), which ranges from 0 (strongest possible support for dimensional structure) through .50 (ambiguous results) to 1 (strongest possible support for categorical structure). One way to use the CCFI would be to interpret any value below .50 as supportive of dimensional structure and any value above .50 as supportive of categorical structure. This single-threshold standard, however, does not recognize any results as ambiguous. Another way to use the CCFI would make reference to two thresholds. For example, any value below .40 would be interpreted as supportive of dimensional structure, any value above .60 as supportive of categorical structure, and any value from .40 to .60 as ambiguous. These thresholds can be located closer to .50 (e.g., .45 and .55) if one is willing to draw some conclusions on the basis of weaker evidence. The choice of how far from .50 a CCFI value must fall to warrant a structural conclusion is analogous to choosing an α level when testing a null hypothesis; there are two types of error one might make, and the threshold one adopts apportions the likelihood of making each type of error. Whether one prefers a single-threshold or dual-thresholds standard, either is consistent with the adoption of the competing-models framework in that it places categorical and dimensional models on equal footing and examines the strength with which one model is supported relative to the other.

The CCFI has been helpful in at least two ways. First, because it is an objective index, it constrains observer subjectivity. The CCFI has never been recommended as a replacement for an investigator's informed judgments, but it can be a useful supplement. If the CCFI and a subjective judgment disagree, it becomes important to reconcile them in a persuasive manner. Second, because the calculation of the CCFI can be automated, it paved the way for extensive, rigorous simulations studies of taxometric methodology. By removing the requirement that taxometric experts subjectively interpret all of the results, simulation studies could produce voluminous output for a wide range of data conditions and procedural implementations. Simulation studies using the CCFI have already helped to establish the data conditions under which taxometric analysis is more (and less) likely to yield informative results as well as the most appropriate ways to implement each taxometric procedure (Ruscio, Ruscio, & Carney, 2011).

The CCFI also forms the basis for the first operationalization of consistency testing in taxometric research that meets the criteria described above. Like a number of prior studies, Ruscio et al. (2010) found that CCFI values calculated from MAXCOV, MAMBAC, and L-Mode results validly differentiated categorical and dimensional data. Each of these procedures' CCFI values also provided incremental validity. There are a number of ways that one can reach a conclusion on the basis of three CCFI values, and the choice among these options involves a trade-off between the accuracy of the conclusions and the proportion of instances in which results are classified as ambiguous. A good approach would yield relatively high accuracy with relatively few results classified as ambiguous. On the basis of their empirical examination of the trade-offs, Ruscio et al. (2010) recommended that the mean of the three procedures' CCFI values be calculated and then this mean CCFI be interpreted using dual thresholds of .45 and .55. For the 100,000 categorical and dimensional samples that varied across a wide range of data conditions in their study, this operationalization of consistency testing attained an accuracy of 99.4% after setting aside only 5.2% of the results as ambiguous.

SEE ALSO: Approaches to Diagnostic Validity; Construct Validity; Dimensional versus Categorical Models of Psychopathology; Distinguishing Categories from Continua; Meehl, Paul E. (1920–2003); Null Hypothesis Significance-Testing Debate

References

Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychological Medicine, 42,* 903–920.

Meehl, P. E. (1965). *Detecting latent clinical taxa by fallible quantitative indicators lacking an accepted criterion* (Report no. PR-65-2). Minneapolis, MN: Department of Psychiatry, University of Minnesota.

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.

Rodgers, J. L. (1999). The bootstrap, the jackknife, and the randomization test: A sampling taxonomy. *Multivariate Behavioral Research, 34,* 441 – 456.

Ruscio, J., & Kaczetow, W. (2008). Simulating multivariate nonnormal data using an iterative algorithm. *Multivariate Behavioral Research, 43,* 355 – 381.

Ruscio, J., & Roche, B. (2012). Variance heterogeneity in published psychological research: A review and a new index. *Methodology, 8,* 1 – 11.

Ruscio, J., Ruscio, A. M., & Carney, L. M. (2011). Performing taxometric analysis to distinguish categorical and dimensional variables. *Journal of Experimental Psychopathology, 2,* 170 – 196.

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.

Ruscio, J., Walters, G. D., Marcus, D. K., & Kaczetow, W. (2010). Comparing the relative fit of categorical and dimensional latent variable models using consistency tests. *Psychological Assessment, 22,* 5 – 21.

Further Reading

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

Waller, N. G. (2006). Carving nature at its joints: Paul Meehl's development of taxometrics. *Journal of Abnormal Psychology, 115,* 210 – 215.

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