TAXOMETRIC ANALYSIS

An Empirically Grounded Approach to Implementing the Method

JOHN RUSCIO
The College of New Jersey

Whether individual differences are treated as categorical or continuous has consequences for theory, assessment, classification, and research in criminal justice. Paul Meehl's (1995) taxometric method allows investigators to test between these two competing structural models. This article provides an overview of the method's inferential framework and data-analytic procedures. Because guidelines for implementing taxometric analyses and interpreting their results have received little research attention, investigators are encouraged to adopt an empirically grounded approach to taxometric analysis rather than following conventions or relying on personal opinion. The guidance afforded by Monte Carlo studies, including the two reported here, can be supplemented by simulating comparison data. This empirically grounded approach, described and illustrated below, helps to implement the taxometric method effectively and to draw valid conclusions.

Keywords: taxometric analysis; consistency tests; simulated comparison data

There are many disputes in basic and applied science about the drawing of boundaries. In the domain of criminal justice, for example, considerable controversy surrounds the structure of psychopathy. Should we classify people into groups? The orderliness and simplicity of typologies (e.g., psychopath versus nonpsychopath) can be appealing, but some qualitative distinctions are more subjective than objective and may obscure or ignore important individual differences. Should we differentiate people along one or more continua (e.g., psychopathic traits such as impulsiveness, social deviance, remorselessness)? The comprehensiveness and fidelity of multidimensional schemes can be appealing, but they can become needlessly complex and may require thresholds for creating groups to facilitate communication or make decisions. Modeling qualitative and quantitative individual differences in reliable, valid, and practically useful ways poses many challenges. For many—but not all—purposes, it can be helpful to adopt the perspective of scientific realism and approach boundary issues armed with knowledge about the underlying structure of individual differences (Meehl, 1992). When distinct groups exist, a typology may be most appropriate, whereas when they do not exist, a dimensional framework may be most appropriate.

OVERVIEW OF MEEHL'S TAXOMETRIC METHOD

Several data-analytic techniques exist for researchers to test competing structural models of individual differences. This article will focus on the taxometric method developed by

AUTHOR'S NOTE: *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.*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 34 No. 12, December 2007 1588-1622
DOI: 10.1177/0093854807307027

Paul Meehl and his colleagues (Meehl, 1995; Meehl & Golden, 1982; Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998), which has been applied in several studies of psychopathy (Edens, Marcus, Lilienfeld, & Poythress, 2006; Guay, Ruscio, Hare, & Knight, in press; Harris, Rice, & Quinsey, 1994; Marcus, John, & Edens, 2004; Skilling, Quinsey, & Craig, 2001). To use this method, one analyzes observed variables for consistent evidence in support of either a taxonic or dimensional model of a construct's latent structure. This description involves four important aspects of the method.

First, individual differences are characterized as categorical (*taxonic*) or continuous (*dimensional*). Taxonic structure means that boundaries can be drawn between groups without recourse to arbitrary distinctions. If instead individuals vary along one or more continua, this corresponds to a dimensional structure. For example, psychopaths might differ qualitatively from nonpsychopaths; psychopathy may be taxonic. Alternatively, individuals might differ from one another along one or more continuous traits; psychopathy may be dimensional. It is possible that the structure of psychopathy may be more complex, containing both taxonic and dimensional features: Within a nonpsychopathic group, a set of dimensions may capture individual differences on psychopathy-relevant traits, but hardcore psychopaths deviate substantially on multiple dimensions and form their own group. In the present article, I will restrict attention to the distinction between taxonic and dimensional structures; see Ruscio, Haslam, and Ruscio (2006) or Ruscio and Ruscio (2004a) for discussions of more complex structures.

Second, the taxometric method examines a construct's latent, rather than manifest, structure. One can choose to measure a variable using either categories or continua, and this influences its manifest (apparent, measured, observable) structure. For example, using a nominal scale (e.g., psychopath versus nonpsychopath) to assess psychopathy will yield data that are taxonic at the manifest level, whereas using a multidimensional scale (e.g., an instrument that assesses multiple facets of psychopathy by summing responses to items pertinent to each facet) will yield data that are dimensional at the manifest level. At the same time, there exists a psychopathy construct whose latent structure is most appropriately modeled as taxonic or dimensional regardless of how it is assessed. How best to model this structure poses an empirical question that can be addressed through analyses of observed variables, which are referred to as *indicators* in the taxometric literature. An analogy may be helpful at this point. In a factor analysis, a covariance or correlation matrix is calculated from observed variables and analyzed to infer the structure of the construct hypothesized to cause the variables' shared variance. Often, factor analysis is used to study the number of latent factors. In a taxometric analysis, relationships among observed variables are analyzed to infer something different about the construct's latent structure—namely, whether it is better modeled as taxonic or dimensional. Meehl (2004) has discussed similarities in the use of the term *latent* in taxometric analysis and more familiar data-analytic techniques, and Meehl (2001) noted that entities in a substantive theory might correspond to latent variables at distinct levels of analysis (e.g., successive steps along a causal pathway that begins with genetic influences and culminates in behavioral outcomes).

Third, the objective of a taxometric analysis is intentionally narrow: determining whether a construct is more appropriately modeled as taxonic or dimensional. The taxometric method consists of tools for assessing the relative empirical support for a taxonic structural model (two latent groups, usually referred to as *taxon* and *complement*, either or both of which may contain dimensional variation) versus a dimensional structural model (one or more latent dimensions). This enables researchers to test focused questions, such as "Do psychopaths represent a qualitatively distinct group of people from nonpsychopaths, or is psychopathy a (multi)dimensional construct?"

It may appear a significant limitation that the taxometric method cannot address as broad a research question as can a factor analysis or a cluster analysis (e.g., how many factors, how many clusters), but requiring researchers to test more focused questions holds some conceptual and empirical advantages (see Ruscio et al., 2006; Ruscio & Ruscio, 2004a, 2004b). A construct may be modeled more accurately when researchers train their attention on individual features of its structure in a series of analyses rather than attempting to model the full complexity in a single analysis. For example, each potential boundary that might be drawn between groups can be tested using a set of indicators carefully chosen to make that particular structural distinction. The potential problem with using a single set of indicators to test multiple boundaries simultaneously is that for any given boundary, only a subset of the indicators may be relevant, and the inclusion of irrelevant variables can yield misleading results (Ruscio & Ruscio, 2004b). For present purposes, the point is to know what the taxometric method is—and is not—designed to accomplish: to determine whether a set of indicators represents individual differences of a categorical or continuous nature. For discussions of how the taxometric method and other latent variable modeling techniques can be used complementarily, see Ruscio and Ruscio (2004a).

Fourth, when performing a taxometric analysis, one seeks results that consistently support an inference of taxonic or dimensional latent structure, and the method contains many tools for checking the consistency of results. An array of data-analytic techniques operates in different ways to provide multiple lines of evidence; the emphasis on consistency checks is a crucial cornerstone of the taxometric method.

GUIDELINES FOR IMPLEMENTATION AND INTERPRETATION

Noting the importance of consistency checks is standard practice in the literature on Meehl's taxometric method. Less attention has been paid to two challenging methodological issues: implementing the data analyses and interpreting their results. Investigators must make many decisions to implement taxometric data-analytic techniques. Virtually no research has examined the many available options or the data conditions under which one choice should be preferred over another. Implementation decisions appear to be guided by conventions and personal preferences, few of which can be justified by citing compelling evidential support. Guidelines for interpreting the results of taxometric analyses may be better developed, but the evidence supporting them remains modest at best.

Most often, guidelines published in the taxometric literature involve rational or intuitive arguments that constitute a loosely theory-based approach. In the absence of rigorous mathematical proof or empirical testing, however, even a plausible argument constitutes a weak form of evidence. In published studies, taxometric procedures have been implemented in many ways, with little attention given to justifying a particular technique when alternatives exist. Likewise, researchers use different standards for interpreting taxometric results without providing a compelling rationale for their preferences. This is clearly an unsatisfactory state of affairs that signals a need for evidence-based guidelines.

One might develop guidelines analytically, via rigorous mathematical proofs, but this approach has not been pursued in earnest. At present, only Maraun and colleagues (Maraun & Slaney, 2005; Maraun, Slaney, & Goddyn, 2003) have used the analytic approach. These researchers challenged some conventional wisdom about the interpretation of taxometric results: They showed that for taxonic data, the results of one widely used taxometric procedure

(MAXCOV, which is described later) do not necessarily conform to expectations. This underscores a major weakness in the theory-based approach—that an apparently plausible argument may be mistaken, at least under certain conditions. As a remedy, Maraun and colleagues called for increased attention to analytic proofs in the taxometric literature.

Alternatively, guidelines for implementation and interpretation could be developed through an empirically grounded approach in a number of ways. Small-scale demonstrations are the simplest way to begin, and analyses of a handful of data sets (real or simulated) can suggest tentative guidelines. Demonstrations can buttress plausible arguments or counter strident but unfounded criticism, because even a single counterexample rebuts an unwarranted generalization. The line separating an empirical demonstration from a Monte Carlo study may be a fuzzy one, but the latter usually includes far more analyses, performed across a wider range of conditions. Even when analytic proof is available, Monte Carlo studies can test assumptions and simplifying approximations. For example, it is not uncommon to set aside the influence of sampling error in analytic derivations, and a Monte Carlo study can test a procedure's performance at realistic sample sizes.

In an ideal world, Monte Carlo research would include all relevant factors and vary each across all realistic values in a fully crossed design. Because this is seldom, if ever, feasible, one can adopt an approach that combines the rigor of a Monte Carlo study with the focus of a demonstration. It is possible to tailor simulations to a particular research situation to address specific questions. This supplements the findings of Monte Carlo studies. Ruscio, Ruscio, and Meron (2007) have presented methods for simulating taxonic and dimensional comparison data sets that reproduce characteristics of one's unique sample of data. Analyzing such comparison data allows an investigator to evaluate empirically the consequences of implementing a taxometric analysis in different ways as well as to obtain results for data known to be taxonic and dimensional, which provides an invaluable interpretive aid. Monte Carlo studies produce a skeletal outline of guidelines for taxometric analysis, and investigators can put flesh on these bones by using simulated comparison data. These complementary techniques constitute the empirically grounded approach to implementing the taxometric method identified in the subtitle of this article.

With this foundation in place, the taxometric method will now be introduced. In the next section, the inferential framework for the method is discussed. Then sections describing taxometric procedures and consistency tests follow. Standard techniques and their rationales will be presented, but they will also be examined. Suggestions will be offered for how to address the plethora of unresolved issues in an empirical manner. Two Monte Carlo studies will be presented both to shed light on some fundamental methodological issues and to illustrate ways that researchers can advance understanding of the taxometric method. To cover such a considerable amount of ground in a single article necessitates an emphasis on breadth rather than depth. Readers interested in learning more about issues raised in this article—as well as important topics that are not addressed because of space constraints, such as the data requirements of taxometric analysis—are referred to Ruscio et al. (2006) for a more thorough treatment.

INFERENTIAL FRAMEWORK FOR TAXOMETRICS

An inferential framework specifies the type(s) of research hypothesis that a particular data-analytic technique can test and, therefore, the nature of the conclusions that can, in

principle, be drawn using this methodology. The inferential framework for taxometrics remains a source of disagreement. The merits and shortcomings of several candidate frameworks are discussed in this section, and readers are encouraged to consider thoughtfully the position that they wish to adopt and defend on logical grounds.

INFERENTIAL FRAMEWORK 1: DETECTION OF TAXONIC STRUCTURE

The inferential framework implicit throughout much of the literature on the taxometric method (e.g., Meehl, 1995; Meehl & Golden, 1982; Waller & Meehl, 1998) involves the detection of taxonic structure. One hypothesizes that a taxon can be distinguished from its complement and performs a series of taxometric analyses to assess the empirical support for this hypothesis. This requires knowledge of the results expected for taxonic data, and researchers are urged to reach a taxonic conclusion when results are consistent with these expectations. Often—but not always—the results expected for dimensional data are considered as well. This calls attention to features believed to be uniquely indicative of taxonic structure. The greatest strength of this inferential framework is its emphasis on the distinction between results expected for taxonic and dimensional data, which encourages a comparative rather than an absolute judgment. There are a few potential weaknesses to consider.

First, the question of when results are sufficiently consistent with the expectations for taxonic data to infer taxonic structure has not been addressed adequately. Evaluating results in terms of their consistency with expectations provides poor protection against the operation of confirmation bias (Nickerson, 1998). An investigator who entertains little doubt that a construct is (or is not) taxonic might present apparently supportive taxometric results through selective retention, presentation, and interpretation. Even if the pressure to confirm an initial hypothesis operates outside of awareness, there is a discomfiting possibility that confirmation bias can influence conclusions.

Second, dimensional data can yield taxometric results that mimic those expected for taxonic data. For example, early work suggested that some taxometric procedures should yield peaked or cusped graphs for taxonic data (e.g., Meehl & Yonce, 1994, 1996), in which a peak features lower values on both sides and a cusp results from an increase in values all the way to the end of a graph. It was shown more recently that when indicators are skewed, dimensional data often yield cusped curves (e.g., Ruscio, Ruscio, & Keane, 2004). This illustrates one weakness of an inferential framework that uses researchers' expectations to guide the interpretation of results. The meager Monte Carlo literature, replete with idealized data conditions, provides little assurance that researchers' expectations are well justified.

Third, under the taxon-detection framework, it is unclear how to interpret results that are not consistent with expectations for taxonic variables. The absence of evidence for taxonic structure does not logically support an inference of dimensional structure because this alternative does not exhaust the possibilities. For example, the indicators may have been insufficiently valid (e.g., too low a value of Cohen's *d* for the separation between groups) to detect taxonic structure. The ambiguity of apparently nontaxonic results is a weakness of this inferential framework, and it leads some to conclude that one can never reach a conclusion of dimensional structure. This would be an unsatisfactory state of affairs because a taxometric analysis could only yield one type of structural conclusion (taxonic), with all other results deemed inconclusive. This could lead to a publication bias in favor of taxonic results, which in turn could create or exacerbate confirmation biases in favor of taxonic results. Also, given

the recently renewed interest in dimensional models of mental disorders (see, e.g., the special section of the *Journal of Abnormal Psychology* edited by Krueger, Watson, & Barlow, 2005), an inferential framework that disallows dimensional conclusions would render the taxometric method unappealing to many researchers.

INFERENTIAL FRAMEWORK 2: DIMENSIONAL STRUCTURE AS A NULL HYPOTHESIS

This framework is a rather bold departure from the approach pioneered by Meehl, who stressed the virtues of *not* testing null hypotheses and favored a methodological emphasis on consistency testing. Nonetheless, some individuals (e.g., Beauchaine, 2003) have proposed that dimensional structure be treated as a taxometric null hypothesis (H_0). When results diverge from those expected for dimensional data, researchers are urged to reject H_0 and reach a taxonic conclusion. Despite familiarity to those who have performed other tests of statistical significance to evaluate null hypotheses, there are a number of important limitations to this framework. The first three parallel those outlined for the taxon-detection framework: (a) The problem of how much results must depart from those expected for dimensional structure to reject H_0 is not specified, raising concerns about confirmation bias; (b) the approach depends critically on researchers' knowledge of results expected for H_0, which is incomplete and can lead to misinterpretations when expectations are mistaken; and (c) the prohibition on reaching a conclusion of dimensional structure is even stronger because accepting H_0 conjures a deeply ingrained taboo.

An additional problem merits consideration: Rejecting H_0 (dimensional structure) does not logically support a taxonic conclusion, for this does not exhaust the alternative hypotheses. The method was designed with the narrow goal of differentiating two structural models that by no means span the full range of possibilities. Eliminating one of these does not necessarily provide much support for the other. Whereas the taxon-detection framework at least encouraged a comparative judgment (the relative support for taxonic and dimensional structures), treating dimensional structure as H_0 forces researchers to commit a logical error to reach a conclusion of taxonic structure.

INFERENTIAL FRAMEWORK 3: TWO COMPETING STRUCTURAL HYPOTHESES

Rather than seeking evidence that supports one structural model (taxon-detection) or that refutes another structural model (dimensional structure as H_0), one can treat taxonic and dimensional structural models as two competing hypotheses and evaluate the relative support for each. Instead of attempting to reach an absolute conclusion that latent structure is either taxonic (by detecting a taxon) or not dimensional (by rejecting H_0), one attempts to reach a relative conclusion that one of these represents a more valid structural model for the target construct than the other. Adopting such a competing-models inferential framework not only appears more consistent with the goals of the taxometric method but it also may ease the problems posed by the other inferential frameworks (Ruscio et al., 2007).

First, demonstrating that evidence not only supports one structural model but also that it fails to support another provides some protection against confirmation bias. Biases are constrained by the requirement that results simultaneously corroborate one structural model and refute another. This is a more rigorous standard because it disallows the interpretation of results that are equally consistent with taxonic and dimensional data, which can arise when

one analyzes data that are insufficiently valid to distinguish taxonic and dimensional data or when one implements taxometric analyses inappropriately.

Second, the competing-models framework lends itself to an empirically grounded approach to implementing the taxometric method that overcomes the limitations of our present knowledge of the expected results for taxonic and dimensional data. As discussed earlier, one can use simulated comparison data to provide an interpretive aid that fills gaps in the Monte Carlo literature (Ruscio et al., 2007). This also avoids the need to analytically prove what results should look like for either structure. The key to the empirically grounded approach is that discernibly different results would be expected for taxonic and dimensional data, but neither this difference nor the results for either structure need remain constant across all data conditions. For example, Maraun and Slaney (2005) proved that some taxometric analyses can yield either single- or double-peaked curves for taxonic data, and they argued that this illustrates a serious deficiency in the taxometric method because the predominant expectation is single-peaked curves. For those operating under the taxon-detection or dimensional-structure-as-H_0 inferential frameworks, this indeed poses a significant problem: When one's expectations are wrong, the chances of misinterpreting results increase. However, for those treating the two structural models as competing hypotheses, it is unimportant whether analyses of taxonic data yield single- or double-peaked curves. What is critical is that there be some discernible difference in results across taxonic and dimensional data when other important characteristics are held constant (e.g., sample size, number of indicators, indicator correlations, implementation of the analysis). These differences inform the interpretation of results with no need for a universal, analytically proven set of expectations.

Third, there would be no prohibition on reaching conclusions of either taxonic or dimensional structure. To the extent that results are more consistent with one structural model than the other, this supports a preference for that model. When results are equally consistent with both models, one should be reluctant to draw any conclusion. In fact, this is a significant strength of this framework: The ambiguities that led to interpretational difficulties or taboos under the other frameworks can, at least sometimes, be disentangled under this one. For example, if results for taxonic comparison data and dimensional comparison data cannot be distinguished, this suggests that the data, analyzed in the chosen way, do not afford a powerful test between these structural models. Whereas the taxon-detection framework confounds inadequate data with dimensional structure when results appear nontaxonic and the dimensional-structure-as-H_0 framework prohibits all inferences of dimensional structure and requires a logical error to infer taxonic structure, the competing-models framework supports inferences of taxonic or dimensional structure while providing a mechanism for withholding judgment in cases of ambiguity because of problematic data or implementation.

TAXOMETRIC PROCEDURES

To distinguish taxonic and dimensional structure, the taxometric method contains multiple data-analytic procedures. In this section, the most popular procedures will be introduced and illustrated through analyses of several artificial data sets generated using latent structures known to be taxonic or dimensional. Artificial data sets are used because there may be no construct in criminal justice that commands sufficient agreement regarding the taxonic versus dimensional nature of its latent structure; for present purposes, it is essential to know the

latent structure of each data set. The rationale by which procedures operate will be described along with the decisions that one must make to implement them.

TAXONIC AND DIMENSIONAL SAMPLES

To demonstrate the way that taxometric analyses can distinguish taxonic and dimensional data, the first two illustrative data sets were generated with the goal of providing unambiguous results. Taxonic Sample 1 (TS1) was created by drawing $n = 300$ cases from each of two populations (taxon and complement), which yields a taxon base rate of $P = .50$. Four indicators varied along normal, continuous distributions and were uncorrelated within each population, and means on each indicator were separated by 2.00 within-group standard deviations ($\sigma = 1.00$ for each indicator in each population, yielding $d = 2.00$ for each indicator); data were restandardized so that $SD = 1.00$ for each indicator in the final sample of $N = 600$. Dimensional Sample 1 (DS1) also contained $N = 600$ cases, but scores on the four indicators were drawn from a single multivariate normal population ($\sigma = 1.00$ for each indicator), with indicator correlations equal to those in a single population containing an equal mixture of the two populations sampled to create TS1 ($\rho = .50$). Indicators in TS1 were restandardized to equate variances with those in DS1. Thus, TS1 and DS1 share the same sample size, number of indicators, indicator variances, expected indicator correlation matrices, and expected within-group indicator distributions.

Because TS1 and DS1 involve a number of idealizations, the taxometric results obtained through their analysis cannot safely be generalized to less-ideal data conditions. To demonstrate that taxometric results can be affected by data conditions, a second series of illustrative data sets was generated. The most salient ways in which Taxonic Sample 2 (TS2) differs from TS1 are that $P = .25$, indicators are positively skewed within groups, indicator variances are unequal across groups, and indicators vary along five ordered categories rather than continuous scales. Dimensional Sample 2 (DS2) also contains indicators that vary along five ordered categories in positively skewed distributions. Whereas TS1 and DS1 are very similar to one another, differing primarily in their latent structures, TS2 and DS2 differ from one another a bit more and may be more representative of actual research data. It should be more challenging to correctly identify the latent structures of TS2 and DS2 than those of TS1 and DS1. For further details regarding these four samples, see Figure 1 and Table 1 for indicator distributions and Table 2 for indicator correlations.

The taxometric method includes a number of procedures, three of which will be described and illustrated here: mean above minus below a cut (MAMBAC; Meehl & Yonce, 1994), maximum covariance (MAXCOV; Meehl & Yonce, 1996), and maximum eigenvalue (MAXEIG; Waller & Meehl, 1998). These procedures form the core of what Meehl referred to as his "coherent cut kinetics" approach. Each word captures an important element of the approach, which involves searching for predictable patterns in the results ("coherent") that are obtained as a threshold ("cut") slides along one of the variables ("kinetics"). There are at least two other taxometric procedures that belong to the coherent cut kinetics family, the maximum-slope (Grove, 2004) and consistency-hurdles (Golden, 1982) procedures. The maximum-slope procedure has been recommended as useful in lieu of MAXCOV or MAXEIG when only two indicators are available for analysis (Meehl, 1999), but it appears that this is rarely the case. The consistency-hurdles procedure shares some important similarities to MAMBAC but has seldom been used in the past couple of decades. Waller and Meehl (1998)

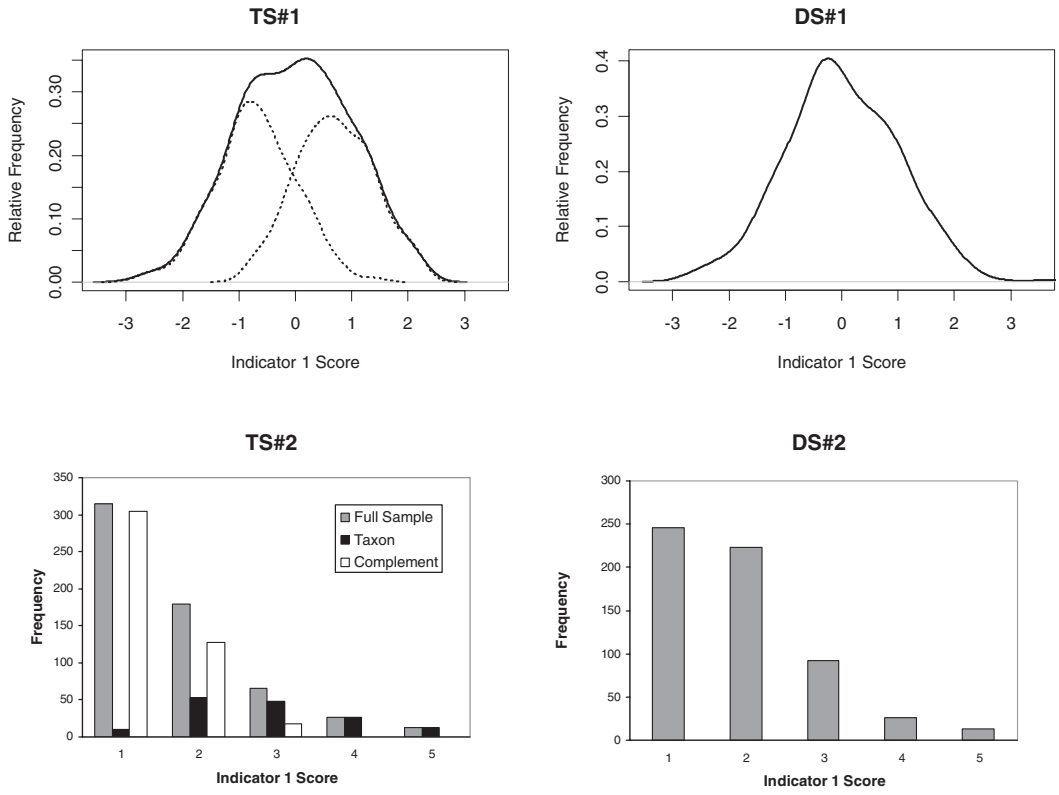

Figure 1: Distribution of the First Indicator in Each of the Four Samples
Note. For TS1 and DS1, indicators vary along continuous scales. For TS1, the full-sample distribution is plotted as a solid curve and within-group distributions are plotted as dashed curves. For TS2 and DS2, indicators vary across five ordered categories. For TS2, distributions for the full sample and within each group are plotted separately. TS1 = Taxonic Sample 1; DS1 = Dimensional Sample 1; TS2 = Taxonic Sample 2; DS2 = Dimensional Sample 2.

introduced a taxometric procedure called latent mode (L-mode) that operates via factor analysis. Because L-mode does not belong to the coherent cut kinetics family of taxometric procedures and has been studied less intensively than MAMBAC, MAXCOV, and MAXEIG, it will not be discussed further here; interested readers should consult Waller and Meehl (1998) for additional information about L-mode.

MAMBAC

This procedure requires at least two indicators, one referred to as the "input" indicator and the other referred to as the "output" indicator, and it involves the search for an optimal cut score to differentiate two groups (Meehl & Yonce, 1994). To begin, cases are sorted using the input indicator. Next, a cut score is located near the bottom of the range of scores on the input indicator, and a mean difference is calculated on the output indicator. Specifically, the mean output score for cases that are less than the cut is subtracted from the mean output score for cases greater than the cut. The cut score is then increased a bit along the input indicator, and a new mean difference is calculated using the output scores greater than and

TABLE 1: Indicator Distributions in the Illustrative Data Sets

	Taxonic Sample 1					*Taxonic Sample 2*				
	M	SD	*Skew*	*Kurtosis*	d	M	SD	*Skew*	*Kurtosis*	d
Indicator 1	0.00	1.00	–0.05	–0.50	2.00	1.74	0.97	1.41	1.64	2.09
Taxon	*0.71*	*0.69*	*0.05*	*–0.50*		*2.86*	*1.06*	*0.39*	*–0.51*	
Complement	*–0.71*	*0.72*	*–0.03*	*0.12*		*1.36*	*0.56*	*1.25*	*0.60*	
Indicator 2	0.00	1.00	0.04	–0.53	1.81	1.56	0.87	1.70	2.53	2.39
Taxon	*0.67*	*0.73*	*0.01*	*0.11*		*2.65*	*0.98*	*0.42*	*–0.26*	
Complement	*–0.67*	*0.75*	*0.28*	*0.10*		*1.19*	*0.41*	*1.84*	*2.22*	
Indicator 3	0.00	1.00	0.00	–0.51	2.21	1.62	0.90	1.63	2.49	1.92
Taxon	*0.74*	*0.68*	*0.03*	*–0.33*		*2.61*	*1.07*	*0.46*	*–0.35*	
Complement	*–0.74*	*0.67*	*–0.17*	*0.45*		*1.28*	*0.50*	*1.63*	*2.43*	
Indicator 4	0.00	1.00	–0.06	–0.67	2.00	1.55	0.88	1.78	2.86	2.09
Taxon	*0.71*	*0.68*	*–0.12*	*–0.33*		*2.57*	*1.07*	*0.41*	*–0.47*	
Complement	*–0.71*	*0.74*	*0.21*	*0.21*		*1.20*	*0.43*	*1.88*	*2.60*	

	Dimensional Sample 1				*Dimensional Sample 2*			
	M	SD	*Skew*	*Kurtosis*	M	SD	*Skew*	*Kurtosis*
Indicator 1	0.00	1.00	0.03	–0.04	1.90	0.96	1.10	0.98
Indicator 2	0.00	1.00	0.20	0.43	2.01	1.04	1.05	0.68
Indicator 3	0.00	1.00	–0.04	–0.18	1.92	1.03	1.10	0.67
Indicator 4	0.00	1.00	–0.10	–0.07	1.72	0.80	1.03	0.97

Note. *d* = Cohen's *d*, the standardized separation between taxon and complement groups. Correlations calculated within groups are italicized, and all other correlations were calculated in the full sample of data.

less than this cut. These steps are repeated until the cut score has reached a point near the upper end of the range of scores on the input indicator. A graph is constructed to display the results of the MAMBAC analysis. The *x* axis corresponds to the input indicator, and the *y* axis shows mean differences on the output indicator. Each cut score–mean difference pair is plotted as a data point, and these are connected to form a MAMBAC curve. The shape of a MAMBAC curve is examined to determine whether it is more consistent with taxonic or dimensional latent structure.

If two groups are mixed in the full sample (taxonic structure), then there exists a cut score at which these groups can be distinguished most validly. In other words, for taxonic data, a MAMBAC graph should be peaked near the cut score that best differentiates taxon and complement members (Meehl & Yonce, 1994). In contrast, one would not expect a peak to emerge for dimensional data, as there are no groups to be differentiated. Instead, the expected shape of a MAMBAC curve for dimensional data is either relatively flat or concave (sometimes described as "dish-shaped"; Meehl & Yonce, 1994).

Figure 2 shows the results of MAMBAC analyses of the four illustrative data sets. This and subsequent figures include the results for simulated taxonic and dimensional comparison data as an interpretive aid. Each panel in these figures consists of two graphs, both containing the results for the target data set; these are plotted as individual data points connected with dark line segments. The graph on the left contextualizes this with the results from analyses of 10 samples of comparison data generated to reproduce characteristics of the target data using a taxonic structural model. The pair of light lines plotted in this graph represents the sampling distribution of results for the simulated taxonic data. The graph on the right contextualizes the results for the target data, with the results from analyses of 10 samples

TABLE 2: Indicator Correlations in the Illustrative Data Sets

	Taxonic Sample 1: Full Sample				*Taxonic Sample 2: Full Sample*		
	Indicator 2	*Indicator 3*	*Indicator 4*		*Indicator 2*	*Indicator 3*	*Indicator 4*
Indicator 1	.53	.53	.50	Indicator 1	.51	.42	.49
Indicator 2		.51	.47	Indicator 2		.47	.48
Indicator 3			.54	Indicator 3			.44
	Taxonic Sample 1: Taxon				*Taxonic Sample 2: Taxon*		
	Indicator 2	*Indicator 3*	*Indicator 4*		*Indicator 2*	*Indicator 3*	*Indicator 4*
Indicator 1	.11	–.03	–.06	Indicator 1	.16	.04	.12
Indicator 2		–.05	.05	Indicator 2		–.03	–.01
Indicator 3			.01	Indicator 3			.05
	Taxonic Sample 1: Complement				*Taxonic Sample 2: Complement*		
	Indicator 2	*Indicator 3*	*Indicator 4*		*Indicator 2*	*Indicator 3*	*Indicator 4*
Indicator 1	.10	.06	.06	Indicator 1	–.10	–.10	.00
Indicator 2		.12	–.07	Indicator 2		.08	.02
Indicator 3			.06	Indicator 3			–.03
	Dimensional Sample 1				*Dimensional Sample 2*		
	Indicator 2	*Indicator 3*	*Indicator 4*		*Indicator 2*	*Indicator 3*	*Indicator 4*
Indicator 1	.51	.55	.54	Indicator 1	.41	.41	.41
Indicator 2		.55	.51	Indicator 2		.38	.42
Indicator 3			.52	Indicator 3			.41

of comparison data generated to reproduce the same characteristics of the target data using a dimensional model. Comparing the results for the target data to those for both taxonic and dimensional data provides an easy-to-use, visually guided interpretive aid. Often, one can see plainly whether taxonic or dimensional comparison data yield results more similar to those for the target data.

In addition to this visual inspection, which necessitates a subjective judgment that is not always easy to make or defend, a quantitative measure has been developed to provide an objective index of the relative fit of results to those for the taxonic versus dimensional comparison data (Ruscio et al., 2007). This is done as follows. First, calculate the root mean squared residual (RMSR) between the N data points on the curves for the target data ($y_{res.data}$) and the simulated data ($y_{sim.data}$):

$$Fit_{RMSR} = \sqrt{\frac{\sum (y_{res.data} - y_{sim.data})^2}{N}}, \quad (1)$$

This is done once using the averaged curve for the simulated taxonic data, yielding $\text{Fit}_{RMSR-tax}$, and then using the averaged curve for the simulated dimensional data, yielding $\text{Fit}_{RMSR-dim}$. Then, combine these two measures to yield a comparison curve fit index (CCFI):

$$\text{CCFI} = Fit_{RMSR-dim} / (Fit_{RMSR-dim} + Fit_{RMSR-tax}) \quad (2)$$

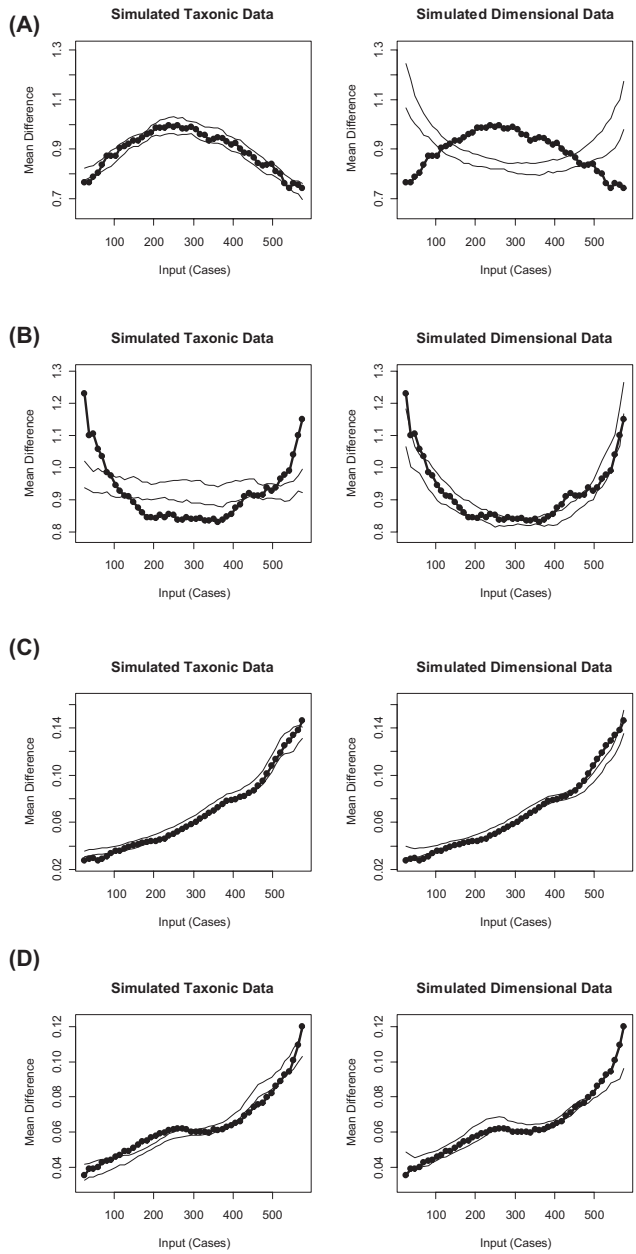

Figure 2: MAMBAC Analyses of the Four Illustrative Data Sets
Note. MAMBAC = mean above minus below a cut.

The CCFI can range from 0 (*most supportive of dimensional structure*) to 1 (*most supportive of taxonic structure*), with .50 representing a maximally ambiguous result—equally good (or poor) fit for both structures. Thus, the CCFI measures the relative fit of the observed results to those for the two structural models that the taxometric method is designed to distinguish. For details on how to generate the necessary comparison data sets as well as a series of Monte

Carlo studies that provide rigorous empirical tests supporting the utility of this technique, see Ruscio et al. (2007) as well as the second study below.

Now we can return to the MAMBAC results. The interpretive aid provided by the generation and analysis of comparison data is not necessary to draw accurate conclusions for the unambiguous curves shown in the top two panels of Figure 2, which contain results for TS1 (Panel A) and DS1 (Panel B). Both of these bear out the expectations noted earlier. In Panel A of Figure 2, the MAMBAC results for TS1 are peaked near the center of the graph, which correctly suggests taxonic structure. Likewise, Panel B of Figure 2 shows that the MAMBAC results for DS1 are concave, which correctly suggests dimensional structure. CCFI values also would have correctly identified the latent structure of these data: For Panel A, CCFI = .923 (which is greater than .50 and indicative of taxonic structure); for Panel B, CCFI = .169 (which is less than .50 and indicative of dimensional structure).

In Panels C (TS2) and D (DS2), the results are more challenging to interpret. In both cases, the MAMBAC curve increases to a cusp at the upper end. The conventional wisdom in the taxometric literature is that a cusp, like a peak, is indicative of taxonic structure (Meehl & Yonce, 1994). Following this interpretive guideline, one would correctly identify the results for TS2 as taxonic but mistakenly identify the results of DS2 as taxonic, too. By generating and analyzing taxonic and dimensional comparison data, it becomes clear that these results are more ambiguous than they appear at first. Most striking is that the results for both taxonic and dimensional data yield similarly cusped curves. With a careful attention to detail, visual inspection of the curves in Panel C reveals that the results for the research data do conform to those for taxonic comparison data more closely than to those for dimensional comparison data; a CCFI value of .657 confirms this impressionistic judgment and correctly identifies the taxonic structure of TS2. Even with a careful inspection, the curves in Panel D remain difficult to interpret with much confidence. Although the CCFI value of .474 correctly identifies the dimensional structure of DS2, it also suggests that these are not nearly as definitive results as one might wish. Whether or not one chooses to refrain from interpreting a CCFI value this close to .50, it clearly prevents the mistaken conclusion of taxonic structure that would almost certainly be reached without the use of simulated comparison data as an interpretive aid.

Implementing the MAMBAC procedure requires careful attention to a series of potentially important issues that are summarized in Table 3. First, one must determine how to assign the available variables to the required roles of input and output indicator. When there are only two variables, this is simple: Treat one as the input indicator and the other as the output indicator for a MAMBAC analysis, and then reverse the assignments for a second MAMBAC analysis. (The fact that one can perform multiple MAMBAC analyses of the same data provides a first illustration of consistency testing, which will be discussed later.) When there are more than two variables available, one can perform MAMBAC analyses using all pairwise combinations of the variables—with each pair assigned in both directions as input/output and output/input indicators. Alternatively, one can perform MAMBAC analyses by assigning one variable to be the output indicator and combining (e.g., summing) the remaining variables to serve as a composite input indicator and then repeating this so that each variable serves once as the output indicator. Or, if there are many variables available, one can form composite input and output indicators using subsets of the variables.

Second, one must decide how to locate cut scores along the input indicator. A certain minimal distance should be maintained from each end of the range of scores on the input indicator to stabilize the mean differences plotted near the ends of the MAMBAC curve.

TABLE 3: Implementation Decisions for Taxometric Procedures

1. Calculating results using the output indicator(s)
   a. One output indicator, mean differences with sliding cut score = MAMBAC
   b. Two output indicators, conditional covariances in successive subsamples = MAXCOV
   c. Two or more output indicators, conditional eigenvalues in successive subsamples = MAXEIG
2. Assigning *k* indicators to input/output roles
   a. MAMBAC
      i. All pairwise combinations; $k(k-1)$ curves
      ii. Composite input indicator, single output indicator (or the reverse); *k* curves
      iii. Composite input and output indicators; intermediate number of curves
   b. MAXCOV/MAXEIG
      i. All possible triplets; $k(k-1)(k-2)/2$ curves
      ii. One input indicator, all others as outputs (MAXEIG only); *k* curves
      iii. Composite input indicator, pair of output indicators; $k(k-1)/2$ curves
      iv. Composite input and output indicators; intermediate number of curves
3. Placing cuts along the input indicator
   a. MAMBAC
      i. Minimal sample size beyond first and last cuts
      ii. Cut between each successive case, at every *n*th case, at intact scale values, or using *SD* units
      iii. Internal replications if cut separates tied scores
   b. MAXCOV/MAXEIG
      i. Minimum subsample size
      ii. Intervals (nonoverlapping)—equal-sized, *SD* units, or intact scale values
      iii. Overlapping windows—number and overlap
      iv. Internal replications if tied scores assigned to different subsamples
4. Graphing and presenting results
   a. Constructing the *x* axis
      i. MAMBAC—case numbers versus input indicator scores
      ii. MAXCOV/MAXEIG—input indicator scores versus subsample numbers
   b. Scaling the *y* axis
      i. Range of values observed in analysis, range specified in advance, or formula-based range
      ii. Range of values held constant across curves (standardize indicators) versus varied by curve
   c. Present "raw" curve, smoothed curve, or both
   d. Present full panel of curves, averaged curve, or both

Note. MAMBAC = mean above minus below a cut; MAXCOV = maximum covariance; MAXEIG = maximum eigenvalue.

For example, one might locate the first (and last) cuts 25 cases from either end of the input range. Intermediate cuts can be located between each successive case, at every *n*th case (equivalent to specifying the number of cuts and locating them at equally-spaced intervals), intact scale values (e.g., between each score value observed on a behavioral checklist), or using *SD* units (e.g., after standardizing the input indicator, placing cuts every .25 *SD*s). When cuts are located between tied scores, this introduces noise into the analysis because the sorted order of tied scores is arbitrary. A remedy for this problem is to perform internal replications (Ruscio et al., 2006) by randomly resorting tied scores and repeating the analysis. This can be done many times, with the averaged results retained for graphing and interpretation. In all analyses of TS2 and DS2 presented in this article, 10 internal replications were used because the coarse indicators, which varied across just five ordered categorical values, resulted in many tied scores.

Third, there are a number of issues to consider when graphing and presenting results. The *x* axis can be constructed using case numbers or input indicator scores. The *y* axis can be scaled using the range of mean differences observed in the analysis or a range specified in advance. The range can be held constant across all curves in an analysis, or it can be

allowed to vary for each curve. For each MAMBAC graph, one can present the "raw" curve, containing the actual mean differences; a smoothed curve (e.g., by using running medians or another technique to produce a less "noisy" curve that more clearly reveals its true shape); or both. Finally, one can present the full panel of curves—recall that even with only two variables, this yields two MAMBAC curves—or an averaged curve, or both (e.g., by superimposing the average on a single graph that contains the full panel of curves). When curves will be averaged, it may be helpful to standardize each indicator to remove fluctuations across curves because of unequal variances.

There are many choices that one must make to implement MAMBAC but no systematic studies of the most appropriate way to make these decisions. Ruscio et al. (2006) have discussed pros and cons of many options and demonstrated the importance of making informed choices. I would like to emphasize the utility of an empirically grounded approach to implementing the taxometric method. Specifically, one can generate taxonic and dimensional comparison data and analyze these samples in various ways to determine which implementation yields the cleanest differentiation between these structures. Then, one can analyze the target data using this technique. This approach replaces much of the guesswork and obedience to tradition with an empirically driven assessment of what is likely to work best in a particular investigation. For example, consider the MAMBAC results shown in Figure 3. Each of these shows a MAMBAC analysis of TS2, but two implementation options were varied. In Panels A and B, 10 cuts were used, whereas in Panels C and D, 50 cuts were used. In Panels A and C, no internal replications were used, whereas in Panels B and D, 10 internal replications were used. The CCFI values were highly ambiguous for Panels A and B (.473 and .491, respectively), and they departed from .50 more conclusively for Panels C and D (.614 and .657, respectively). This suggests that, for this data set, using 50 cuts is better able to differentiate taxonic from dimensional structure than using 10 cuts—all else being equal. Likewise, the CCFI departed from .50 more conclusively for Panel D than for Panel C, which supports the use of 10 internal replications. (The astute reader may have noticed that Panel D of Figure 3 is identical to Panel C of Figure 2; in fact, analyses such as those shown in Figure 3 guided my decision making as I implemented MAMBAC in the analyses of the illustrative data sets, and I used 50 cuts and 10 replications in each analysis.) Had one failed to use a sufficiently large number of cuts, one might have obtained ambiguous results; with a sufficiently large number of cuts, less ambiguous results were obtained, and in this case, they correctly identified the taxonic structure of TS2. Using 10 internal replications enhanced the clarity of results a bit, too. This is a small sampling of the ways that MAMBAC could have been implemented—one might examine the merits of different techniques for assigning variables to the required roles of input and output indicators, of smoothing or averaging curves, and so forth—and the choices that one makes can influence the results. Researchers are strongly encouraged to take advantage of readily available computing power to adopt an empirically grounded approach to implementing taxometric procedures.

MAXCOV

Whether the latent structure of the target construct in a taxometric investigation is taxonic or dimensional, the indicators used to represent it will be positively correlated with one another. In the case of taxonic structure, the positive correlations stem, at least in part,

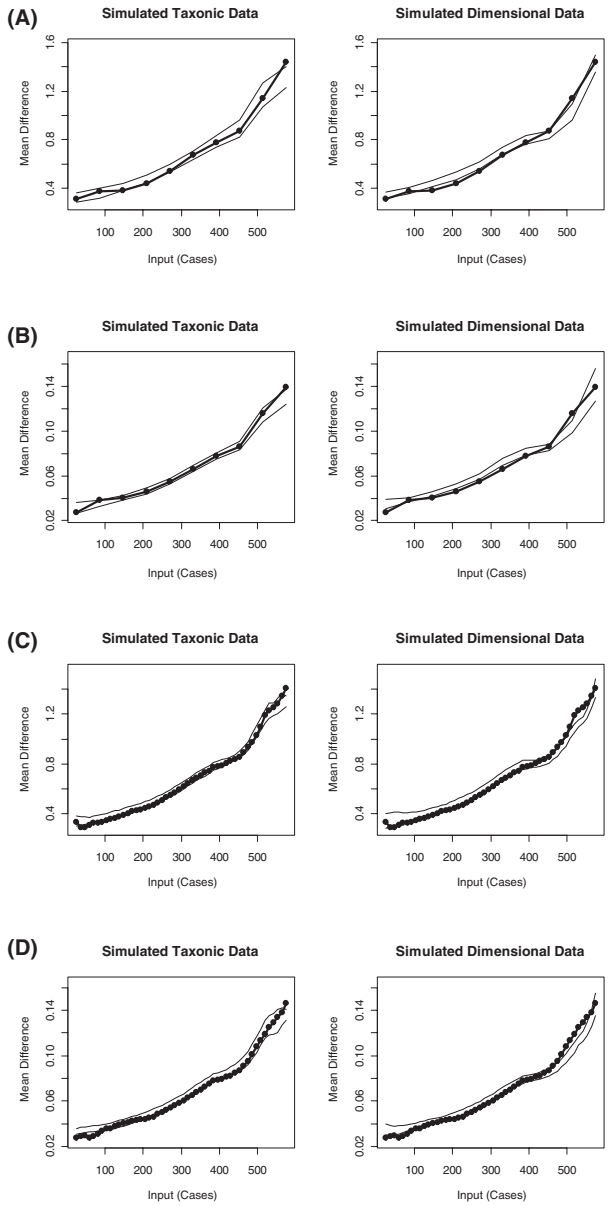

Figure 3: MAMBAC Analyses of TS2, Implemented in Four Different Ways
Note. MAMBAC = mean above minus below a cut; TS2 = Taxonic Sample 2.

from the mixture of groups. Provided that they are keyed in the same direction, any two variables that validly differentiate two groups will be positively correlated because individuals in one group (the taxon) tend to score higher on both variables than individuals in the other group (the complement). In the case of dimensional structure, the positive correlations stem from each variable's loading onto one or more shared factors. It is possible that some correlations among indicators of a multidimensional construct might be negligible in magnitude, or even negative, if the indicators load most strongly onto different factors that are

themselves negligibly or negatively correlated. However, if one observes correlations among indicators that are not positive and substantial in magnitude, submitting them to taxometric analysis will not afford an informative test between taxonic and dimensional structure. In order for the taxonic structural model to be a viable competitor to the dimensional model, it must be conceivable that the indicators distinguish two groups with sufficient validity to be detected in a taxometric analysis. Negligible or negative indicator correlations preclude this possibility.

The MAXCOV procedure can be used to help determine whether positive indicator correlations are or are not attributable to the mixture of two groups (Meehl & Yonce, 1996). Whereas MAMBAC requires at least two indicators and operates by sliding a cut score to examine the pattern of mean differences higher and less than the cut, MAXCOV requires at least three indicators and operates by calculating the covariance between two variables (the output indicators) within subsamples ordered along another variable (the input indicator). Because the covariance is calculated within a subsample, it is referred to as a conditional covariance. As in MAMBAC, one begins by sorting cases along the input indicator. Then the full sample is divided into a series of subsamples ordered along the input indicator. The conditional covariance between the output indicators is calculated within each subsample, and the results are plotted. In a MAXCOV graph, the *x* axis denotes scores on the input indicator, and the *y* axis denotes (conditional) covariances. Curve shape is used to infer the latent structure of the target construct.

In the case of taxonic structure, indicator covariance arises, at least in part, because of the mixture of groups. In subsamples that contain a mixture of taxon and complement members, the covariance between output indicators should be large; it should reach a maximum when groups are mixed in equal proportions. To the extent that a subsample is more homogenous, containing predominantly taxon or complement members, the covariance between output indicators should decrease. Therefore, taxonic data should yield a peaked or cusped MAXCOV graph (Meehl & Yonce, 1996). A peak would emerge if there are sufficiently homogeneous subsamples on both sides of the subsample in which group members are mixed most evenly. A cusp would emerge if there are too few members of one group to populate homogeneous subsamples on one side of the most heterogeneous subsample. For example, if there are only 50 taxon members in a sample of 1,000 individuals, even the uppermost subsamples may contain more complement than taxon members. Even though this would not produce a peaked curve, it could produce a cusped curve, as the groups are mixed in more even proportions in higher-scoring subsamples. Hence, either a peaked or cusped MAXCOV curve is conventionally interpreted as evidence in favor of taxonic structure (Meehl & Yonce, 1996).

In the case of dimensional structure, indicator covariance arises because of loadings onto one or more shared factors. Because there are no groups being mixed, it is not expected that the covariance between output indicators will vary systematically across ordered subsamples. By convention, a relatively flat MAXCOV curve is interpreted as evidence in favor of dimensional structure (Meehl & Yonce, 1996).

Figure 4 shows MAXCOV results for the four illustrative data sets, once again accompanied by results for 10 samples apiece of simulated taxonic and dimensional comparison data. The results for TS1 (Panel A) look unambiguously taxonic; the CCFI value of .880 confirms this impression. The results for DS1 (Panel B) are not as simple because the curve for the target data is not as flat as one might expect for dimensional data. Nonetheless, it is clear that the results are more supportive of dimensional than taxonic structure; the CCFI value of .271

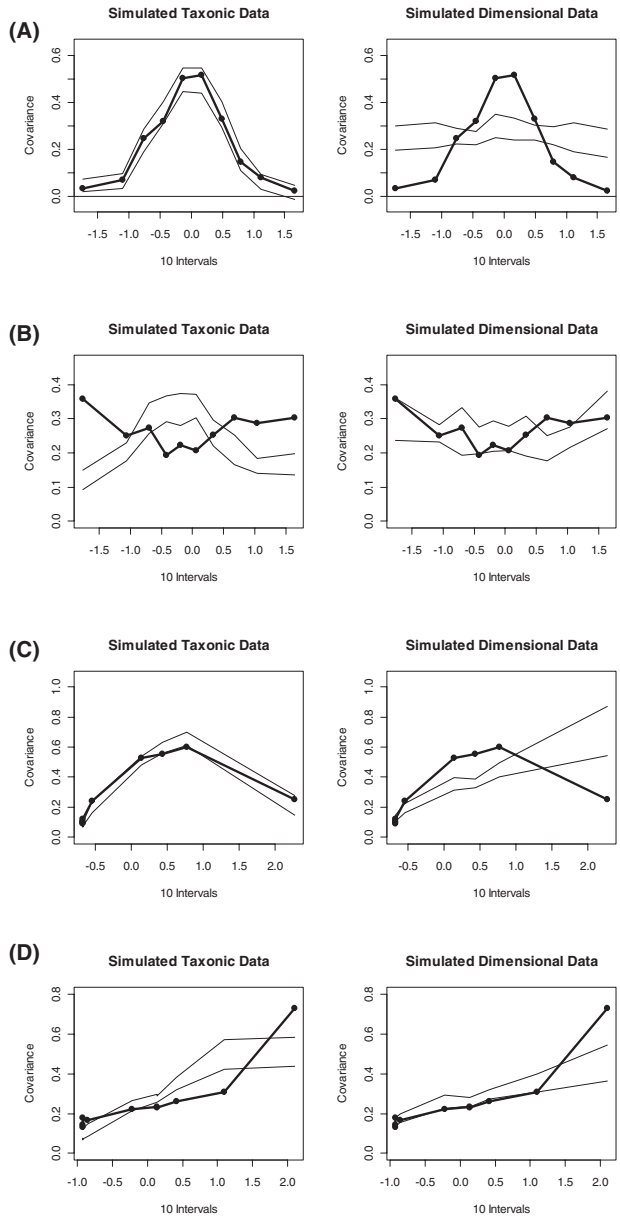

Figure 4: MAXCOV Analyses of the Four Illustrative Data Sets
Note. MAXCOV = maximum covariance.

corroborates this impression. The results for TS2 (Panel C) correctly identify its taxonic structure, as does the CCFI value of .807. As was the case for MAMBAC, the results for DS2 (Panel D) could easily be misread as evidence of taxonic structure if simulated comparison data were not used as an interpretive aid. According to conventional standards, the cusped MAXCOV curve is suggestive of a small taxon. Given the results for the comparison data, however, this would be more difficult to maintain. Even dimensional data yielded a cusped curve, and in fact, the right-end cusp was more dramatic for the dimensional than the taxonic

comparison data. A visual inspection might suggest that the evidence is more supportive of dimensional structure. Even though the CCFI value of .466 underscores the ambiguity of reaching any conclusion, as it is very close to .50, if one were willing to reach a conclusion on the basis of this value, it would correctly identify the dimensional structure of DS2. Although the use of comparison data does not afford a strong conclusion, the ambiguous results at least give one pause in reaching an unwarranted inference of a small taxon.

To implement the MAXCOV procedure, there are at least as many decisions to make as for MAMBAC. Because the following procedure (MAXEIG) shares many similarities to MAXCOV, the implementation of both procedures will be discussed after MAXEIG has been described and illustrated.

MAXEIG

This procedure was introduced by Waller and Meehl (1998) as a multivariate generalization of MAXCOV. Although there are several differences in the ways that these procedures have been implemented, there is only one fundamental difference between these procedures: whether covariances or eigenvalues are calculated to measure the strength of the association between output indicators. To perform MAXCOV, one calculates the conditional covariance between two indicators within each ordered subsample of cases. To perform MAXEIG, one calculates the first (largest) conditional eigenvalue of the covariance matrix constructed from two or more indicators within each ordered subsample of cases. The covariance matrix is the usual variance–covariance matrix with the diagonal of variances replaced by zeros to leave only off-diagonal covariances. The first eigenvalue of such a matrix indexes the extent to which the indicators covary. The MAXEIG procedure takes advantage of the availability of more than three indicators to yield results that can, under many conditions, more validly distinguish taxonic and dimensional structure. For example, suppose that a researcher has six variables. Using MAXCOV, only two variables per analysis can serve as output indicators to calculate covariances. Using MAXEIG, on the other hand, all but one of the available variables can serve as output indicators to calculate eigenvalues (the other variable is required to serve as the input indicator). Provided that each indicator is sufficiently valid, calculating conditional eigenvalues using five output indicators is likely to yield more informative results than calculating conditional covariances using two output indicators (Waller & Meehl, 1998). By convention, the interpretation of MAXEIG results follows the same logic as for MAXCOV: Peaked or cusped curves are suggestive of taxonic structure, and comparatively flat curves are suggestive of dimensional structure (Waller & Meehl, 1998).

Figure 5 shows MAXEIG results for the four illustrative data sets. The results for TS1 (Panel A) appear taxonic, and the CCFI value of .901 strongly supports this. Likewise, the results for DS1 (Panel B) appear dimensional, and the CCFI value of .288 strongly supports this. The results for TS2 (Panel C) correctly identify its taxonic structure, with confirmation provided by the CCFI value of .755. As was true for MAMBAC and MAXCOV, the MAXEIG results for DS2 (Panel D) remain the most ambiguous. The use of simulated comparison data once again prevents the mistaken interpretation of the cusped curve as indicative of a small taxon, and dimensional structure can be identified correctly—albeit with only a modest level of support based on a visual inspection or an interpretation of the CCFI value of .441.

Implementing either MAXCOV or MAXEIG involves a series of decisions with even more options than are available for MAMBAC. The first decision is whether to perform

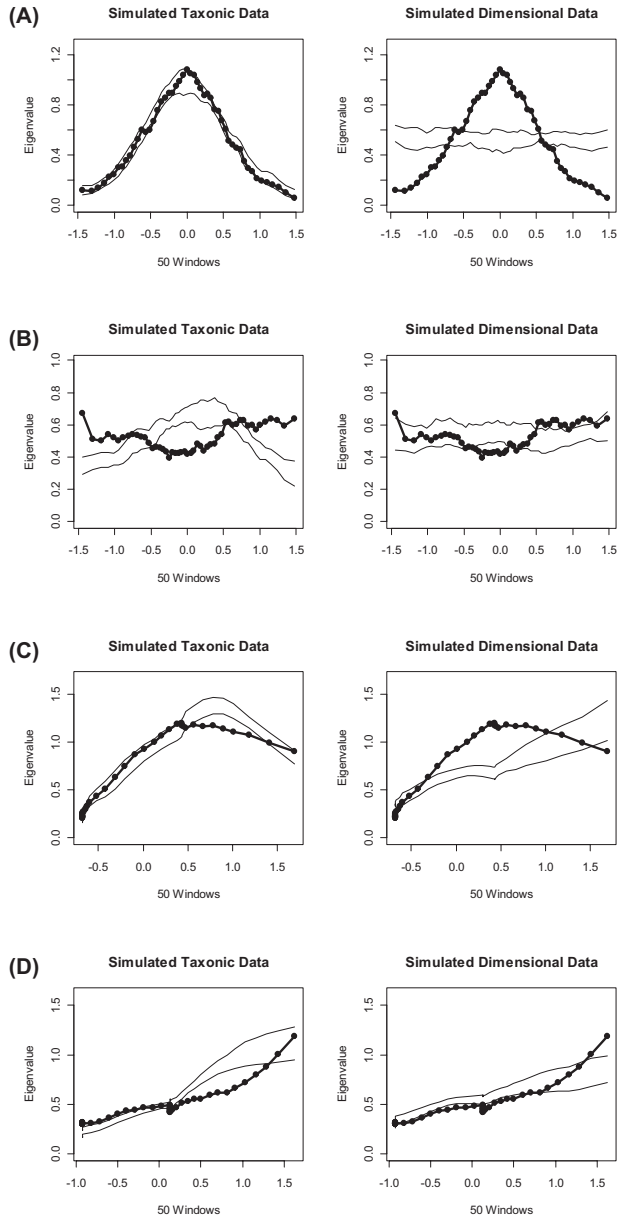

Figure 5: MAXEIG Analyses of the Four Illustrative Data Sets
Note. MAXEIG = maximum eigenvalue.

MAXCOV or MAXEIG, and the choice depends on whether conditional covariances or eigenvalues will serve as the measure of association between output indicators. Once this decision is made, MAXCOV and MAXEIG present nearly identical implementation options. A second decision is how to assign the available variables to the required roles of input and output indicators. There are many ways to do this. Variables can be assigned in all possible configurations of input/output/output indicator triplets; this is the traditional way to implement

MAXCOV. One variable per analysis can serve as the input indicator, with all others serving as output indicators; this is the traditional way to implement MAXEIG (and it cannot be done for MAXCOV because the calculation of covariances limits analysis to a pair of output indicators). Two variables per analysis can serve as output indicators, with the remaining variables combined to form a composite input indicator. Finally, subsets of variables can be assigned as output indicators and composite input indicators. For example, with six variables available, one could assign three (or four) to serve as output indicators and sum the remaining three (or two) to serve as a composite input indicator.

A third decision is how to place cuts along the input indicator to divide the full sample into ordered subsamples. It is useful to establish a minimal size below which no subsample will be allowed to fall (e.g., $n \geq 25$) and ensure that this is followed. The subsamples may consist of a series of intervals, which do not overlap with one another, or windows, which do overlap. Traditionally, MAXCOV has been implemented using intervals. These may be constructed by dividing the sample into a specified number of equal-sized intervals (e.g., deciles), by placing cuts at specified *SD* values (e.g., every .25 *SD*) or by placing cuts at intact scale values (e.g., at scores of 5, 10, 15, and so forth on a behavioral checklist). Traditionally, MAXEIG has been implemented using windows. Three interrelated values determine the placement of cuts to create a series of windows for a sample of size *N*: the number of windows (*W*), the proportion of overlap between adjacent windows (*O*), and the subsample size within each (n_w). By choosing any two of these values, the third can be determined (Waller & Meehl, 1998, p. 42):

$$n_w = \frac{N}{W \times (1 - O) + O} \tag{3}$$

Waller and Meehl (1998) used $O = .90$ for most of their analyses, and Ruscio et al. (2006) argued that there is little, if any, reason to use a lower value. Treating *O* as a constant, the only tradeoff that remains is between the number of windows and the size of each. Larger numbers of windows tend to yield more interpretable curves, at least until the sample size within each becomes so small that sampling error adds too much noise. When this occurs depends on the sample size and other factors; there is no universally optimal number of windows. Trial and error using the empirically grounded approach with simulated comparison data can help to determine an appropriate number of windows for a particular analysis. Finally, the use of internal replications is advised whenever cases with tied scores might be assigned to different subsamples (i.e., when using equal-sized intervals or windows).

A fourth set of decisions involves the graphing and presentation of results. When constructing the *x* axis, one can use either input indicator scores or consecutive numbers denoting subsamples. When scaling the *y* axis, one can use the range of values observed in the analysis, a prespecified range of values, or a range that is based on—but not equal to—the values observed in the analysis. For example, Ruscio et al. (2006) presented a formula that accentuates the difference between peaked/cusped curves and comparatively flat curves. Because the former are often indicative of taxonic structure and the latter are usually indicative of dimensional structure, this formula may assist in the visual inspection of results. Whichever technique that one uses, the range of values can be held constant across curves or allowed to vary. When comparison data are used to obtain a panel of graphs, the range should be held constant across panels; in other situations, a case can be made for either approach. (It is worth mentioning that the CCFI is independent of the scaling of the *x* or *y* axes, and hence,

not subject to potentially poor choices that investigators might make when constructing their graphs.) Finally, one can present "raw" curves, smoothed curves, or both, and one can present either a full panel of curves, an averaged curve, or both.

As noted earlier, this is simply an overview of the choices that must be made and many of the options that are available. Ruscio et al. (2006) discussed advantages and disadvantages of these options, including data conditions under which certain strategies may be more or less appropriate than others.

CONSISTENCY TESTS

Confidence in the conclusions drawn from a taxometric analysis accumulates as results from nonredundant analyses consistently indicate either taxonic or dimensional structure. Of course, "nonredundant" is itself a matter of degree. For example, the differences between the MAMBAC and MAXEIG procedures are greater than the differences between the MAXCOV and MAXEIG procedures. A taxometric study in which MAMBAC and MAXEIG yield consistent results would provide stronger support for a conclusion than would a study in which MAXCOV and MAXEIG yielded equally consistent results. In addition to performing multiple taxometric procedures, there are many ways that researchers can check the consistency of results. Unfortunately, guidelines for how to check the consistency of results do not exist. For a particular consistency test, what threshold should be used to distinguish taxonic and dimensional data? To what extent does this depend on characteristics of the data? Which tests provide incremental validity beyond the results of the primary taxometric procedures? How should the results of multiple procedures and consistency tests be integrated to reach a structural inference? These are basic but crucial questions, and they have received little research attention. Later, I will report the results of two Monte Carlo studies of taxometric consistency testing, with a special emphasis placed on the use of base rate estimates. Here, the available consistency tests are grouped into three broad categories.

PERFORMING TAXOMETRIC PROCEDURES IN MULTIPLE WAYS

In addition to using multiple taxometric procedures, each of them can be performed repeatedly. To implement each of the procedures described in this article, one assigns variables to each input/output indicator configuration to produce a panel of curves. This affords an examination of the consistency of their shapes. Multiple samples of data can be analyzed, and if sample size is sufficiently large, one might even analyze multiple subsamples. Although samples (or subsamples) drawn from the same population provide only an exact replication, with results differing only because of sampling error, samples drawn from different populations provide a more informative consistency check. Within any sample, the available data can be used to construct multiple indicator sets that represent the same target construct in different ways. For example, one set of indicators might be calculated in accordance with a theory of the construct, whereas another might be based on a factor-analytic study of the relations among items on a scale.

A final test based on the repeated performance of the same taxometric procedure is the inchworm consistency test introduced by Waller and Meehl (1998). To conduct this test, one repeats a MAXCOV or MAXEIG analysis using an increasing number of windows. For taxonic data, one expects to see a peak emerge more clearly with larger numbers of windows,

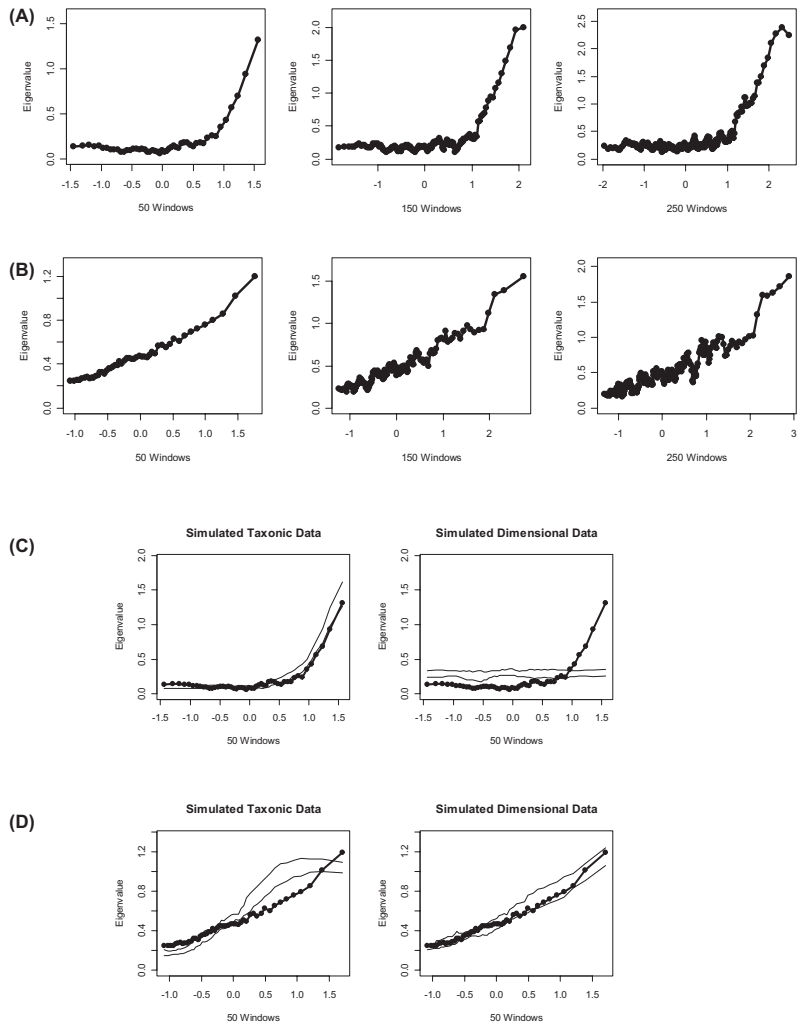

Figure 6: MAXEIG Analyses of Taxonic and Dimensional Samples That Present the Challenge of Disambiguating Cusped Curves

Note. MAXEIG = maximum eigenvalue.

whereas for dimensional data one would not expect to see an orderly progression toward a clearly defined peak (Waller & Meehl, 1998). Ordinarily, it might be advisable to perform a single analysis that uses the most appropriate number of windows. However, when a MAXCOV or MAXEIG curve is cusped, this is interpretationally ambiguous, and the inchworm test can clarify matters. Ruscio et al. (2004) drew attention to the ambiguity of cusped curves, noting and demonstrating that in addition to obtaining cusped curves with a very small (or large) taxon, analyses of skewed indicators can produce cusped curves even for dimensional data. Positively skewed indicators can yield rising curves with right-end cusps, and negatively skewed indicators can yield falling curves with left-end cusps. Ruscio et al. recommended that researchers use the inchworm consistency test—along with simulated comparison data—as an interpretive aid.

To illustrate the power of this test, consider the curves in Figure 6. Panel A shows the results of a MAXEIG analysis of a taxonic data set containing 50 taxon members in a sample of 1,000 individuals. With 50 windows (left graph), the curve was cusped and therefore ambiguous. As the number of windows was increased to 150 (middle graph) and then 250 (right graph), a peak emerged. This is the prototypical progression of curves that the inchworm test can provide for a small taxon. Panel B shows the results of a MAXEIG analysis of a dimensional data set with $N = 1{,}000$ and positively skewed indicators that vary across five ordered categories. Here, too, the curve is cusped and ambiguous with 50 windows (left graph). Even when the number of windows increased to 150 (middle graph) and 250 (right graph), the curve remained cusped, with no movement toward a better defined peak. These results are suggestive of dimensional structure. Because it is pertinent to the disambiguation of cusped curves, simulated comparison data were generated and analyzed. Panel C shows the results of a MAXEIG analysis of the same data from Panel A. Even with just 50 windows, the comparison data allow the easy identification of taxonic structure; the CCFI value of .805 strongly supports this conclusion. Panel D shows the results of a MAXEIG analysis of the same data from Panel B. Here, too, even with just 50 windows, the comparison data correctly suggest the dimensional structure of these data; the CCFI value of .216 supports this impression.

Beach, Amir, and Bau (2005) took issue with the recommendations of Ruscio et al. (2004), particularly the notion that simulated comparison data help to distinguish small taxa from dimensional constructs. They alleged that comparison data fail to reproduce data characteristics without bias and, more important, that the use of comparison data as an interpretive aid failed to detect the small taxa in a series of taxonic data sets that they created. Ruscio and Marcus (2006) addressed conceptual and methodological problems with this work and reanalyzed Beach et al.'s data sets to afford an objective interpretation of results using the CCFI. Taxonic structure was correctly identified for each sample, with a CCFI substantially in excess of .50 in most every case. In addition, Ruscio and Marcus noted that Beach et al. implemented the inchworm consistency test poorly, failing to use a sufficiently large number of windows to allow cusped curves to become peaked curves. With a larger number of windows, unambiguously peaked curves emerged. At present, there appears to be no more powerful way to disambiguate cusped curves than to use the inchworm consistency test and simulated comparison data.

EXAMINATION OF LATENT PARAMETER ESTIMATES AND CLASSIFIED CASES

Without question, the most common advice about consistency testing proferred in the taxometric literature is to check the agreement among multiple estimates of the taxon base rate. Each MAMBAC, MAXCOV, or MAXEIG curve can be used to calculate a base rate estimate (for details, see the publications in which these procedures were introduced: Meehl & Yonce, 1994, 1996, and Waller & Meehl, 1998, respectively, or Ruscio et al., 2006). These estimates can be compared across curves within a panel produced by a single procedure as well as across taxometric procedures, but the critical questions about how to make these comparisons have not been addressed. How should base rate consistency be quantified? Often, researchers calculate the *SD* of base rate estimates. What threshold should be applied to reach a conclusion of taxonic versus dimensional structure? Based on MAXCOV analyses of 20 taxonic samples drawn from a single set of population parameters plus 10 dimensional samples drawn from another population, Schmidt, Kotov, and Joiner (2004) recommended .10 as optimal, with *SD*'s below this level indicative of taxonic structure. To what extent does an

appropriate threshold vary with data conditions or taxometric procedures? How much weight should be assigned to this test relative to other taxometric results, such as the curve shapes themselves? Questions like this, which involve incremental validity, have not received nearly the attention that they warrant. One might even wonder whether an examination of base rate consistency provides any incremental validity, as each base rate estimate distills the complex information contained in a taxometric curve down to a single number that may obscure a pattern more validly interpreted by examining the curve itself. Schmidt et al. reported that this test "offered little incremental improvement" (p. 47) over an interpretation based on the MAXCOV curve shapes themselves. The Monte Carlo studies introduced later in this section examine this test more broadly, including multiple taxometric procedures and a very wide array of data conditions.

Although examining the coherence of base rate estimates is of questionable utility, a case removal consistency test that involves a predicted change in base rate estimates can be helpful (Meehl & Yonce, 1994; Ruscio, 2000). By systematically removing a subset of cases and re-running an analysis, one can test whether the base rate estimate changes in a manner more consistent with taxonic or dimensional structure. For example, suppose that a MAXEIG analysis yields a cusped curve and a base rate estimate of .10. This might indicate a small taxon, but it may be that skewed indicators of a dimensional construct produced the cusp. Consider what would happen if one removed the bottom quartile of cases, individuals with the lowest total scores on all indicators. If the construct is taxonic, the base rate estimate should be higher in this subsample, as most of the bottom quartile should be complement members. If the entire bottom quartile belongs to the complement, then the new base rate estimate should approach .10/.75 = .13. On the other hand, if the construct is dimensional, removing low-scoring cases would not be expected to increase the base rate estimate; in fact, it often lowers it further (Ruscio, 2000; Ruscio et al., 2006). Both the magnitude and direction of any change in the taxon base rate estimate can be informative.

There has been no systematic exploration of the utility of estimating other latent parameters (e.g., indicator validity) to examine their consistency. However, using a taxon base rate estimate as well as estimates of each indicator's valid and false positive rates, one can calculate each case's probability of taxon membership using Bayes' theorem (see Waller & Meehl, 1998, for details). Many authors have suggested that these probabilities should cluster around 0 and 1 for taxonic data and be more evenly distributed along this range for dimensional data. Before the validity of this proposed test was examined, Beauchaine and Beauchaine (2002) used it as one of two criteria for the successful detection of taxonic structure. Specifically, they calculated the proportion of Bayesian probabilities of taxon membership in the outermost deciles (< .10 or > .90) and applied a threshold of .80, with values greater than this indicative of taxonic structure. Subsequent testing has shown that this test distinguishes taxonic and dimensional structure with above-chance, but very modest, validity (Ruscio et al., 2007 press). At present, using this consistency test would be difficult to justify, particularly as other tests have performed much better under the same conditions. This issue is revisited in the second Monte Carlo study to be presented shortly.

ASSESSING MODEL FIT

Waller and Meehl (1998) introduced an index of fit between an observed statistical summary of the data and one predicted by a taxonic structural model. Having estimated the taxon

base rate as well as each indicator's variance within each group and validity in distinguishing groups, one can generate a predicted variance–covariance matrix and compare this to the matrix observed in the full sample. Waller and Meehl quantified this comparison using the goodness of fit index (GFI; Jöreskog & Sörbom, 2001). Based on analyses that were not described, they reported that taxonic data usually produced a GFI greater than .90, but dimensional data seldom did so. Beauchaine and Beauchaine (2002) relied on this test and threshold as their other criterion for the successful detection of taxonic structure. Several studies have revealed that the GFI does not distinguish taxonic and dimensional data very well (Cleland, Rothschild, & Haslam, 2000; Haslam & Cleland, 2002; Ruscio et al., 2007). One potential weakness of the GFI in a taxometric analysis is that when the observed results are equally consistent (or inconsistent) with taxonic and dimensional structure, this index can be misleading. For example, fit may be good for both models, but even though the GFI exceeds .90, fit may be stronger for the dimensional model; interpreting the high GFI as supportive of taxonic structure would be problematic. Likewise, fit may be poor for both models, yielding a GFI less than .90 yet better for taxonic structure. The GFI is included in the second Monte Carlo study presented below.

Whereas the GFI quantifies the fit of a taxonic model, the CCFI draws on simulated taxonic and dimensional comparison data to quantify the relative fit of taxonic and dimensional models. This addresses the problem identified above for the GFI, and Ruscio et al. (2007) found that from the results of the same MAXEIG analyses, the CCFI distinguished taxonic and dimensional structure much better.

MONTE CARLO STUDY 1: THE BASE RATE CONSISTENCY TEST (BRCT)

The BRCT has been recommended without hesitation in virtually every publication on the taxometric method, and some version of this test was used in about three quarters of the taxometric studies published prior to the review of Ruscio et al. (2006). It may surprise some readers to learn that, until very recently, the performance of this consistency test had never been studied. The small-scale study of Schmidt et al. (2004), mentioned earlier, involved 20 taxonic and 10 dimensional samples and did not vary data conditions or taxometric procedures. Ruscio et al. presented a more extensive Monte Carlo study of consistency tests based on MAXCOV results that included the BRCT. In this study, 10,000 samples of data (5,000 taxonic and 5,000 dimensional) were generated using a random sampling design by which data parameters varied within ranges considered acceptable for taxometric analysis: sample size ($N = 300$ to 1,000), taxon base rate ($P = .10$ to .50), indicator validity ($d = 1.25$ to 2.00), within-group correlations ($r = .00$ to .30), indicator skew ($S = 0$, 1, or 2), and number of indicators ($k = 3$ to 8); all values were drawn at random from uniform distributions. For dimensional data, indicator correlations were set to the expected value for the randomly sampled configuration of P, d, and r using a formula adapted from Meehl and Yonce (1994, 1996).

Each sample was analyzed using MAXCOV, and the *SD* of the base rate estimates was recorded. To determine the extent to which this BRCT distinguished taxonic and dimensional data, receiver operating characteristic (ROC) analyses were performed. The area under the ROC curve (A) represents the probability that a taxonic sample chosen at random yields a lower *SD* of base rate estimates than a dimensional sample chosen at random. This assesses the validity of the BRCT independent of the threshold that might be used to reach conclusions,

in which $A = 1.00$ indicates perfect discrimination and $A = .50$ indicates chance-level discrimination. For the BRCT, Ruscio et al. (2006) found that $A = .66$, a modest improvement over chance and far from the impressive validity that one might infer from the BRCT's ubiquitous endorsement in the literature. For comparison, the same study found $A = .70$ for the GFI and $A = .60$ for the proportion of Bayesian probabilities at the extremes.

In a study using a very similar design, Ruscio et al. (2007) found $A = .38$ for the BRCT, which is worse than chance-level performance. By comparison, $A = .74$ for the GFI, .61 for the proportion of Bayesian probabilities at the extremes, and .93 for the CCFI. The strong discrimination achieved by the CCFI suggests that the poor performance of the other tests cannot be attributed to overly challenging data conditions or a poorly implemented analysis plan. The discrepancies in results between this and the previous study stem from two differences in their designs: MAXEIG, rather than MAXCOV, was used, and a wider range of indicator skew values was included. It seems that the BRCT fares especially poorly with skewed indicators, which is consistent with Ruscio et al.'s (2004) observation that dimensional data can produce cusped curves and highly consistent base rate estimates when indicators are skewed.

The present study involves a crossed factorial design in which a broad range of data conditions is covered and a much larger number of samples are generated and analyzed. The factors and levels are as follows:

1. Latent structure—2 levels (taxonic, dimensional). The taxonic model includes two groups, and the dimensional model includes a single factor.
2. Sample size—3 levels ($N = 300$, 600, 1,000). These values include Meehl's (1995) recommended minimum of $N = 300$, a value approximating the median ($N = 585$) in the 66 taxometric studies reviewed by Haslam and Kim (2002), and a still larger value of $N = 1{,}000$.
3. Taxon base rate—4 levels ($P = .50$, .25, .10, and .05). These values include the three base rates studied by Meehl and Yonce (1994, 1996; $P = .50$, .25, and .10) plus an even lower value.
4. Indicator validity—5 levels ($d = 2.00$, 1.75, 1.50, 1.25, and 1.00). These values range from the large validity used in most of Meehl and Yonce's (1994, 1996) conditions ($d = 2.00$), through Meehl's (1995) recommended minimum of $d = 1.25$, to an even smaller value ($d = 1.00$) that should seriously challenge taxometric procedures.
5. Within-group correlation—4 levels ($r = .00$, .10, .20, and .30). These values range from the literal correctness of the taxonic model ($r = .00$) to Meehl's (1995) suggested upper limit to the robustness of taxometric procedures ($r = .30$).
6. Indicator skew—3 levels ($S = 0$, 1, and 2). Although most published simulation studies of taxometric analyses have used normal indicator distributions, research data often deviate from normality. Micceri (1989) showed that even rigorously developed psychometric measures and achievement tests yielded asymmetric score distributions, frequently skewed at values on the order of 1 or 2 and sometimes even more. Indicators for the present study were drawn either from normal distributions (skew = 0) or χ^2 distributions with skew levels (1 or 2) often encountered in research data.
7. Number of indicators—2 levels ($k = 4$ and 6). These values include the number of indicators typical in Monte Carlo studies of taxometric procedures ($k = 4$; e.g., Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998) as well as a larger value ($k = 6$) that yields substantially more taxometric curves and may approach the limit of conceptually and empirically nonredundant indicators available in most taxometric studies—an essential data requirement for informative results (Cole, 2004; Ruscio et al., 2006; Ruscio & Ruscio, 2004a). Because k was hypothesized to have little effect, only two levels were used so that other factors believed to be more important could vary across more levels without yielding an overwhelming number of cells in the overall design. (In supplementary analyses performed after this study was completed, the addition of a condition with $k = 8$ indicators—which was fully crossed with all factors in the design—did not alter any of the results described below.)

Factorially varying these seven variables yields 2,880 cells in the experimental design. With 100 replication samples generated for each condition, a total of 288,000 data sets were analyzed. Submitting each sample to MAMBAC, MAXCOV, and MAXEIG, and estimating the base rate in two ways for each MAXEIG curve, yielded a total of 19,296,000 base rate estimates (with the extra $k = 8$ conditions, the total exceeds 30 million). By any measure, this is the most extensive study of taxometric analyses that has been performed to date. Interested readers can contact the author for a full report with methodological details and analyses of the accuracy of base rate estimates, which are omitted in this abbreviated report.

To test the performance of the BRCT, ROC analyses were performed within each condition. Three measures were obtained: (a) *A* was calculated to assess discriminating power independent of threshold, (b) the threshold that correctly classified the most samples was identified as the optimal threshold, and (c) the percentage of correct classifications at the optimal threshold was recorded.

To test for differences across conditions, a series of 3 (*N*) × 4 (*P*) × 5 (*d*) × 4 (*r*) × 3 (*S*) × 2 (*k*) ANOVAs was performed, using *A* as the dependent variable. There were a total of five such ANOVAs, one for each of the four procedures (MAMBAC, MAXCOV, and the two MAXEIG techniques) plus one for a cross-procedure BRCT consisting of the *SD* of the *M* estimates of the individual taxometric procedures. Rather than relying on tests of statistical significance with the extraordinarily large sample size, ω^2 was used to identify effects of nontrivial magnitude; effects with $\omega^2 \geq .01$ are shown in Table 4.

Sample size and the number of indicators had relatively little influence on *A*. Table 4 reveals just one $\omega^2 \geq .01$ for a main effect of *N* and only 10 interaction terms involving *N* (of 75 tested), and all 11 $\omega^2 \leq .03$. The number of indicators had even less influence: There were no interactions involving *k* and only one small main effect ($\omega^2 = .01$; in MAMBAC analyses, *A* was a bit smaller with six indicators [$M = .32$] than with four [$M = .36$]). The bottom of Table 4 also shows that effects for *N* and *k* were much smaller than those for other factors, especially *P* and *S*. Because the effects of *N* and *k* were negligible, ROC analyses were rerun with data collapsed across levels of these factors. This yielded 600 samples apiece for taxonic and dimensional structure within each of the remaining 240 cells—4 (*P*) × 5 (*d*) × 4 (*r*) × 3 (*S*)—of the experimental design. The resulting values of *A* are plotted in Figure 7.

Across data conditions the BRCT was of variable, but generally low, discriminating power. For each procedure, *A* spanned the full range from 0 to 1 and was often at or less than chance level. MAMBAC yielded the lowest *A* values, and for 80% of conditions, MAMBAC produced results poorer than chance. Cross-procedure *SD*s performed more poorly than those for MAXCOV or MAXEIG but not quite as poorly as those for MAMBAC. Whereas MAXCOV and MAXEIG yielded BRCT results worse than chance for one quarter to one third of experimental conditions, the cross-procedure BRCT performed worse than chance in just more than one half of all conditions. For more data conditions than not, MAMBAC and cross-procedure BRCTs would have yielded more valid results if lower *SD*s of base rate estimates were interpreted as evidence of dimensional, rather than taxonic, structure.

Indicator skew markedly decreased *A* values. This factor, alone or through its interaction with other factors, explained an average of 50% of the variance in *A* values. Figure 7 reveals that larger *S* reduced *A* almost without exception across conditions. Often the decline was substantial. Across all estimation procedures, *A*s averaged .72, .54, and .40 for $S = 0$, 1, and 2, respectively.

TABLE 4: Effect Sizes (w2) for ANOVAs Performed on the Areas Under ROC Curves for the Base Rate Consistency Test

Effect	*MAMBAC* SD	*MAXCOV* SD	*MAXEIG (Hitmax)* SD	*MAXEIG (GCMT)* SD	*Cross-Procedure* SD	*Row* M
N	.02					
P	.23	.25	.36	.31	.07	
d	.01	.02		.02	.14	
r		.05	.08	.12	.08	
S	.50	.24	.24	.23	.14	
k	.01					
N × *P*	.02	.03			.02	
N × *S*		.02			.02	
P × *d*	.03	.04	.04	.05	.03	
P × *r*	.02			.01	.01	
P × *S*	.02	.21	.15	.15	.27	
d × *r*					.03	
d × *S*		.02	.02	.02	.03	
r × *S*	.02			.02	.02	
N × *P* × *S*	.01	.02			.02	
P × *d* × *S*	.01				.02	
P × *r* × *S*	.02	.01		.01	.02	
Sum Across All Main Effects and Interactions With $\omega^2 \geq .01$						
N	.05	.07			.06	.04
P	.33	.56	.55	.53	.46	.49
d	.05	.08	.06	.09	.25	.09
r	.06	.06	.08	.16	.16	.10
S	.58	.52	.41	.43	.54	.50
k	.01					.00

Note. ANOVAs were performed using 1,440 areas under ROC curves across experimental conditions. Effects that yielded $\omega^2 < .01$ for all four procedures are omitted from the table, as are ω^2 values < .01 for effects that are listed. *N* = sample size (300, 600, 1,000); *P* = taxon base rate (.50, .25, .10, .05); *d* = indicator validity (1.00, 1.25, 1.50, 1.75, 2.00); *r* = within-group correlation (.00, .10, .20, .30); *S* = skew (0, 1, 2); *k* = number of indicators (4, 6). ROC = receiver operating characteristic; MAMBAC = mean above minus below a cut; MAXCOV = maximum covariance; MAXEIG = maximum eigenvalue.

The influence of taxon base rate rivaled that of indictor skew. For each procedure, *A*s were larger when *P* was smaller. Averaged across estimation techniques, *A* averaged .36, .49, .68, and .68 at *P* = .50, .25, .10, and .05, respectively. Not only was the main effect for *P* large, but so was its interaction with *S*. Although larger *S* reduced *A* for all levels of *P*, it did so to a lesser extent with smaller *P*.

The results also show that, with one exception, larger indicator validity increased *A*. The exception is that when *P* = .50 or .25, MAMBAC yielded lower *A*s with increasing values of *d*. Without exception, smaller within-group correlations increased *A*. Neither effect was especially large: The averaged total effect sizes for *d* (.09) and *r* (.10) were much smaller than those for *S* (.50) and *P* (.49).

One final issue crucial to applications of the BRCT is what threshold best distinguishes taxonic and dimensional data. Schmidt et al. (2004) are the only authors to offer a specific recommendation. The present study addresses this issue across a wide range of data conditions. The distribution of optimal thresholds for each cell in the design (after collapsing across levels of *N* and *k*) was examined. Schmidt et al.'s recommended threshold of .10 for MAXCOV is near the low end of the distribution for this procedure. For each procedure,

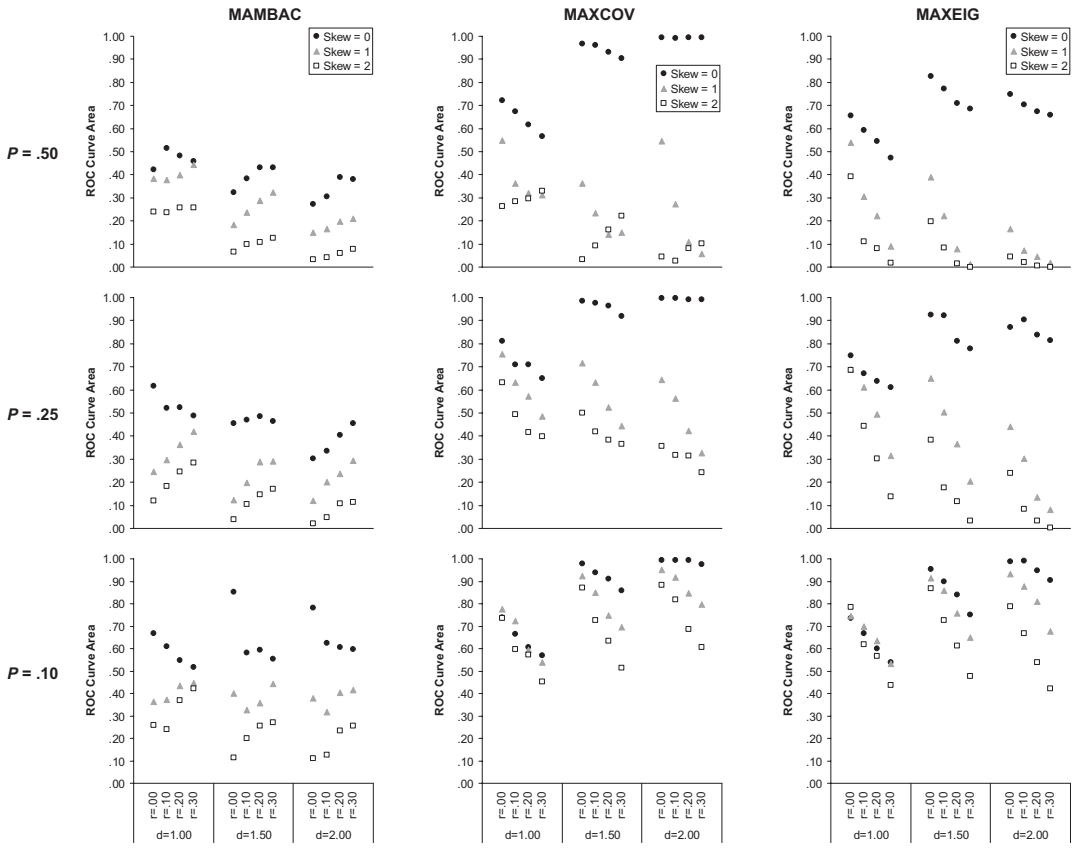

Figure 7: Areas Under the ROC Curves That Represent the Discriminating Power of the Base Rate Consistency Test

Note. Chance-level accuracy corresponds to *A* = .50, the middle value on the *y* axis of each graph. Each data point represents the area under the ROC curve calculated for one experimental condition, collapsed across levels of *N*and *k*. The *x* axes contain configurations of indicator validity (*d* = 1.00, 1.50, 2.00) and within-group correlation (*r* = .00, .10, .20, .30). Results for *P* = .50, .25, and .10 are shown; results for *P* = .05 were similar those for *P* = .10.MAMBAC = mean above minus below a cut; MAXCOV = maximum covariance; MAXEIG = maximum eigenvalue; ROC = receiver operating characteristic.

optimal thresholds varied widely. No threshold appears to be generally applicable across data conditions or taxometric procedures.

The distributions of optimal thresholds include those for many conditions under which the BRCT performed poorly. This leaves open the possibility that for one or more procedures, there may exist a threshold applicable across the subset of conditions for which discriminating power was good. To test this possibility, conditions in which each procedure correctly classified at least 90% of data sets were identified and their optimal thresholds were examined. The percentage of conditions that reached this level of accuracy and a summary of the optimal thresholds under those conditions are as follows: MAMBAC, < 1% of conditions (1 out of 240), optimal threshold = .15; MAXCOV, 17% of conditions, *M* = .17, *SD* = .05; MAXEIG (hitmax), 10% of conditions, *M* = .08, *SD* = .01; MAXEIG (GCMT), 10% of conditions,

$M = .06$, $SD = .01$; cross-procedure, 3% of conditions, $M = .09$, $SD = .02$. The conditions included in these evaluations were chiefly those with $S = 0$ and low P; to a lesser extent, large d and small r were associated with good performance. Even within this select subset of conditions, there was sufficient variability across procedures to suggest that no single threshold is generally applicable. The results of this study strengthen and extend the central conclusion of previous investigations of the BRCT: It seldom works very well, and it would be difficult to justify its use on empirical grounds. Possible exceptions would include situations in which data conditions are highly favorable (e.g., nonskewed indicators, small taxon base rate, large indicator validity, small within-group correlations). However, under these conditions, taxometric curves usually are easy to interpret correctly, leaving little room for incremental validity of the BRCT. The next study deals with this issue more directly.

MONTE CARLO STUDY 2: CONSISTENCY TESTING AND INCREMENTAL VALIDITY

Although the previous study used a crossed factorial design to address a relatively focused set of questions about one particular consistency test, this study used a random-sampling design to explore a broader range of questions about the general practice of consistency testing. To what extent do various taxometric procedures or consistency tests differentiate taxonic and dimensional data? More important, to what extent do these tests provide incremental validity when used in conjunction with one another? Given the emphasis placed on the virtues of consistency testing in the taxometric literature, it is surprising that only one published study has examined incremental validity. Cleland et al. (2000) found that when MAMBAC and MAXCOV suggested the same structural conclusion (which occurred in 78% of the analyses), this was correct 90% of the time. Because their study required the visual inspection of taxometric graphs, its design was limited to a small number of data conditions. The present study includes a wider range of taxometric tests and data conditions, but it still constitutes only an initial foray into territory that will require a concerted research effort to understand well.

The design of this study is similar to those of Ruscio et al. (2006, in press), with an additional factor and more samples included. Twenty-five thousand samples of data (12,500 taxonic and 12,500 dimensional) were generated using a random sampling design with the following variables: sample size ($N = 300$ to 1,000), taxon base rate ($P = .10$ to .50), indicator validity ($d = 1.25$ to 2.00), within-group correlations ($r = .00$ to .30), indicator skew ($S = 0$ to 6, whole numbers only), number of indicators ($k = 3$ to 8), and indicator distributions (6, 9, 12, or 15 ordered categories, or continuous scales); all values were drawn at random from uniform distributions, with the exception of S, for which lower values were drawn with greater probability. Having drawn 12,500 sets of data parameters to generate the samples using a taxonic model, these same 12,500 sets of parameters were used to generate the data using a dimensional model; indicator correlations were set to the expected value for the configuration of P, d, and r. Each sample was analyzed using MAMBAC and MAXEIG, and for each procedure, 10 samples apiece of taxonic and dimensional comparison data were used to calculate a CCFI value. Several other tests were included in the study: the BRCT, the GFI, and the proportion of extreme Bayesian probabilities (PBayes). The BRCT was performed separately using MAMBAC and MAXEIG results, and the GFI and PBayes tests

TABLE 5: Percentage of Samples Classified Correctly as Taxonic or Dimensional

Taxometric Test	*Taxonicity Threshold*	*All Data (N = 25,000)*	*Taxonic Data (N = 12,500)*	*Dimensional Data (N = 12,500)*
Recommended thresholds				
CCFI-M	≥ .50	96.6	99.4	93.8
CCFI-X	≥ .50	83.3	76.8	89.8
BRCT-M	≤ .10	51.1	79.1	23.1
BRCT-X	≤ .10	50.6	79.5	21.7
GFI	≥ .90	68.8	83.1	54.4
PBayes	≥ .80	49.1	38.2	60.0
Sample-optimized thresholds				
CCFI-M	≥ .495	96.6	99.6	93.6
CCFI-X	≥ .444	85.5	89.0	82.0
BRCT-M	≤ .039	61.6	52.9	70.2
BRCT-X	≤ .157	54.0	95.4	12.7
GFI	≥ .931	70.1	68.9	71.4
PBayes	≥ .544	65.6	93.3	37.9

Note. CCFI-M = comparison curve fit index, MAMBAC, CCFI-X = comparison curve fit index, MAXEIG; BRCT-M = base rate consistency test; MAMBAC; BRCT-X = base rate consistency test, MAXEIG ; GFI = goodness of fit index; PBayes = proportion of Bayesian probabilities in extreme deciles (< .10 or > .90); MAMBAC = mean above minus below a cut; MAXEIG = maximum eigenvalue.

were performed using the MAXEIG results. In all, six tests were included: CCFI-M (for MAMBAC), CCFI-X (for MAXEIG), BRCT-M, BRCT-X, GFI, and PBayes.

The results were analyzed in several ways. To begin, an ROC analysis was performed for each test to determine how well it distinguished the 12,500 taxonic and 12,500 dimensional samples, with *A* the measure of discriminating power. Four tiers of results emerged: (a) the CCFI-M (*A* = .99) and CCFI-X (*A* = .94) performed very well, (b) the GFI (*A* = .75) performed moderately well, (c) the BRCT-M (*A* = .63) and PBayes (*A* = .60) surpassed chance by narrower margins, and (d) the BRCT-X (*A* = .36) fared worse than chance. These results are consistent with those reported earlier, with the exception that CCFI-M had not been studied previously—and it stands out as an exceptionally good test. The percentage of samples classified correctly as taxonic or dimensional is shown in Table 5, first for a threshold recommended in the literature and then for the sample-optimized threshold determined in the ROC analysis. These results suggest that only CCFI-M achieves an exemplary classification accuracy, with CCFI-X in second place and the other tests far behind. Even with the application of a sample-optimized threshold, none of the other tests approaches the validity of either CCFI.

What makes this study novel is the exploration of incremental validity: Analyses examined the extent to which classification accuracy would improve by using more than one test. To avoid potential capitalization on chance, only the recommended (rather than sample-optimized) thresholds were used in these analyses. In the 80.7% of cases when the two CCFIs suggested the same conclusion, this was correct 99.0% of the time. The fact that taxometric curve shapes, when interpreted using an objective index based on simulated comparison data, so seldom led to mistaken conclusions means that there may be little room for improvement using additional consistency tests.

Given the results in Table 5, the only other test in this study that one might consider including, along with one or both CCFIs, would be the GFI. Using the CCFI-M plus the GFI would afford 98.0% accuracy but only for the 68.1% of cases when both tests suggested the same structural conclusion; this is inferior to the CCFI-M plus CCFI-X combination in that it is less

accurate even when a larger proportion of cases are left undecided. Using the GFI in addition to both CCFI values would yield a slightly more complicated array of results. If one drew conclusions only for the 58.4% of cases when all three tests agreed, accuracy would be 99.4%; if one drew conclusions for all samples by following the majority rule of the tests' indications, accuracy would be 91.3%. The latter option is inferior to using the CCFI-M test by itself and the former option increases an extremely high accuracy (99.0% for the CCFI-M plus the CCFI-X) by a small amount (to 99.4%) at the expense of withholding judgment in a much larger proportion of cases (41.6% rather than 19.3%). Considering the desire of researchers to draw conclusions from their studies, these data suggest that one should either use CCFI-M exclusively or use both the CCFI-M and the CCFI-X. Regardless of how one navigates this trade-off, the results of this study reveal the poor incremental validity provided by the other four tests included in this study. The GFI achieved some increment in validity—but the cost was a substantial proportion of inconclusive results.

No combinations of tests that did not include either CCFI value approach their level of performance. To provide a clear sense for how poorly the remaining tests performed in comparison to the CCFIs, consider the results of a logistic regression analysis in which all four of the other tests were entered to predict taxonic versus dimensional structure. Such an analysis presents the results very favorably because it assigns sample-optimized weights to each test and then determines an optimal threshold for classification. Even allowing capitalization on chance to assist the tests in these ways, the logistic regression analysis correctly classified only 73.3% of all samples.

CONCLUDING THOUGHTS

In everyday discourse, the language of qualitative distinctions (e.g., whether a defendant is "a psychopath") is usually less cumbersome than the language of quantitative differences (e.g., whether a defendant is "an individual high in psychopathic traits"). However, when a construct is continuous in nature, using simple but imprecise phrasing may reinforce legal decision makers' tendency to think in terms of categories. Several studies have investigated psychopathy using the taxometric method, and most suggest that its latent structure is dimensional (Edens et al., 2006; Guay et al., in press; Marcus et al., 2004). This suggests that forensic assessment should locate a defendant in a multidimensional space defined by continuously varying psychopathic traits, rather than identifying an individual as psychopathic or not, and that decision makers should judge whether this constellation of traits meets an applicable legal criterion, such as "dangerous and severe personality disorder," rather than relying on an artificially categorical assessment procedure that may or may not address the legal question at hand. Taxometric research could inform the science and practice of criminal justice by clarifying the latent structure of many other constructs that appear to be presumed taxonic in nature. For example, describing some individuals as "sexually violent predators" or "incompetent to stand trial" implies a qualitative distinction, as does the application of a quasi-diagnostic label such as "battered women syndrome."

This article has focused on conceptual issues such as how a taxometric examination of a construct's latent structure might be used to address questions in basic and applied science and the inferential framework by which conclusions are drawn from taxometric analyses. Data-analytic tools within the taxometric method were reviewed, with attention to the decisions

that must be made to implement each as well as consideration of the empirical status of the taxometric procedures and consistency tests. In the space of a single journal article, it is impossible to provide sufficient information to enable researchers to become sophisticated data analysts. Readers interested in conducting taxometric investigations are encouraged to consult additional resources for further details, especially with regard to the data requirements of the method (e.g., Meehl, 1995; Meehl & Yonce, 1994, 1996; Ruscio et al., 2006). Programs for performing taxometric analyses, including the simulation of comparison data, can be downloaded along with a detailed user's manual at http://www.taxometricmethod.com.

A pair of new Monte Carlo studies was presented not only to tackle important methodological issues (e.g., the utility of the BRCT, the validity of structural inferences drawn from consistency tests used alone or in combination) but also to demonstrate approaches to doing so rigorously and empirically. Whether one generates and analyzes taxonic and dimensional comparison data in an individual taxometric study or performs a Monte Carlo study with a crossed factorial or random sampling design, there are many ways to help place the taxometric method on a more solid evidential foundation. For too long, researchers using the method have been guided largely by tradition, anecdote, and untested opinion. It is time to adopt more fully an empirically grounded approach.

REFERENCES

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.

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

Beauchaine, T. P., & Beauchaine, R. J. (2002). A comparison of maximum covariance and *k*-means cluster analysis in classifying cases into known taxon groups. *Psychological Methods*, *7*, 245-261.

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

Cole, D. A. (2004). Taxometrics in psychopathology research: An introduction to some of the procedures and related methodological issues. *Journal of Abnormal Psychology*, *113*, 3-9.

Edens, J. F., Marcus, D. K., Lilienfeld, S. O., & Poythress, N. G., Jr. (2006). Psychopathic, not psychopath: Taxometric evidence for the dimensional structure of psychopathy. *Journal of Abnormal Psychology*, *115*, 131-144.

Golden, R. R. (1982). A taxometric model for the detection of a conjectured latent taxon. *Multivariate Behavioral Research*, *17*, 389-416.

Grove, W. M. (2004). The MAXSLOPE taxometric procedure: Mathematical derivation, parameter estimation, consistency tests. *Psychological Reports*, *95*, 517-550.

Guay, J., Ruscio, J., Hare, R., & Knight, R. A. (in press). A taxometric study of the latent structure of psychopathy: Evidence for dimensionality. *Journal of Abnormal Psychology*.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1994). Psychopathy as a taxon: Evidence that psychopaths are a discrete class. *Journal of Consulting and Clinical Psychology*, *62*, 387-397.

Haslam, N., & Cleland, C. (2002). Taxometric analysis of fuzzy categories: A Monte Carlo study. *Psychological Reports*, *90*, 401-404.

Haslam, N., & Kim, H. C. (2002). Categories and continua: A review of taxometric research. *Genetic, Social and General Psychology Monographs*, *128*, 271-320.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8.51 user's manual*. Lincolnwood, IL: Scientific Software.

Krueger, R. F., Watson, D., & Barlow, D. H. (2005). Introduction to the special section: Toward a dimensionally based taxonomy of psychopathology. *Journal of Abnormal Psychology*, *114*, 491-493.

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.

Marcus, D. K., John, S. L., & Edens, J. F. (2004). A taxometric analysis of psychopathic personality. *Journal of Abnormal Psychology*, *113*, 626-635.

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. (2001). Primary and secondary hypohedonia. *Journal of Abnormal Psychology*, *110*, 188-193.

Meehl, P. E. (2004). What's in a taxon? *Journal of Abnormal Psychology*, *113*, 39-43.

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: John Wiley.

Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*, *74*, 1059-1274.

Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports*, *78*, 1091-1227.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156-166.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175-220.

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., Haslam, N., & Ruscio, A. M. (2006). Introduction to the taxometric method: A practical guide. Mahwah, NJ: Lawrence Erlbaum.

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. (2004a). 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. (2004b). A nontechnical introduction to the taxometric method. *Understanding Statistics*, *3*, 151-193.

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.

Schmidt, N. B., Kotov, R., & Joiner, T. E., Jr. (2004). *Taxometrics: Toward a new diagnostic scheme for psychopathology*. Washington, DC: American Psychological Association.

Skilling, T. A., Quinsey, V. L., & Craig, W. M. (2001). Evidence of a taxon underlying serious antisocial behavior in boys. *Criminal Justice and Behavior*, *28*, 450-470.

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