Journal of Abnormal Psychology
2004, Vol. 113, No. 1, 24–38

Copyright 2004 by the American Psychological Association, Inc.
0021-843X/04/$12.00 DOI: 10.1037/0021-843X.113.1.24

Clarifying Boundary Issues in Psychopathology: The Role of Taxometrics in a Comprehensive Program of Structural Research

John Ruscio
Elizabethtown College

Ayelet Meron Ruscio
The Pennsylvania State University

Despite decades of debate, important questions about the boundaries that separate psychological disorder from normality and that distinguish 1 disorder from another remain largely unanswered. These issues pose empirical questions that may be addressed by assessing the latent structure of psychopathological constructs. Because these constructs are likely to be structurally complex, and no single statistical tool addresses all structural questions, it is proposed in this article that boundary issues be examined through programmatic research grounded in the taxometric method and elaborated by complementary analyses. The authors describe how such a program could delimit the structure of disorders and test competing explanations of diagnostic co-occurrence, emphasizing the potential to enhance the reliability and validity of assessment, maximize the power of research designs, and improve diagnostic classification.

From the earliest attempts to classify psychopathology, mental health professionals have struggled to understand, organize, and represent the pathological entities that underlie presentations of mental disorder. Although our nosological systems have grown increasingly sophisticated, important questions remain about the fundamental nature and structure of these entities, particularly with respect to the boundaries that define them. Such questions about boundaries usually take one of two forms. First, what is the nature of the boundary that separates disorder from normality? Second, what is the nature of the boundary that distinguishes one disorder from another? Both of these questions are central to our understanding of psychopathology and have stimulated considerable debate within the mental health community, yet both are still largely unanswered.

These empirical questions are most appropriately answered by exploring the structure underlying surface-level symptoms or features of mental disorders. We describe the possible varieties of latent structure and their implications for the nature of psychopathology. Because the latent structure of disorders may be highly complex, we suggest that boundary issues may be addressed most powerfully through the systematic application of multiple, complementary analytic approaches designed to delineate latent structure. At the same time, because unique features of the taxometric method make it particularly well suited to answer the most basic boundary questions, we emphasize the role that taxometric studies can play within a comprehensive program of structural investigation.

The Variety and Complexity of Potential Latent Structures

Latent Versus Manifest Levels of Analysis

An informed discussion of boundary issues usefully begins by recognizing the important distinction between latent and manifest structure. *Latent structure* refers to the fundamental nature of a construct, the underlying categorical or continuous structure that exists regardless of how we might choose to measure the construct. In contrast, *manifest structure* refers to characteristics of observable measures of a construct, the surface structure that depends—among other things—on how the construct is assessed. Observable scores, such as psychological test results or diagnostic interview ratings, provide information at the manifest level of analysis, and we use these empirical data to draw inferences about hypothetical constructs (MacCorquodale & Meehl, 1948) at the latent level.

Because membership in latent classes and positions along latent continua influence, but do not completely determine, individuals' scores on manifest variables, structure may be different at the latent and manifest levels of analysis. For example, a distribution of manifest scores in a sample consisting exclusively of two homogeneous latent classes will often appear continuous, unimodal, and even bell-shaped because the separation between groups does not offset the measurement error within them. Figure 1 illustrates a case in which two latent classes differ substantially from one another (2 within-group *SD*) yet produce a joint distribution appearing fairly normal in shape. Indeed, for a joint distribution to show the first hint of bimodality (a slight dip toward the center of the distribution), equal-sized groups must be separated by more than 2 *SD;* for unequal-sized groups, the separation must be even higher (Murphy, 1964). Thus, latent classes can easily give rise to manifest continua.

The reverse is also true: Latent continua may give rise to manifest types through threshold effects, sampling error (particu-

John Ruscio, Department of Psychology, Elizabethtown College; Ayelet Meron Ruscio, Department of Psychology, The Pennsylvania State University.

This article was supported by National Research Service Award MH12675 to Ayelet Meron Ruscio from the National Institute of Mental Health and by summer research support to John Ruscio from Elizabethtown College. We are grateful to T. D. Borkovec, David A. Cole, Nick Haslam, and Scott O. Lilienfeld, for their valuable suggestions on previous drafts of this article.

Correspondence concerning this article should be addressed to John Ruscio, Department of Psychology, Elizabethtown College, Elizabethtown, PA 17022. E-mail: rusciojp@etown.edu

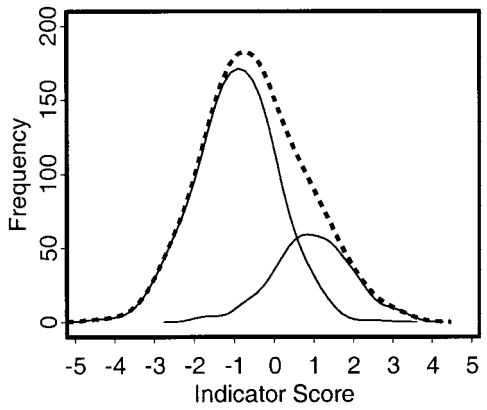

Figure 1. Frequency distributions for a manifest indicator of two latent taxa (solid lines) and their joint distribution (dashed line). The smaller group is one-third the size of the larger, and the groups are separated by 2 within-group *SD*.

larly when samples are small), selective sampling from the extremes of a continuum, and a variety of other causes. For example, Haslam (1997) suggested that male sexual preferences may be continuous at the latent level but appear polarized at the manifest level because of norms against reporting nonexclusive sexual patterns. Thus, there are a great many ways by which latent types may give rise to an observed continuum and by which a latent continuum may give rise to apparent types (Grayson, 1987; Haslam, 1999; Murphy, 1964). Because boundaries are theoretically meaningful only at the latent level of analysis, it is essential that psychopathology researchers empirically determine the latent structure of their constructs rather than presume this structure on the basis of manifest distributions.

In addition to distinguishing between latent and manifest structure, theories may postulate constructs at multiple latent levels of analysis (Meehl, 2001b). For example, in his etiological theory of schizophrenia, Meehl (1962, 1990) hypothesized that the vulnerability to schizophrenia involves the interplay of entities and processes at four distinct latent levels. Closest to the manifest level is primary hypohedonia, characterized by behavioral and phenomenological signs and symptoms of underlying deficits. At a more latent level is secondary hypohedonia, an impaired hedonic capacity associated with central nervous system dysfunction. Still deeper levels include the biological basis of secondary hypohedonia and, ultimately, the conjectured genetic-specific etiology. Similarly, Haslam and Kim (2002) distinguished between two latent levels of analysis in their attempt to reconcile apparently discrepant findings concerning the structure of male sexual orientation. They argued that Haslam (1997), who obtained evidence for a latent continuum of sexual orientation using relatively unobtrusive items with low face validity, tapped the construct at a deeper level (closer to the causal origins of sexual orientation) than a study by Gangestad, Bailey, and Martin (2000), whose more face-valid measures may have polarized responses such that they suggested the existence of two latent classes by cuing social norms against bisexuality. To simplify our discussion, we focus on only one latent level of analysis for the remainder of the present article. However, we recommend that researchers consider both relevant theory and the nature of their variables to determine the latent level that is targeted by their analyses and to contextualize the ensuing results appropriately.

Dimensions Versus Taxa

Thus far we have used colloquial language to describe different kinds of latent structures. We now introduce the terminology that is used most often in the growing structural literature. Psychological constructs have traditionally been regarded as either dimensional (continuous, quantitative, latent factor, latent trait) or taxonic (categorical, discrete, qualitative, latent class). The term *dimensional* refers to constructs along which individuals differ only quantitatively, such that any groups that might be formed are arbitrary. Clear examples of dimensional constructs include barometric pressure and temperature, which can be scaled along the continua of millimeters of mercury and degrees Centigrade, respectively. High-pressure weather and hot objects are not naturally occurring taxa but are instead distinctions superimposed on dimensions for pragmatic reasons (J. Ruscio & Ruscio, 2002).

The term *taxonic* refers to constructs in which individuals or objects are separated into nonarbitrary groups at the latent level. That is, one or more latent boundaries exist such that individuals either belong or do not belong to each group; in the two-group case, these are often referred to as the *taxon* (e.g., disordered group) and its *complement* (e.g., nondisordered group). We conceptualize these boundaries as Kendell (1975) does, but whereas he differentiates classes by discontinuities along manifest distributions of scores, we do so by discontinuities at the latent level. Although taxa may produce overlapping score distributions on manifest variables and classification errors are inevitable in practice, none of this implies that the taxa themselves are not genuinely discrete.

The precise meaning of taxonic latent structure has proven very difficult to specify in conceptual terms. Hence, researchers often implicitly define taxon in terms of the mathematical model that they use to evaluate latent structure. The definition that we have adopted is grounded in the General Covariance Mixture Theorem (GCMT)—the structural model introduced in Meehl's (1962) first technical report on the taxometric method and articulated in numerous subsequent publications.[1] According to this model, two variables that are valid indicators of a latent dimension will covary at a fairly constant level across subsamples of cases that have been ordered along a third indicator variable, whereas two valid indicators of a taxon will be associated with one another differentially as the proportion of taxon and complement members changes across ordered subsamples. Thus, dimensional structure denotes a single group of individuals whose manifest scores covary because of shared loadings on a common latent factor, whereas taxonic structure denotes distinct groups whose manifest scores covary more strongly in a mixed sample than within either group.

At present, few indisputably taxonic constructs have been identified in psychology, though other sciences can provide less contentious examples. For example, molecules, chemical elements and isotopes, and subatomic particles all represent classes of objects that differ from one another in a categorical fashion (e.g., glucose vs. ammonia, hydrogen vs. carbon, proton vs. electron). However, other taxonic constructs are somewhat fuzzy conceptually because

[1]For readers interested in a more thorough treatment of the challenges of defining taxonic latent structure and of the important relation between this definition and corresponding mathematical models, we recommend the insightful discussions in Meehl (1992, 1995, 1999) and Meehl and Golden (1982).

of definitional difficulties (e.g., biological species are notoriously difficult to specify with perfect precision) or are challenged by exceptional cases that cannot be classified neatly (e.g., human biological sex is complicated by androgynous births). Matters may be complicated further by the fact that taxa themselves may not represent entirely homogeneous classes of individuals. Instead, there may be meaningful variation within groups. For example, if schizophrenia were found to be taxonic, this would not preclude substantial differences in symptomatology, severity, or other characteristics among schizophrenic individuals or among nonschizophrenic individuals. Thus, as we discuss further below, it is possible to conceive of additional construct-relevant dimensional variation superimposed on taxonic structure. Moreover, it should be obvious that members of a taxon can differ from one another on factors unrelated to the target construct. For example, within the relatively discrete taxa corresponding to men and women, individuals differ from each other on a host of variables other than biological sex. In sum, the presence of taxonic latent structure should not be taken to suggest that all taxon members are identical to one another, nor that latent structures more complex than two taxa or a single dimension are impossible (J. Ruscio & Ruscio, 2002, in press).

In a domain as complex and multidetermined as psychopathology, it seems likely that etiological factors may combine and interact (via catalytic, synergistic, or other moderated effects) in elaborate ways to produce many or even most mental disorders. For example, consider the potentially complex latent structure of social anxiety. First, social anxiety may consist of a single latent dimension along which individuals vary in severity. Second, a pathological social anxiety taxon, or social phobia, may be distinct from a complement class of normal shyness. Third, complexity may increase through the combination of taxonic boundaries with dimensional variation; that is, within any taxon, there may be construct-relevant variation among individuals that is not due to measurement error. Social phobia may be taxonically distinct from normal shyness, with dimensional variation remaining among the socially anxious, the normally shy, or both. Fourth, the potential for a hierarchical arrangement of higher- and lower-order taxa or dimensions within a construct (e.g., a higher order social phobia taxon subsuming lower-order generalized and nongeneralized taxonic subtypes) further increases the possible complexity of latent structures.

Whereas some psychologists have criticized psychiatric nosology for classifying mental disorders according to exclusively categorical models in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [*DSM–IV*]; American Psychiatric Association, 1994; e.g., Carson, 1991, 1996; Widiger & Clark, 2000), others have been critical of psychologists' tendency to presume that psychopathological constructs are uniformly dimensional (e.g., Dahlstrom, 1995; Gangestad & Snyder, 1985; Meehl, 1992, 1995). Blanket statements about latent structure stemming from experiential impressions, disciplinary preferences, or other a priori notions may undermine rather than foster understanding of the nature of psychopathology. Instead, advances may be most likely to ensue from programmatic empirical research specifically designed to test the latent structure of important constructs in the domain of psychopathology.

Why Study the Latent Structure of Psychopathology?

Having argued that the latent structure of mental disorders is best determined through empirical investigation, we now turn to the question of why such investigation is worth pursuing. In this section, we briefly outline the implications of structural knowledge for some of the major tasks facing mental health professionals: performing psychological assessments, selecting appropriate research designs, and classifying mental disorders.

Psychological Assessment

The reliability and validity of assessment may be improved by matching measurement models to the latent structure of target constructs. Forcing dimensional variation into spurious taxa can discard valuable information (Cohen, 1983; MacCallum, Zhang, Preacher, & Rucker, 2002) whereas scaling taxonic constructs along a dimensional distribution[2] can increase measurement error and make it far more difficult to select an optimal cutting score when (for theoretical or practical purposes) a categorical distinction must be made (J. Ruscio & Ruscio, 2002). Thus, knowledge of a construct's latent structure should guide the selection of an appropriate assessment approach.

Classical test theory (Guilford, 1954; Gulliksen, 1950) and item response theory (IRT; Embretson, 1996; Hambleton, Swaminathan, & Rogers, 1991; Lord, 1980) are by far the most widely used measurement models in psychology because of their familiarity, power, and utility. However, these models work best with constructs that meet the assumption of dimensional latent structure. For taxonic constructs, categorical approaches such as Bayes's Theorem hold several advantages over dimensional models. For example, by estimating the probability that each individual is a member of the taxon (or complement), Bayes's Theorem facilitates the location of an optimal cut score along the distribution of probability values for assigning cases to classes. Not only can Bayesian probabilities yield greater classification accuracy than dimensional measurements when the optimal cutting score is used, they also yield a bimodal score distribution that ensures robustness to suboptimal cuts (J. Ruscio & Ruscio, 2002) and that allows one to explicitly consider the relative costs of false-positive errors and false-negative errors (Swets, Dawes, & Monahan, 2000). Thus, the notion that dimensions retain more information (and are therefore always preferable to categories) may be overly simplistic. Instead, psychological assessment premised on a construct's latent structure may yield greater reliability, validity, and statistical power.[3] At the same time, research is needed to specify more precisely the conditions under which categorical measurements are more useful than dimensional scores, a task that has

[2] We refer here to traditional methods of obtaining dimensional scores, such as summing or averaging responses to a series of items or using IRT models to estimate values along a latent trait, which typically produce unimodal distributions regardless of the latent structure of the construct being assessed. In contrast, one can use Bayes's Theorem to generate a bimodal distribution of taxon membership probabilities. A bimodal distribution readily suggests an optimal cutting score and is more robust to suboptimal choices.

[3] It is also worth noting that although computerized adaptive testing (CAT; Embretson & Herschberger, 1999; Wainer et al., 1990) is frequently paired with IRT in the assessment literature, categorical as well as dimensional measurement models can take advantage of CAT to shorten testing, reduce fatigue, eliminate hand-scoring errors, and provide immediate results (see J. Ruscio & Ruscio, 2002, for a description of how CAT may be used to maximize classification accuracy of taxonic constructs).

proved conceptually and practically challenging (Grove, 1991b; MacCallum et al., 2002).

Research Design

Information about the latent structure of a construct can also facilitate the selection of suitable research designs to further study that construct. Taxonic disorders may be most appropriately studied using dispositional group comparison designs. For example, if social phobia can be taxonically distinguished from normal shyness, one could sample individuals from these two groups and then compare them on other variables of interest. If a correlational design were used in the absence of dimensional variation, individuals in the disordered and nondisordered groups would be blended into a single, spurious, heterogeneous continuum of social anxiety. Much of the variability along this continuum might represent measurement error around the mean scores of socially phobic and normally shy individuals, thereby reducing statistical power and increasing the likelihood that important effects will be missed (Cohen, 1983; MacCallum et al., 2002).

In contrast, dimensional disorders may be most appropriately studied using correlational designs that include individuals at all severity levels of the construct. These research designs will be considerably more powerful for dimensional constructs than will comparative designs in which the continuum of social anxiety severity is artificially divided into groups (Cohen, 1983; MacCallum et al., 2002). Although statistical power may be increased when—in the interest of creating homogeneous and well-differentiated groups—researchers exclude individuals who exhibit moderate levels of symptom severity, selective sampling from the extremes of a continuum may limit our understanding of the construct and may obfuscate nonlinear relationships between symptom severity and other variables (A. M. Ruscio, Borkovec, & Ruscio, 2001). Finally, assigning individuals who vary along a latent dimension into groups can yield spurious statistical significance and a corresponding increase in Type I errors (Maxwell & Delaney, 1993).

For disorders with more complex latent structures, more complex research designs may be necessary. For example, if reliable dimensional variation is found within normally shy and socially phobic taxa, a combination of group comparison and correlational designs may be warranted. Careful sampling would be needed to ensure that participants are selected from both taxa as well from a range of severity levels within each. One (categorical) variable could be used to represent membership in the socially phobic taxon versus the normally shy complement, whereas a second (continuous) variable could be used to represent the severity of social anxiety. Analyses may consider these variables simultaneously or may begin by comparing means across taxa and proceed to correlational analyses within taxa. Thus, as with psychological assessment, the appropriate match of study design to latent structure may lead to increasingly informative research investigations that more rapidly advance knowledge of psychopathology.

Classification

The latent structure of a mental disorder may also have important implications for how it is classified and diagnosed. If a taxon is found, research can examine how closely it corresponds to prevailing diagnostic categories and consider adjustments to the diagnostic criteria that would enable them to better characterize the underlying construct (Meehl, 1986). For example, does the taxon represent one or more existing disorders, a new disorder, a variant of an existing disorder, or a subtype of an existing disorder? Are there any ways in which relevant diagnostic criteria can be refined to afford more reliable, valid, and efficient classification? Such refinements might include changes to the symptom list itself, to the diagnostic algorithm (e.g., the number of symptoms from different criteria that must be present for a diagnosis to be made), to exclusionary provisions, or to the subtypes or specifiers that are listed for the disorder. At present, all mental disorders are listed in the *DSM–IV* as discrete categories to be diagnosed on the basis of signs and symptoms that distinguish individuals with and without the disorder. Research on latent structure can help to identify disorders that are best conceived within such a categorical framework, and follow-up analyses can identify the most valid indicators and methods of combining them to achieve the greatest classification accuracy.

In addition, structural research may support the contention that some—perhaps even many or most—disorders are best conceived as continuous (e.g., Widiger & Clark, 2000). Acknowledging the existence of dimensional disorders would increase the difficulty of reaching value judgments that sometimes can be made more simply with categorical classifications. For example, mental health professionals would need to decide on what basis (e.g., specific symptoms, severity of presentation) resources such as psychotherapy, disability compensation, and other services should be provided (A. M. Ruscio et al., 2002). In such cases, "concern for defining a precise boundary between normality and psychopathology might become less important than determining the appropriate professional response to different variants and degrees of psychopathology" (Widiger & Clark, 2000, p. 950). Epidemiological research on the association between dimensional severity levels and clinically relevant outcomes might be used to establish consensual cutoffs (Kessler, 2002), and diagnostic decision-making tools such as receiver operating characteristic curves may help balance the costs and benefits of different decision thresholds for different purposes in different populations (Swets et al., 2000). Because categorical decisions are often unavoidable in routine clinical work (e.g., one can either accept a client or refer the individual elsewhere, reach a particular diagnosis or not, continue or terminate therapy), psychopathologists will need to make reasonable, defensible cuts along latent dimensions and make necessary value judgments in a rigorous and principled manner. Ideally, this process would involve explicit criteria that are empirically supported for the specific decision at hand.

Statistical Approaches to Testing for Taxonic Boundaries

Having introduced the distinctions between latent and manifest levels of analysis and dimensional and taxonic latent structure, reviewed the potentially complex forms that latent structures may take, and described several significant implications of latent structure for research and practice, we now apply these ideas to a fundamental question in psychopathology: the nature of the boundary separating mental disorder from normal functioning. In setting an agenda for the next century of research on psychopathology, Widiger and Clark (2000) noted that "the challenge facing the developers of *DSM–V* may not be to differentiate more clearly between normal and pathologic expressions of behavior; rather, it

may be to determine *whether or not a qualitative distinction can in fact be made*" (p. 950; emphasis added). Although we emphasize how studies of latent structure can help resolve the boundaries of disorders, the points raised herein apply equally well to other psychological constructs within and beyond the realm of psychopathology.

To engage in this research process, we need powerful statistical tools that can correctly uncover latent structures. Fortunately, there exist several analytic techniques that may be used for this purpose. Many of these approaches are complementary in the sense that they address different questions and are therefore useful at different stages of a research program designed to resolve latent structure. We begin by discussing three broad families of analytic techniques that empirically test for taxonic boundaries rather than beginning with the assumption of taxa or dimensions: hierarchical cluster analyses, finite mixture models, and the taxometric method.[4] Later, we discuss how the complementary use of additional procedures can clarify the number and nature of taxa and dimensions underlying a construct or set of related constructs.

Hierarchical Cluster Analyses

Cluster analysis represents an extremely diverse array of statistical techniques that has been used to determine how many relatively homogeneous groups of cases can be distinguished within a given sample (Arabie, Hubert, & DeSoete, 1996; Everitt, 1993; Lorr, 1994; McLachlan & Peel, 2001). We provide a brief overview of two approaches to cluster analyses that have been frequently used in psychological research. In this section, we describe hierarchical clustering procedures; in the next section, we describe finite mixture models.

Several of the most popular clustering techniques involve a process analogous to using multiple manifest variables to create a multidimensional scatterplot, then looking to see whether cases are distributed haphazardly throughout this scatterplot or whether clouds of similar cases tend to clump together. When the number of cases and indicator variables is relatively large, the number of distinct clustering solutions will far exceed the limits of what can feasibly be tested. Thus, many different clustering algorithms have been devised to serve as heuristic search procedures, all of which involve a two-step process. First, a measure of similarity or distance is used to quantify the relations between all cases in a sample. For example, one might compute pairwise correlations between the score profiles of different cases to capture their similarity (and hence relative closeness in the multidimensional scatterplot), or compute squared or unsquared Euclidean distances between pairs of cases to capture their distance from one another. Second, a mathematical rule is applied to parse the resulting matrix of similarity or distance values into clusters. For example, Ward's (1963) criterion of minimizing within-cluster variance is a popular method that tends to yield equal-sized clusters. When using a hierarchical cluster analysis procedure, it is more common to begin by treating each case as a cluster and then sequentially fuse similar clusters until all cases have been joined into one cluster (an *agglomerative* method; Everitt, 1993; Lorr, 1994) than to begin with one cluster and sequentially remove the least similar cases until each has been placed into its own cluster (a *divisive* method).

The most significant problem with using these clustering procedures to empirically assess latent structure is the difficulty of determining the number of clusters that corresponds to the latent structure of the target construct. Because a hierarchical cluster analysis yields multiple solutions (e.g., a 1-cluster solution, a 2-cluster solution, and so forth up to an *N*-cluster solution in which each case is treated as a cluster), the investigator must determine the most appropriate number of clusters. When the number of clusters is itself the central research question, this task requires an empirical index or *stopping rule* that will reveal a construct's true latent structure. Unfortunately, cluster analyses often yield a high rate of false positives (i.e., illusory taxa) because the stopping rules designed to identify the appropriate number of clusters are highly fallible (Grove, 1991a; Milligan & Cooper, 1985). Stopping rules perform particularly poorly in the situation that is the focal concern of this article; namely, determining the presence or absence of a single taxonic boundary purported to separate genuine clusters (Lorr, 1994). Many authors have observed that identifying the true number of clusters remains one of the most challenging problems in cluster analysis (e.g., Everitt, 1993; Lorr, 1994; Milligan, 1996). For this and other reasons, Everitt (1993) recommended that hierarchical cluster analysis be viewed as a descriptive tool, potentially useful for summarizing data or for posing a structural hypothesis. This suggests that the obtained clusters should be submitted to another analytic procedure that provides a powerful test between taxonic and dimensional latent structure.

Finite Mixture Models

Finite mixture models attempt to determine the parameters of hypothetical subgroup distributions, or *components*, that combine to produce the observed distributions of manifest variables (e.g., Everitt & Hand, 1981; McLachlan & Basford, 1988; McLachlan & Peel, 2001). By systematically varying both the number of components and some or all of their parameters (e.g., base rate, mean, standard deviation, skew), a variety of tests can be used to identify a best-fitting mixture model. When finite mixture models are used to assess latent structure, the components in the best-fitting model are interpreted to represent latent classes.

As with hierarchical cluster analyses, it is unclear whether the correct number of components in a mixture can be reliably determined. McLachlan and Basford (1988) describe a number of indices that have been proposed for this purpose and note that this remains a "very difficult problem which has not been completely resolved" (p. 21), a conclusion later echoed by Bock (1996) and subsequently repeated verbatim by McLachlan and Peel (2001, p. 175). Using several simulated data sets, McLachlan and Peel demonstrated that many available indices tend to overestimate the correct number of components. Moreover, because manifest structure may differ from latent structure for numerous reasons (e.g., the "lumpiness" of chance, selective sampling from the extremes of a distribution), even a mixture model that fits the manifest data well may not accurately reflect a construct's true latent structure. In particular, McLachlan and Peel note that although normal mixture models can adequately model skewed manifest distributions, the best-fitting model will often include extra (false) com-

[4] Although they can be applied to either taxonic or dimensional constructs, techniques such as factor analysis and latent class analysis do not address the more fundamental question of whether or not a taxonic boundary exists. These techniques can be quite valuable for addressing other structural questions and will be discussed as complementary procedures that build upon the results of initial structural research.

ponents to account for this skew. This may be a significant concern, given that nonnormal distributions are ubiquitous in psychological research, with skew observed even when rigorously developed achievement and psychometric measures are administered in nondisturbed populations (Micceri, 1989). When psychopathological constructs are assessed—especially in community or analogue populations—the rarity of the phenomena may result in more markedly skewed distributions on symptom measures (cf. J. Ruscio, Ruscio, & Keane, 2004). Thus, the tendency of finite mixture models to uncover too many components in the presence of skewed data seems to be an important limitation of these procedures.

Another difficulty arises when using any clustering approach (hierarchical, mixture model, or otherwise) that searches for multiple taxonic boundaries using a single set of manifest indicators. For example, such approaches might evaluate the latent structure of major depressive disorder (MDD) by submitting a broad array of affective, cognitive, and somatic depressive symptoms—including those relevant to both MDD and its theoretical subtypes—to one analysis. Cluster analysis may appear to offer a great deal of information in an analytically simple package by searching for all taxonic boundaries (e.g., depressive types and subtypes) in a single sweep. However, this simplicity may also pose a liability. A single set of variables is unlikely to be equally valid for detecting multiple, heterogeneous boundaries, and boundaries represented by a subset of the variables may not be detected when variables irrelevant to that boundary are nonetheless included in all calculations or when variables unique to that boundary are excluded from analysis. For example, the search for a boundary delineating a putative melancholic or endogenous subtype of depression may fail if it only includes variables that distinguish depressed from nondepressed individuals. It may be unrealistic to expect any analytic approach to simultaneously uncover all types and their subtypes in a single analysis. Thus, we suggest that analyses of latent structure focus on just one putative taxonic boundary at a time, using variables that sensitively and specifically demarcate that particular boundary.

The Taxometric Method

The taxometric method consists of several diverse analytic procedures that were expressly designed to distinguish taxa from dimensions (e.g., Meehl, 1995, 1999; Meehl & Golden, 1982; Waller & Meehl, 1998). The mathematical model that underlies the method, the GCMT, describes how manifest variables interact to reveal a latent discontinuity between members of a taxon and its complement:

$$\mathrm{cov}(xy) = P\mathrm{cov}_t(xy) + Q\mathrm{cov}_c(xy) + PQ(\bar{x}_t - \bar{x}_c)(\bar{y}_t - \bar{y}_c). \quad (1)$$

This equation expresses the covariance between two indicators x and y as a function of the covariance within the taxon—$\mathrm{cov}_t(xy)$—the covariance within the complement—$\mathrm{cov}_c(xy)$—and the mean difference between the taxon and complement on each of the indicators (i.e., indicator validity), with each term weighted by the base rate(s) of the relevant class(es). If indicators are chosen that covary only negligibly with one another within the taxon and complement classes, the first two terms approach zero and can be dropped from the GCMT, which then simplifies to a function of the taxon and complement base rates and the validity of the two indicators:

$$\mathrm{cov}(xy) = \mathrm{PQ}(\bar{x}_t - \bar{x}_c)(\bar{y}_t - \bar{y}_c). \quad (2)$$

Several taxometric procedures have been independently derived from this mathematical model and can be used to test latent structure. Each nonredundant procedure serves as a consistency check for the results provided by others. For example, the MAXCOV (MAXimum COVariance; Meehl & Yonce, 1996) and MAXEIG (MAXimum EIGenvalue; Waller & Meehl, 1998) procedures work by examining the association among two or more indicators within ordered subsamples along another indicator. In the presence of latent classes, this association fluctuates as the base rates of class members change across subsamples, reaching a maximum value—or peak—in the subsample containing equal numbers of taxon and complement members. The absence of latent classes produces a relatively constant association across subsamples, resulting in a plot that is relatively straight. By contrast, the MAMBAC (Mean Above Minus Below A Cut; Meehl & Yonce, 1994) procedure works by sorting cases on one indicator, then using a series of sliding cuts on a second indicator to search for an optimal cutting score that separates taxon from complement cases. Taxonic structure yields an optimal cut score and hence a peaked MAMBAC curve, whereas dimensional structure produces a concave curve.

Many other taxometric procedures and consistency tests can be used to check the results yielded by MAXCOV/MAXEIG and MAMBAC and build confidence in a structural solution. For example, one can use the *inchworm consistency test* (Waller & Meehl, 1998) to clarify an ambiguous MAXEIG curve by systematically increasing the number of ordered subsamples in the analysis. This test allows researchers to determine whether rising MAXEIG curves with a right-end cusp reflect the presence of a small latent class or the influence of positively skewed indicators of a latent dimension (J. Ruscio, Ruscio, & Keane, 2004). One can also use the L-Mode (Latent Mode; Waller & Meehl, 1998) procedure to examine the number of latent modes present in the distribution of estimated scores on the first principal factor of all indicators. Likewise, one can obtain an estimate of the taxon base rate from each taxometric curve, then compare these estimates within and across procedures to determine whether they closely converge on a single, plausible rate of taxon members in the sample. For a more detailed discussion of the taxometric procedures and consistency tests that may be appropriate for a particular data set, as well as cautionary notes on consistency tests that have gained some popularity despite a poor ability to distinguish taxonic from dimensional structure (e.g., distributions of Bayesian probabilities of taxon membership; the goodness-of-fit index described in Waller & Meehl, 1998), see J. Ruscio, Haslam, and Ruscio (2003) or J. Ruscio & Ruscio (in press).

Unlike most other approaches to assessing latent structure, the taxometric method tests for one taxonic boundary at a time using a set of variables specifically chosen to identify that boundary. A growing body of Monte Carlo studies attests to the ability of taxometric procedures to distinguish the presence or absence of a taxonic boundary when suitable data are available (see Haslam & Kim, 2002, for a review). At the same time, because each taxometric analysis involves multiple indicator variables, the use of taxometrics is limited to constructs that are represented by multiple

criteria.[5] Moreover, because taxometric analyses test for just one taxonic boundary at a time, they may need to be performed sequentially to detect multiple boundaries, each time using a set of indicators appropriate to the boundary in question.

Conclusions

Each of the statistical approaches described above has a unique set of strengths and weaknesses. Although they may serve as useful tools in an exploratory mode or for addressing other kinds of research questions, hierarchical cluster analyses and finite mixture models appear to have questionable utility for testing the number of taxonic boundaries underlying a construct, and therefore for distinguishing between the presence and absence of a single taxonic boundary. Because the taxometric method focuses on one boundary at a time, has a strong track record of correctly distinguishing the presence versus absence of such a boundary, and provides the added safeguard of assessing agreement among multiple consistency tests, we advocate its use when addressing the most basic of structural questions—that of the existence of boundaries related to the construct of interest.

Conceptual and Methodological Issues in Taxometric Research

Because the initial stages of structural research will influence the form and content of subsequent investigations, it is essential that foundational taxometric studies be conducted using appropriate samples of data with indicator variables that afford a powerful test of the putative taxonic boundary. We review several of the most significant issues that must be considered carefully before conducting a taxometric analysis.

Sampling

When selecting or constructing a sample for a taxometric investigation, care must be taken to avoid sampling approaches that may yield spuriously taxonic results. The ideal sample is one in which participants are drawn randomly from the population of interest. If the putative taxon represents an exceedingly small class of individuals (e.g., when studying a rare form of psychopathology or when conducting psychopathology research in nonclinical samples), a particularly large sample may be required. Although it may be tempting to construct a sample for taxometric analysis comprising approximately equal numbers of suspected taxon and complement members, this approach can lead to spuriously taxonic results if individuals with intermediate symptom severity are systematically excluded or if the selection criteria are correlated with additional, confounding factors that distinguish the two groups. For example, Grove (1991a) discussed problems of *institutional pseudotaxa* that can result from nonrandom selection on one or more variables (e.g., hospitalization) related to the construct under investigation.

Assessment of the Construct

Within an appropriate sample, the indicator variables to be analyzed must provide adequate coverage of the target construct. This is critical because, in a taxometric analysis, the latent taxon under study is identified by the set of manifest indicators chosen to search for it (cf. Meehl, 1986), and a poor choice of indicators may yield misleading results (Widiger, 2001). For example, an analysis including indicators of binge eating alone would identify the latent structure of binge eating disorder, whereas an analysis including indicators of both binge eating and inappropriate compensatory behaviors would identify the structure of bulimia nervosa. It is therefore important that the chosen indicator set capture as many signs and symptoms of the putative psychopathological taxon as possible, whether these are drawn from *DSM–IV* criteria, from another nosological system, or from other contemporary theories of the construct. By the same token, taxometric results may be seriously confounded if a significant number of indicators inadvertently triangulate on another construct. For example, given the substantial symptom overlap and diagnostic co-occurrence (Lilienfeld, Waldman, & Israel, 1994) of MDD and generalized anxiety disorder (GAD), a taxometric investigation of GAD would require careful selection of a set of indicators that characterize this disorder but that are not equally (or almost equally) characteristic of MDD. Just as the interpretation of a latent factor depends on the particular variables submitted to a factor analysis and the relative loadings of these variables on the latent factor, the construct about which structural conclusions can be drawn depends on the particular set of indicators submitted to a taxometric analysis.

It is also important that each indicator be sensitive to the putative taxon at the appropriate severity level (P. E. Meehl, personal communication, May 10, 2001). Thus, a variable reflecting abnormally and persistently euphoric mood would be more likely to isolate a putative mania taxon than a variable reflecting positive mood, which many individuals—manic and nonmanic—would be expected to endorse.

Suitability of Data for Taxometric Analysis

Sufficient representation of the putative taxon. To be useful, a sample must contain enough members of the putative taxon to allow this taxon to be detected. Although published Monte Carlo studies have not examined the performance of taxometric procedures with taxon base rates lower than .10 (e.g., Meehl & Yonce, 1994, 1996), it may be that the absolute number of taxon members in the sample is equally or more important than the base rate of taxon membership in determining whether a small taxon will be detected (J. Ruscio & Ruscio, in press). It may also be the case that some taxometric procedures are considerably more sensitive to small taxa than others. To examine these possibilities, we simulated data sets in which taxon members were separated by 2 *SD* from complement members on each of four indicators. Data were generated using the procedure described by Meehl and Yonce (1994, 1996), and five samples were generated at each of a number of sample sizes. With the taxon held constant at $n = 100$, sample sizes ranged from $N = 200$ to $N = 80{,}000$; with the taxon held constant at $n = 200$, sample sizes ranged from $N = 400$ to $N = 160{,}000$ (see Table 1). Thus, in each series of simulated data sets, taxon base rates ranged from .50 to .00125, the latter falling well below the conventionally accepted limit of taxometric sensitivity.

Three taxometric procedures were used to analyze these data: MAMBAC, MAXEIG, and L-Mode. MAMBAC and MAXEIG were conducted using a composite input indicator method that has

[5] Among taxometric procedures, there is one exception: The *normal minimum chi-square method* (Golden & Meehl, 1973) is essentially a mixture model and requires just one indicator variable.

Table 1
Detection of Increasingly Small Taxa Using MAMBAC, MAXEIG, and L-Mode

	$N_{taxon} = 100$				$N_{taxon} = 200$			
P	*N*	MAMBAC	MAXEIG	L-Mode	*N*	MAMBAC	MAXEIG	L-Mode
.50	200	X	X	X	400	X	X	X
.25	400	X	X	X	800	X	X	X
.10	1,000	X	X	X	2,000	X	X	X
.05	2,000	X	X		4,000	X	X	X
.025	4,000		X		8,000	X	X	
.01	10,000		X		20,000		X	
.005	20,000		X		40,000		X	
.0025	40,000		X		80,000		X	
.00125	80,000				160,000		X	

Note. Each of four indicators was simulated as a random normal deviate ($M = 0$, $SD = 1$), with 2 *SD* added to cases in the simulated taxon. An "X" represents the successful identification of the taxon in at least three of the five samples for that *P* (taxon base rate) and *N* (sample size). For MAMBAC and MAXEIG, the criterion was a peak in the graph (not merely a right-end cusp); for L-Mode, the criterion was a bimodal distribution of factor scores. MAMBAC = means above minus below a cut; MAXEIG = maximum eigenvalue; L-Mode = latent mode.

been found to increase statistical power and assist in the detection of small taxa (J. Ruscio, 2002; J. Ruscio & Ruscio, in press), and curves were averaged to reduce the effects of sampling error. The taxon was judged to be successfully detected if the graphical results for at least 3 of the 5 samples with a given sample size yielded a discernible peak in MAMBAC and MAXEIG curves—not merely a right-end cusp, which is easily produced by dimensional data in the presence of significant indicator skew (A. M. Ruscio & Ruscio, 2002; J. Ruscio, Ruscio, & Keane, 2004)—and yielded discernibly bimodal L-Mode curves. For each taxon size ($n = 100$ or 200), each procedure was conducted in successively larger samples (which therefore possessed lower taxon base rates) until the taxon could no longer be detected in three or more of the five samples.

Results of these analyses appear in Table 1, and illustrative taxometric curves for the lowest taxon base rate detected by each procedure appear in Figure 2. Of the three procedures, L-Mode was the least sensitive to small taxa, MAMBAC performed slightly better, and MAXEIG was extraordinarily sensitive when a sufficiently large number of overlapping windows was used. In addition, each procedure was able to detect a lower taxon base rate when the taxon contained a larger number of cases. With a taxon of $n = 100$, L-Mode was sensitive down to a base rate of .10 ($N = 1{,}000$), MAMBAC to a base rate of .05 ($N = 2{,}000$), and MAXEIG to a base rate of .0025 ($N = 40{,}000$); with a taxon of $n = 200$, L-Mode was sensitive down to a base rate of .05 ($N = 4{,}000$), MAMBAC to a base rate of .025 ($N = 8{,}000$), and MAXEIG to a base rate of .00125 ($N = 160{,}000$). Thus, with twice as large a taxon, taxonic structure was detected by each procedure even when the taxon base rate itself was halved. Although the parameters of these data were highly favorable for taxometric analysis (e.g., high validity, no within-group nuisance covariance, no indicator skew), it is nonetheless remarkable that 200 taxon members were detected correctly in a sample of 160,000 individuals and that the three taxometric procedures evidenced consistent and substantial differences in sensitivity to small taxa despite their application to the same data sets.

While systematic Monte Carlo research is needed to specify the number of taxon members that are required for taxon detection by each taxometric procedure under a variety of realistic research conditions, this simple demonstration suggests some tentative conclusions. First, efforts to detect small taxa will likely require large samples. Second, there may be substantial differences in the sensitivity of various taxometric procedures to small taxa. In particular, MAXEIG appears to be an especially powerful technique when a very small taxon or low taxon base rate is suspected. Third, the absolute number of taxon members appears to be an important determinant of whether a taxon will be detected in taxometric analyses. This argues against the practice of discarding putative complement members from the sample to boost the taxon base rate, because this does nothing to increase the number of taxon members in the analysis.

Adequacy of other data parameters. Having sufficient representation of putative taxon members is only one important aspect of a sample of data. It is also important, for example, that indicator variables separate the putative groups with sufficient validity, that indicators are relatively uncorrelated within putative groups, and that indicators are not too highly skewed. Because it can be difficult, if not impossible, to judge whether the specific configuration of relevant parameters of a research data set is appropriate for taxometric analysis, we have developed an approach that can be used to empirically test the suitability of a set of research data for taxometric analysis (J. Ruscio, Ruscio, & Meron, 2003). Rather than relying on general rules of thumb or presumptions about the adequacy of one's data, we advocate that researchers simulate taxonic and dimensional comparison data sets that closely match the parameters of the research data. We have created a simulation procedure that matches all distributional properties of a data set and reproduces indicator correlations as closely as sampling error allows; this procedure can be used to generate one or more data sets with taxonic latent structure as well as one or more data sets with dimensional latent structure (J. Ruscio, Ruscio, & Meron, 2003). The resulting simulated data sets are submitted to the taxometric analyses planned for the research data. If parallel analyses of these simulated taxonic and dimensional data sets yield discernibly different results, the parameters of the research data can be considered capable of distinguishing between these structures, and should therefore be suitable for taxometric analysis.

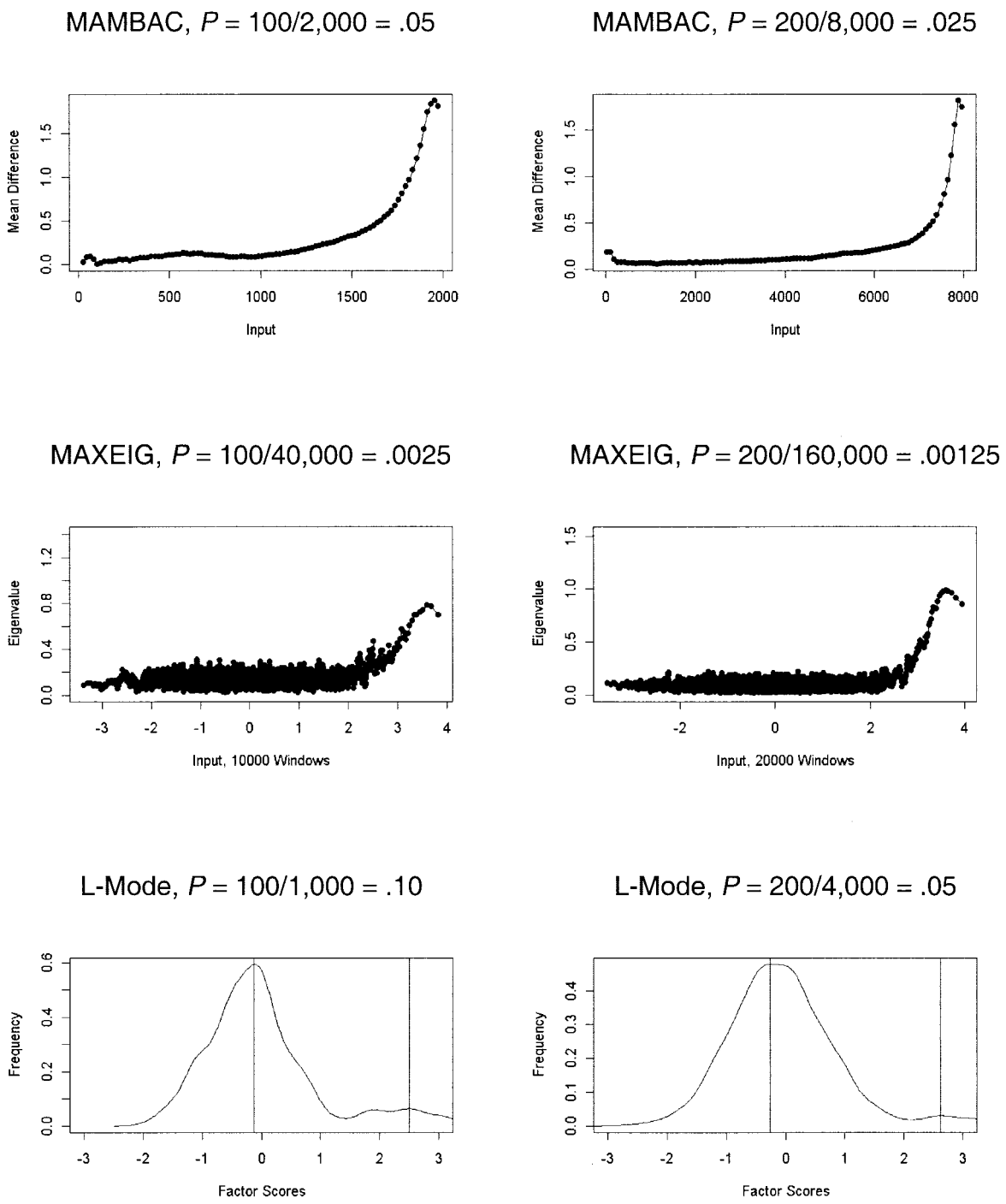

Figure 2. Illustrative results for the lowest detectable taxon base rate using MAMBAC, MAXEIG, and L-Mode. Each of the four indicators was simulated as a random normal deviate ($M = 0$, $SD = 1$), with 2 *SD* added to cases in the simulated taxon. P = taxon base rate, N = sample size. For MAMBAC and MAXEIG, the criterion for successful taxon detection was a peak in the averaged curve (not merely a right-end cusp) for at least three of the five samples of a given size; for L-Mode; the criterion was a bimodal distribution of factor scores in at least three of the five samples. MAMBAC = mean above minus below a cut; MAXEIG = maximum eigenvalue; L-Mode = latent mode.

The results yielded by simulated comparison data can also greatly facilitate the interpretation of research results. By holding constant the unique distributional and correlational properties of the research data across parallel series of taxometric analyses, all that systematically varies is the latent structure of each data set. Thus, differences in curve shape between taxonic and dimensional data can be assessed independent of factors that may distort the shapes of taxometric curves (e.g., indicator skew; J. Ruscio, Rus-

cio, & Keane, 2004). If the research results appear considerably more similar to those yielded by one or the other known latent structure, this affords greater confidence in a particular structural conclusion. A number of recent taxometric investigations have taken advantage of simulated comparison data to test the suitability of data for analysis and to provide an interpretive aid (e.g., Rothschild, Cleland, Haslam, & Zimmerman, 2003; A. M. Ruscio, & Ruscio, 2002; A. M. Ruscio, Ruscio, & Keane, 2002; J. Ruscio, Ruscio, & Keane, 2004).[6]

Complementarity of Approaches to Latent Structure

When a number of taxometric analyses and consistency tests are performed in an appropriate sample of data using empirically suitable indicators that adequately represent the target construct, the results should clearly support or refute the existence of a putative taxonic boundary. However, because the latent structure of a disorder may be more complex than the simple distinction between two taxa and one dimension, initial taxometric research addressing this fundamental question can be supplemented by additional procedures that address remaining structural and conceptual questions about the construct under investigation. In this section, we describe how follow-up analyses, drawing upon a variety of analytic techniques, can build on taxometric results to elaborate the full latent structure of a mental disorder.

Further Exploration of Dimensional Results

If taxometrics suggest that a construct is dimensional, its structure can be tested further using statistical methods that are familiar to most psychologists. For example, exploratory or confirmatory factor analysis (e.g., Byrne, 2001; Gorsuch, 1983; Long, 1983) can be used to determine whether the construct is uni- or multidimensional, to identify the indicators that best assess and differentiate the factors, or to test for the existence of higher and lower order factors. In addition, latent trait analyses such as IRT can be used to examine critical properties of the indicators of each latent dimension. Results of these analyses can identify indicators that provide comparable measurement reliability across the full range of trait levels (for general assessment or research purposes) or that focus measurement precision near a particular trait level (for decision-making purposes).

Further Exploration of Taxonic Results

If taxometric analyses suggest the presence of a taxonic boundary, there are a number of promising avenues for follow-up investigation. These avenues are not mutually exclusive; indeed, the most profitable approach may be to apply a variety of complementary analytic techniques within a systematic program of research aiming to unearth the full latent structure of a construct.

First, following the identification of one taxonic boundary, the same or a highly similar set of indicators can be submitted to subsequent taxometric analyses to test for further boundaries hypothesized to exist within the taxon or complement. For example, an initial analysis may distinguish a taxon corresponding to Bipolar I disorder from a complement of individuals without this disorder; subsequent analysis within the complement may identify a second, cyclothymic taxon distinct from individuals who are free of symptoms.

Second, one could subject new sets of indicators to subsequent taxometric analyses to test for subtypes within the taxon. For example, having used one set of indicators to identify a bipolar disorder taxon, new indicators could be used to test for the presence of a rapid-cycling subtype within the sample of cases classified into the bipolar taxon. Although there is nothing inherent in the mathematics of taxometrics to favor searching for lower-order taxa (subtypes) within a type rather than searching for a higher-order taxon (type) that subsumes the putative subtype, each of these approaches presents strengths and weaknesses. It should be noted, however, that when a latent taxon is identified, it may be problematic to refer to it as a "subtype"—rather than simply a "type"—until a higher order taxon is also identified and shown to subsume the alleged subtype. For example, when investigators conclude that there is evidence for a melancholic or endogenous depression taxon (e.g., Ambrosini, Bennett, Cleland, & Haslam, 2002; Grove et al., 1987; Haslam & Beck, 1994), it does not necessarily follow that this taxon is a subtype of MDD. However, the premature employment of subtype terminology can lead to unwarranted conclusions about the latent structure of MDD (J. Ruscio & Ruscio, 2000). Referring to a melancholic type of depression rather than a melancholic subtype of depression would reflect the categorical nature of melancholia without implying that an overarching depressive construct is itself taxonic.

Another way to avoid semantic confusion is by searching for a putative type before searching for its putative subtypes. An added advantage of this approach is that once a type is identified, it defines the population within which one or more putative subtypes may be investigated. On the other hand, if a number of heterogeneous subtypes exist, it may not be possible to begin by studying the higher-order type that encompasses them all. This task would require indicator variables which distinguish the taxon (individuals who belong to any of the subtypes) from its complement (individuals who belong to none of the subtypes), yet which do not distinguish among the various subtypes. Such indicators may be conceptually and methodologically difficult to obtain, making a bottom-up approach necessary.

Whether one adopts a top-down or bottom-up approach, we suggest testing for one type or subtype at a time using a set of indicator variables carefully chosen for that purpose. Although there are procedures (e.g., cluster analysis) that can be used to search for multiple taxa simultaneously, simultaneous searches overlook the critical importance of careful indicator selection for detecting an intended taxon. As noted previously, each putative taxon is represented by a specific set of indicator variables, and it seems highly unlikely that a single indicator set will apply equally well to multiple taxonic boundaries. We recommend a more deliberate, theory-driven plan of attack: Choose a set of indicators to assess one conjectured taxonic boundary, test for it, and then repeat the process for the next boundary using a new (though potentially overlapping) set of indicators that is appropriate for that boundary.

It is also worth noting that the search for boundaries between subtypes is similar to the search for an overarching type in that each analysis requires phenotypically distinct indicator variables.

[6] Program code to simulate comparison data, as well as to perform taxometric analyses, is available online in both the R and S+ languages at http://www.etown.edu/psychology/faculty/ruscio.htm

That is, in any taxometric investigation, indicators should reflect multiple facets of a construct and be nonredundant with one another. Strategies that force phenotypically similar items onto separate indicators (e.g., combining every third item into one of three composite indicators) are likely to produce indicators that are correlated within the putative taxon and complement groups to a problematic extent. If a construct itself is not multifaceted, or if available measures do not adequately assess multiple facets of the construct, the taxometric method is not an appropriate statistical tool.

Third, just as factor analysis or IRT can be used to identify the most efficient and least redundant indicators of a latent dimension, the same can be done for taxonic constructs using statistical procedures designed for this purpose. In addition to estimates of indicator validity provided by taxometric procedures themselves, other indicator-related information may be obtained by conducting latent class analysis (e.g., Green, 1951; Lazarsfeld & Henry, 1968), a categorical analogue of factor analysis that uses manifest categories as indicators of latent taxa, or latent profile analysis (e.g., Muthén, 2001), which uses manifest continua as indicators of latent taxa. Because latent class and latent profile analysis begin with the presumption that latent structure is taxonic, these procedures are not likely to be as appropriate as taxometrics at the initial phase of differentiating taxa from dimensions. Nevertheless, they can provide useful information about indicator validity once the structure underlying these indicators is shown to be taxonic.

Finally, each taxon that is uncovered—including types and subtypes—may be assessed for dimensional variation. Procedures such as factor analysis or analyses of internal consistency can be used to determine whether reliable residual variation exists among members of the taxon or complement. If reliable variance is identified, one can proceed with any of the research strategies described previously for clarifying and elaborating dimensional results (e.g., exploratory or confirmatory factor analysis, IRT modeling).

Delineating the Boundaries Between Disorders

Up to this point, we have limited our discussion of boundary issues to questions about the structure of individual disorders. An entirely different boundary question that has received considerable attention concerns the widespread and substantial levels of diagnostic co-occurrence among mental disorders (Widiger & Clark, 2000). Because the term *comorbidity* has been used to represent differing concepts, and because this term may blur the critical distinction between latent and manifest levels of analysis that is focal to the present paper, we instead use the term *diagnostic co-occurrence* to refer to this boundary issue (Lilienfeld, Waldman, & Israel, 1994). The term draws attention to the fact that overlap has been observed at the manifest level, where fallible signs, symptoms, and inclusion/exclusion criteria define diagnostic categories that may or may not accurately reflect the true nature of psychopathology. Thus, the most rapid advances in our understanding of relationships between disorders and the implications of these relationships for theory, research, and practice, may be attained by focusing once again on the latent level (Meehl, 2001a; Waldman & Lilienfeld, 2001).

A useful framework for considering the potential causes of diagnostic co-occurrence is provided by Klein and Riso (1993), who list nearly a dozen candidate explanations (see Table 2). The first nine of these explanatory models are artifactual in that diagnostic co-occurrence exists only at the manifest level and could, in principle, disappear with improvements in sampling, nosology, or measurement. However, co-occurrence that stems from the final two models (one disorder is a risk factor for the other; both disorders share a common etiology) is more theoretically and clinically relevant because it involves meaningful, etiological, beyond-chance associations between latent constructs.

Klein and Riso (1993) advocate family studies as the method of choice for comparing explanatory models of diagnostic co-occurrence. We propose an additional avenue through which these explanations can be distinguished, one which builds on the results revealed by studies of latent structure. As we note in Table 2, each of Klein and Riso's explanatory models affords a specific, testable prediction about the relation between the latent constructs represented by co-occurring disorders. Thus, to explicate the boundary between co-occurring disorders, we recommend the following three-stage approach. First, investigate the latent structure of each disorder individually to determine whether or not it can be distinguished from normality (and/or from subclinical levels of the disorder). Second, assign individuals to taxa or locate them along a dimension as reliably and validly as possible. Third, using individuals' estimated class membership or dimensional scores for each disorder, estimate the degree of association between the two latent constructs.

Figure 3 depicts several patterns of results that may arise when cases are assigned to the taxon or complement of two taxonic disorders. For example, if schizophrenia and major depression were both taxonic, individuals could be schizophrenic and depressed, schizophrenic and nondepressed, nonschizophrenic and depressed, or nonschizophrenic and nondepressed. Each of the panels illustrates a crosstabulation of classification frequencies for two taxonic disorders whose observed base rates of diagnosis are .20 and .10, respectively. In Panel A, the rate of co-occurrence precisely matches the product of the base rates, suggesting the presence of two constructs that do not co-occur beyond chance levels (a solution consistent with Models 1–4, 6, and 7 in Table 2). In Panel B, all cases diagnosed with Disorder B are also diagnosed with Disorder A (but not vice versa), suggesting that one disorder is subsumed by the other (Model 5). In Panel C, many of the diagnosed cases of Disorders A and B belong to neither latent construct, suggesting that cases diagnosed with both conditions are actually members of a third, undiagnosed disorder (Model 8). In Panel D, Disorders A and B are alternate expressions or different phases of the same latent construct (Model 9). All four of these patterns of results suggest that the observed co-occurrence between disorders may be attributed to artifactual factors rather than to meaningful association at the latent level. By contrast, Panel E depicts two latent constructs whose co-occurrence exceeds the product of their base rates, suggesting that the two disorders are etiologically related (Model 10 or 11). Each of these panels presumes uncorrelated errors in the classification of individuals to Disorders A and B. To the extent that errors are in fact correlated, this may skew the results toward those of Panel E and hence blur the distinction between explanatory models.

Additional factors may need to be considered to rule out explanatory models on the basis of sampling and base rates (Models 1–3). As we have already noted, co-occurrence due to chance may be ruled out by testing the degree of association between disorders; a statistically significant deviation from independence suggests

Table 2
Klein and Riso's (1993) Models of Diagnostic Co-Occurrence and Their Implications for Latent Structure

Model	Description	Latent structure
1. Chance	Two disorders co-occur by chance (e.g., taxa co-occur at an expected rate equal to the product of their base rates).	Two independent constructs
2. Sampling bias	Treatment seeking increases with number of disorders (Berkson, 1946).	Two independent constructs
3. Population stratification	Risk factors not randomly distributed in the population (e.g., intermarriage among at-risk/disordered individuals).	Two independent constructs
4. Overlapping criteria	Shared signs and symptoms in disorders' diagnostic criteria	Two independent constructs
5. Subset of cases	One disorder is encompassed by another.	Two concentric constructs
6. Multiformity	"Comorbid" condition represents an atypical form of one disorder that is similar to—and mistaken for—another.	Two independent constructs
7. Heterogeneity	"Comorbid" condition represents atypical forms of two disorders.	Two independent constructs
8. Third condition	"Comorbid" condition is actually a third independent disorder.	Three independent constructs
9. Alternate expressions	"Pure" and "comorbid" conditions are different phases or alternate expressions of same disorder.	One construct
10. Risk factor	One disorder is a risk factor for another.	Two associated constructs
11. Overlapping etiologies	Two disorders share risk factor(s).	Two associated constructs

theoretically meaningful overlap. By contrast, co-occurrence due to sampling bias can only be ruled out by conducting the test of association in an unselected (e.g., community) sample. This sample need not be the same one in which the structural research was conducted; indeed, because of the far higher base rates of most psychological disorders in clinical settings, it may be preferable to perform structural analyses in clinical samples and then test the association between disorders in a separate community sample. Next, as Klein and Riso (1993) noted, population stratification is a particularly difficult explanation to rule out without conducting a family study. Finally, even if all artifactual models can be ruled out with confidence, additional information is required to determine which of the two etiologically based models can account for the relationship between the disorders. For example, one useful clue may be the temporal order of onset of the two disorders, such that a disorder posing a risk factor for another consistently occurs earlier in time.

Two additional issues must be considered when using structural investigations to explicate the boundaries between disorders. First, we do not believe that taxometrics should be used to examine the boundary between disorders directly (i.e., by conceptualizing one disorder as the taxon and the other as the complement). One reason is the challenge of selecting appropriate indicators to differentiate the disorders. The boundaries that have generated the greatest interest are those between disorders whose clinical features partially overlap. However, to the extent that some of the signs and symptoms that represent the putative taxon (one disorder) are shared by its complement (another disorder), researchers may be forced to choose between (a) including shared (and thus poorly differentiating) indicators in the analysis whose insufficient validity may undermine the taxometric results, and (b) including only indicators that are unique to each disorder in the analysis, with this narrowed content coverage raising questions about the constructs that are actually being studied.[7] Second, there is the challenge of selecting an appropriate sample. The taxometric method presumes that all individuals are members of the conjectured taxon or its complement class, but not both. Excluding individuals who are diagnosed with both disorders would satisfy this methodological requirement, but would fail to address the original question by systematically excluding the very co-occurrence to be explained. Third, and perhaps most important, is that the taxometric model presumes that any groups detected by the method are mutually exclusive. If the taxometric method does not permit meaningful relationships to exist between taxon and complement at the latent level, the method cannot be used to address questions about the nature of these relationships. Thus, neither taxonic nor dimensional results could be used to disentangle the models of co-occurrence compiled by Klein and Riso (1993).

A second issue concerns whether diagnostic co-occurrence is theoretically and empirically meaningful when the disorders involved are dimensional rather than taxonic. From our perspective, what is most important is the causal mechanism responsible for the observed co-occurrence. Even if the relevant disorders are dimen-

[7] A similar problem could emerge when testing for subtypes of a disorder within a previously identified taxon. As the pool of indicators that validly distinguish between putative taxa shrinks due to overlapping symptoms, it will become increasingly difficult to identify indicators that unambiguously target a particular taxonic boundary.

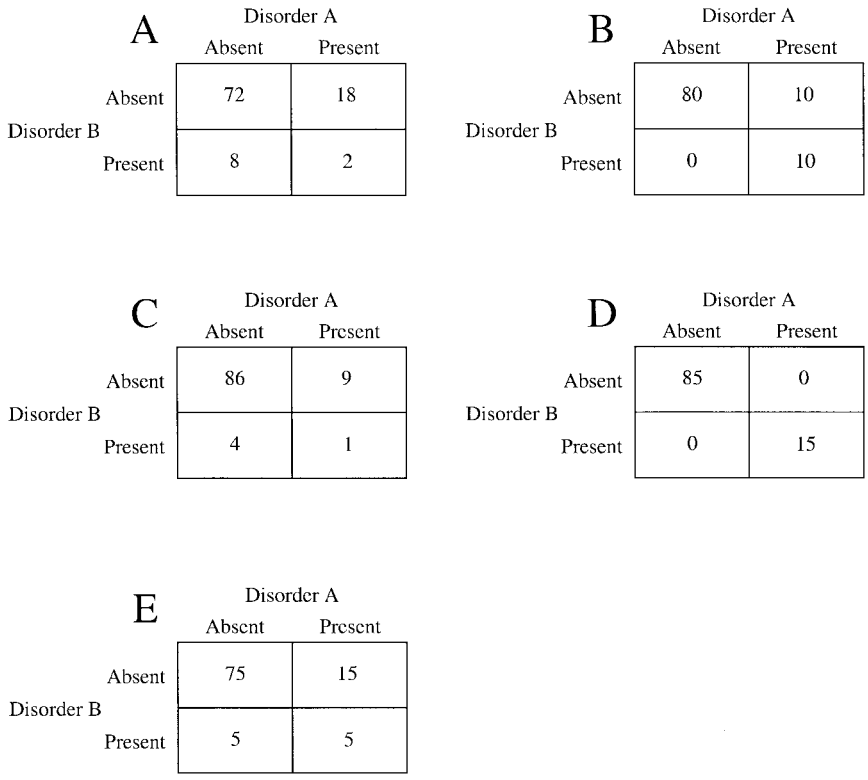

Figure 3. Patterns of co-occurrence between cases infallibly assigned to the latent taxon (present) and complement (absent) classes corresponding to two hypothetical disorders. The diagnostic base rates in a sample of 100 cases are .20 for Disorder A and .10 for Disorder B. Panel A depicts two disorders that co-occur at purely chance levels ($.02 = .20 \times .10$). Panel B depicts one disorder subsumed within another; everyone with Disorder B also has Disorder A, which encompasses other individuals as well. Panel C depicts three independent disorders; many of the cases diagnosed with Disorder A or Disorder B actually belong to neither latent class but to a third (undiagnosed) disorder, causing the rates with which cases are assigned to latent taxa to be less than the diagnostic base rates of .20 and .10. Panel D depicts a single disorder; Disorders A and B are synonymous at the latent level, and the rates with which cases are assigned to taxa differ from diagnostic base rates of .20 and .10 because of classification errors at the manifest level. Panel E depicts two disorders that co-occur at greater-than-chance levels ($.05 > .20 \times .10$).

sional in nature, the discovery that a shift along one dimension causally influences variation along the other provides important information that may enhance our understanding of psychopathology, improve prediction of symptom presentation and course, and facilitate the development of increasingly effective clinical interventions. Thus, we give no special priority to taxonic disorders in the study of diagnostic co-occurrence, and assert that it is worth exploring why individuals simultaneously meet the diagnostic criteria for multiple disorders regardless of these disorders' latent structures. In sum, investigations of latent structure may provide a novel and pragmatically viable way to address the rapidly expanding literature on diagnostic co-occurrence.

Conclusion

Questions about the boundaries that separate disorders from normality and from one another are fundamental to contemporary psychopathology theory, research, and practice. Thus, it behooves psychologists to test the latent structure of psychopathological constructs directly rather than to presume what this structure may be. Given unique strengths of the taxometric method for distinguishing the presence versus absence of a taxonic boundary, we suggest that this method be used early in a structural research program to construct a skeletal outline of the boundaries of each disorder. This outline may then be filled in, elaborated, and refined by complementary statistical approaches until the structure of the disorder is delineated in its entirety. Once the boundary between disorder and normality is resolved, the boundary between pairs of disorders can be examined to rule out artifactual associations and to identify meaningful etiological relationships among disorders. However, these considerable research efforts will be wasted if the structural knowledge that they produce is not applied to the major tasks of our profession. To the extent that this knowledge is used to enhance the reliability and validity of psychological assessments, maximize the power of research designs, and improve diagnostic classification systems, fundamental research on the structure of psychopathology will promote theoretically, practically, and clinically significant progress.

References

Ambrosini, P., Bennett, D. S., Cleland, C. M., & Haslam, N. (2002). Taxonicity of adolescent melancholia: A categorical or dimensional construct? *Journal of Psychiatric Research, 36,* 247–256.

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

Arabie, P., Hubert, L. J., & DeSoete, G. (1996). *Clustering and classification.* River Edge, NJ: World Scientific Publishing.

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

Bock, H. (1996). Probability models and hypothesis testing in partitioning cluster analysis. In P. Arabie, L. J. Hubert, & G. DeSoete (Eds.), *Clustering and classification* (pp. 377–453). River Edge, NJ: World Scientific Publishing.

Byrne, B. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Carson, R. C. (1991). Dilemmas in the pathway of the *DSM–IV*. *Journal of Abnormal Psychology, 100,* 302–307.

Carson, R. C. (1996). Aristotle, Galileo, and the *DSM* taxonomy: The case of schizophrenia. *Journal of Consulting and Clinical Psychology, 64,* 1133–1139.

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

Dahlstrom, W. G. (1995). Pigeons, people, and pigeon-holes. *Journal of Personality Assessment, 64,* 2–20.

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

Embretson, S. E., & Herschberger, S. (1999). *The new rules of measurement: What every psychologist and educator should know.* Mahwah, NJ: Erlbaum.

Everitt, B. S. (1993). *Cluster analysis* (3rd ed.). London: Edward Arnold.

Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions.* New York: Chapman & Hall.

Gangestad, S. W., Bailey, J. M., & Martin, N. G. (2000). Taxometric analyses of sexual orientation and gender identity. *Journal of Personality and Social Psychology, 78,* 1109–1121.

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

Golden, R. R., & Meehl, P. E. (1973). Detecting latent clinical taxa, IV: An empirical study of the maximum covariance method and the normal minimum chi-square method using three MMPI keys to identify the sexes. *Reports from the research laboratories of the Department of Psychiatry, University of Minnesota,* Report No. PR-73-2.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Grayson, D. A. (1987). Can categorical and dimensional views of psychiatric illness be distinguished? *British Journal of Psychiatry, 151,* 355–361.

Green, B. F. (1951). A general solution for the latent class model of latent structure analysis. *Psychometrika, 16,* 151–166.

Grove, W. M. (1991a). The validity of cluster analysis stopping rules as detectors of taxa. In D. Cicchetti & W. M. Groves (Eds.), *Thinking clearly about psychology* (Vol. 1, pp. 313–329). Minneapolis: University of Minnesota Press.

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

Grove, W. M., Andreasen, N. C., Young, M., Endicott, J., Keller, M. B., Hirschfeld, R. M. A., et al. (1987). Isolation and characterization of a nuclear depressive syndrome. *Psychological Medicine, 17,* 471–484.

Guilford, J. P. (1954). *Psychometric methods.* New York: McGraw-Hill.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Hambleton, R. K., Swaminathan, H. S., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

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

Haslam, N. (1999). Taxometric and related methods in relationships research. *Personal Relationships, 6,* 519–534.

Haslam, N., & Beck, A. T. (1994). Subtyping major depression: A taxometric analysis. *Journal of Abnormal Psychology, 103,* 686–692.

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

Kendell, R. E. (1975). *The role of diagnosis in psychiatry.* Oxford, England: Blackwell Scientific.

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

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

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin.

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

Long, J. S. (1983). *Confirmatory factor analysis.* Beverly Hills, CA: Sage.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Lorr, M. (1994). Cluster analysis: Aims, methods, and problems. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 179–195). New York: Springer.

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

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55,* 95–107.

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

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering.* New York: Marcel Dekker.

McLachlan, G. J., & Peel, D. (2001). *Finite mixture models.* New York: Wiley.

Meehl, P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist, 17,* 827–838.

Meehl, P. E. (1986). Diagnostic taxa as open concepts: Metatheoretical and statistical questions about reliability and construct validity in the grand strategy of nosological revision. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology: Toward the DSM–IV* (pp. 215–231). New York: Guilford.

Meehl, P. E. (1990). Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. *Journal of Personality Disorders, 4,* 1–99.

Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality, 60,* 117–174.

Meehl, P. E. (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. (2001a). Comorbidity and taxometrics. *Clinical Psychology: Science and Practice, 8,* 507–519.

Meehl, P. E. (2001b). Primary and secondary hypohedonia. *Journal of Abnormal Psychology, 110,* 188–193.

Meehl, P. E., & Golden, R. R. (1982). Taxometric methods. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127–181). New York: Wiley.

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

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

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

Milligan, G. W. (1996). Clustering validation: Results and applications for applied analyses. In P. Arabie, L. J. Hubert, & G. DeSoete (Eds.), *Clustering and classification* (pp. 341–375). River Edge, NJ: World Scientific Publishing.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50,* 159–179.

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

Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Erlbaum.

Rothschild, L., Cleland, C. M., Haslam, N., & Zimmerman, M. (2003). A taxometric study of borderline personality disorder. *Journal of Abnormal Psychology, 112,* 657–666.

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

Ruscio, A. M., & Ruscio, J. (2002). The latent structure of analogue depression: Should the BDI be used to classify groups? *Psychological Assessment, 14,* 135–145.

Ruscio, A. M., Ruscio, J., & Keane, T. M. (2002). The latent structure of post-traumatic stress disorder: A taxometric investigation of reactions to extreme stress. *Journal of Abnormal Psychology, 111,* 290–301.

Ruscio, J. (2002). *Documentation of program code to perform taxometric analyses and simulate comparison data in the R and S+ languages.* Retrieved November 22, 2002, from www.etown.edu/psychology/faculty/ruscio.htm

Ruscio, J., Haslam, N., & Ruscio, A. M. (2003). *Introduction to the taxometric method: A practical guide.* Manuscript in preparation.

Ruscio, J., & Ruscio, A. M. (2000). Informing the continuity controversy: A taxometric analysis of depression. *Journal of Abnormal Psychology, 109,* 473–487.

Ruscio, J., & Ruscio, A. M. (2002). A structure-based approach to psychological assessment: Matching measurement models to latent structure. *Assessment, 9,* 5–17.

Ruscio, J., & Ruscio, A. M. (in press). A nontechnical introduction to the taxometric method. *Understanding Statistics.*

Ruscio, J., & Ruscio, A. M., & Keane, T. M. (2004). Using taxometric analysis to distinguish a small latent taxon from a latent dimension with positively skewed indicators: The case of Involuntary Defeat Syndrome. *Journal of Abnormal Psychology, 113,* 99–108.

Ruscio, J., Ruscio, A. M., & Meron, M. (2003). *Simulating comparison data to test the suitability of data for taxometric analysis.* Manuscript submitted for publication.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1,* 1–26.

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., Jr., Mislevy, R. J., Steinberg, L., et al. (1990). *Computerized adaptive testing: A primer.* Hillsdale, NJ: Erlbaum.

Waldman, I. D., & Lilienfeld, S. O. (2001). Applications of taxometric methods to problems of comorbidity: Perspectives and challenges. *Clinical Psychology: Science and Practice, 8,* 520–527.

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

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58,* 236–244.

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

Widiger, T. A., & Clark, L. A. (2000). Toward *DSM–V* and the classification of psychopathology. *Psychological Bulletin, 126,* 946–963.

Received November 27, 2001
Revision received May 14, 2003
Accepted May 27, 2003 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://watson.apa.org/notify/ and you will be notified by e-mail when issues of interest to you become available!