A Structure-Based Approach to Psychological Assessment

Matching Measurement Models to Latent Structure

John Ruscio
Elizabethtown College

Ayelet Meron Ruscio
The Pennsylvania State University

The present article sets forth the argument that psychological assessment should be based on a construct's latent structure. The authors differentiate dimensional (continuous) and taxonic (categorical) structures at the latent and manifest levels and describe the advantages of matching the assessment approach to the latent structure of a construct. A proper match will decrease measurement error, increase statistical power, clarify statistical relationships, and facilitate the location of an efficient cutting score when applicable. Thus, individuals will be placed along a continuum or assigned to classes more accurately. The authors briefly review the methods by which latent structure can be determined and outline a structure-based approach to assessment that builds on dimensional scaling models, such as item response theory, while incorporating classification methods as appropriate. Finally, the authors empirically demonstrate the utility of their approach and discuss its compatibility with traditional assessment methods and with computerized adaptive testing.

Keywords: taxometrics, latent structure, measurement, classification, scaling

The latent structure of a psychological construct may be either taxonic (categorical, discrete, qualitative, latent class), dimensional (continuous, quantitative, latent factor, latent trait), or some combination of both. Although the importance of latent structure for measurement has been noted in the assessment literature (e.g., Meehl, 1992; Smith & McCarthy, 1995), there has not yet been a systematic effort to present the full range of possible latent structures, discuss how latent structure can inform the choice of measurement models, or articulate the implications of this choice for assessment. In the present article, we assert that the match—or mismatch—between the latent structure of a construct and the model by which that construct is measured affects the accuracy with which individuals are placed along a continuum or assigned to classes. We explore the consequences of this structure-model match for measurement error, statistical power, the search for an efficient cutting score, and statistical relations among constructs.

In what follows, we develop and illustrate the value of a comprehensive, structure-based approach to assessment by highlighting the critical role of latent structure in measurement. First, we explore differences between the latent and manifest levels of analysis and describe the possible types of latent structures. Next, we describe why it is important to match one's measurement approach to the latent

Portions of this article were presented at the annual meeting of the American Psychological Association, Washington, D.C., August 2000. This work was supported in part by National Research Service Award MH12675-01 to Ayelet Meron Ruscio from the National Institute of Mental Health. We would like to express our thanks to Paul E. Meehl and Leslie J. Yonce for access to the Hathaway Data Bank.
Correspondence concerning this article should be addressed to John Ruscio, Department of Psychology, Elizabethtown College, One Alpha Drive, Elizabethtown, PA 17022; e-mail: rusciojp@etown.edu.

Assessment, Volume 9, No. 1, March 2002 4-16
© 2002 Sage Publications

structure of the construct under investigation. We then briefly review various methods for empirically evaluating latent structure and suggest a generalized strategy for applying these methods. Finally, we outline a structure-based approach to assessment that builds on knowledge of latent structure by incorporating dimensional scaling and categorical classification methods as appropriate.

DISTINGUISHING MANIFEST AND LATENT STRUCTURE

The critical distinction between latent and manifest levels of analysis is seldom discussed in the assessment literature. Latent structure refers to the fundamental nature of a construct, the underlying structure that exists regardless of how one might choose to conceptualize or measure it. Manifest structure, in contrast, refers to characteristics associated with observable indicators of a construct, the surface structure that depends—among other things—on how one chooses to conceptualize and assess the construct. For a given construct, latent and manifest structure can differ (Grayson, 1987; Murphy, 1964). Meehl's (1962, 1990) theory of schizophrenia provides one example of how a latent category can give rise to manifest continua. The theory posits the existence of a single dominant gene that causes central nervous system deficits specific to schizophrenia. Those who inherit this gene develop schizotypy, a condition characterized by psychological and behavioral features such as cognitive slippage, social aversiveness, anhedonia, and ambivalence. Though signs such as these are distributed continuously at the manifest level, they have been found to correspond to a class of schizotypes at the latent level (Golden & Meehl, 1979; Korfine & Lenzenweger, 1995; Lenzenweger, 1999; Lenzenweger & Korfine, 1992; Tyrka et al., 1995).

In contrast, any construct that is continuous at the latent level can be made to appear categorical at the manifest level. One way in which this is often done is by applying a median split to a distribution of scores for the purpose of analytic convenience. For example, the continuous scores yielded by Rotter's (1966) Internal-External Locus of Control Scale are typically divided at the median to create groups with an internal or external locus of control, despite the possibility that this construct is continuous at the latent level. Another approach to categorization is demonstrated by the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994), which depicts all psychological disorders as latent taxa. The *DSM-IV* organizes mental disorders within diagnostic categories, each associated with a specific set of criteria that determine whether an individual is, or is not, disordered. Although these diagnostic categories may accurately reflect the taxonic structure of some psychological disorders, there is evidence that at least some of these categories mask underlying continua (A. M. Ruscio, Ruscio, & Keane, 2001; J. Ruscio & Ruscio, 2000). Thus, both statistical and conceptual categorization at the manifest level may correspond to continua at the latent level.

As these examples illustrate, a given manifest structure need not match the underlying latent structure of a construct—latent categories may give rise to an observed continuum, and a latent continuum may give rise to categorical measurements. We focus on the latent level of analysis in the present article due to the importance of understanding the true nature of a construct, regardless of how people have chosen to measure it.

For clarity and consistency with the literature on latent structure, we use the term *taxonic* to refer to a construct in which individuals or objects are separated into nonarbitrary classes, or taxa, at the latent level. That is, one or more qualitative boundaries "carve nature at its joints": Objects either do or do not belong to these taxa regardless of an observer's beliefs or preferences. By contrast, we use the term *dimensional* to refer to a construct along which individuals or objects differ only quantitatively, such that any classes that might be formed are arbitrary. Incontrovertible examples of latent taxa and dimensions exist in many sciences, though comparatively little research has explored the latent structure of psychological variables. Clear-cut examples of taxonic constructs include biological species, chemical elements, and subatomic particles, with representative taxa being the blue-ring octopus, magnesium, and the proton, respectively. Definitive examples of dimensional constructs include body mass, barometric pressure, and temperature, which are scaled along the continua of kilograms, millimeters of mercury, and degrees centigrade, respectively. "Obese" people, "high-pressure" weather, and "hot" objects are not naturally occurring categories but rather distinctions superimposed on dimensions for pragmatic purposes. Within psychology, preliminary evidence suggests that constructs such as psychopathy (Harris, Rice, & Quinsey, 1994), pathological dissociation (Waller, Putnam, & Carlson, 1996; Waller & Ross, 1997), and Type A personality (Strube, 1989) may be taxonic, whereas constructs such as adult attachment (Fraley & Waller, 1998), depression (A. M. Ruscio & Ruscio, 2001; J. Ruscio & Ruscio, 2000), and worry (A. M. Ruscio, Borkovec, & Ruscio, 2001) may be dimensional. Although this is only a partial listing of psychological constructs whose latent structure has been investigated, the overwhelming majority of variables of interest to psychologists have not been studied. Thus, there is an acute need for research that empirically evaluates latent structure using powerful analytic techniques designed expressly for this purpose.

Although our definitions of taxa and dimensions are standard, it is misleading to imply that a construct must be either taxonic or dimensional because structural combinations are possible. This point has been alluded to elsewhere (e.g., Waller & Meehl, 1998) but not elaborated in a rigorous way. In addition to the relatively simple latent structures of pure taxa (latent classes containing individuals whose manifest scores differ only due to measurement error) and pure dimensions (latent continua along which there are no qualitative boundaries), many hybrid latent structures are theoretically possible. This occurs in cases where a latent class or dimension can itself be broken down into additional latent classes and/or dimensions. Thus, the assessment of latent structure can be conceived as a hierarchical, iterative process in which constituent taxa or dimensions are sought until no further subdivisions of any kind are possible. Ultimately, stopping points will be reached whenever a pure dimension (one that is indivisible into constituents) or a pure taxon (one with no reliable residual variation) is uncovered.

To illustrate some of the possibilities, several hypothetical latent structures for the construct of depression are depicted in Figure 1. Panel A depicts depression as having no qualitative boundaries whatsoever, a pure dimension. Panel B shows depression divided into two latent classes, with no reliable residual variation within either latent class. Panel C shows a simple combination: There is one pure latent class, whereas the other class contains reliable residual variation and is thus a dimension. Panel D depicts a more complex combination: One latent class consists of three subtypes, and the other is a dimension. There are, of course, far more possibilities than the small sampling presented here.

Although a recent literature review (Flett, Vrendenburg, & Krames, 1997) and empirical investigations (A. M. Ruscio & Ruscio, 2001; J. Ruscio & Ruscio, 2000) suggest that depression may best be represented by a single dimension, the nature and complexity of most psychological constructs remain an unexplored empirical question. Waller et al. (1996) noted that as psychologists, "we too often presuppose that our data are unquestionably scaleable along latent dimensions or latent traits (factors or continua)" (p. 317). Dahlstrom (1995), Gangestad and Snyder (1985), and Meehl (1992, 1995) also discussed strong biases against latent taxa. Although structure is often presumed to match the manifest measurement scale employed or the presupposition of the researchers, the determination of a construct's true latent structure poses an empirical question that can be addressed using appropriate methods.

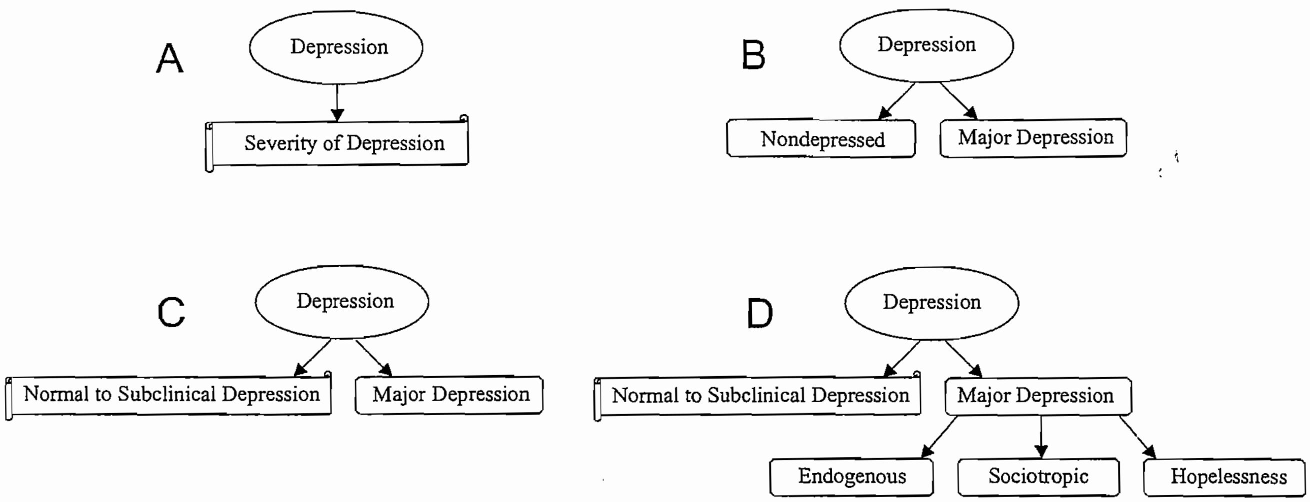
LATENT STRUCTURE AND PSYCHOLOGICAL ASSESSMENT

Understanding latent structure has significant implications for psychological assessment. A measurement model based on dimensional scaling will best locate an individual's position along a continuum, whereas a measurement model based on classification into taxa will best assign individuals to groups. As Meehl (1992) has noted, these disparate measurement goals lead to considerably different assessment guidelines and approaches, making an appropriate match between latent structure and measurement model particularly important.

When assessing a latent dimension, the goal of measurement is to most precisely determine the value of each individual's true score (in classical test theory) (Guilford, 1954; Gulliksen, 1950) or latent trait (in item response theory [IRT]) (Embretson, 1996; Hambleton, Swaminathan, & Rogers, 1991; Lord, 1980). In this context, any model that classifies individuals into groups is inappropriate. Dimensional measurement of a dimensional construct results in maximal measurement precision and statistical power, whereas spurious classification may have devastating consequences. Cohen (1983) has shown that when computing the statistical association between two continuous variables, dichotomizing one of them throws away 36% of the systematic variance, whereas dichotomizing both of them throws away nearly 60% of the systematic variance. In this way, research employing categorical diagnoses to study the comorbidity among psychological disorders may dramatically underestimate the co-occurrence of conditions that exist along a latent continuum. Indeed, preliminary evidence of dimensional structure for several depressive and anxiety disorders (A. M. Ruscio, Ruscio, & Keana, in press; J. Ruscio & Ruscio, 2000) suggests that this weakening of statistical power may systematically distort our understanding of the controversial relationship between these constructs (e.g., Clark & Watson, 1991; Foa & Foa, 1982; Maser & Cloninger, 1990). The rise in Type II errors associated with decreased power led Fraley and Waller (1998) to argue that spurious classification can cripple a field of research in the long run.

Another deleterious effect of spurious classification is that it may alter—not just weaken—statistical relations and inferred theoretical links between constructs. For example, the common practice of measuring adult attachment styles by the popular three-group scheme (secure, insecure-avoidant, and anxious-ambivalent) (Ainsworth, Blehar, Waters, & Wall, 1978), rather than by the two dimensions suggested by an examination of latent structure (anxiety and avoidance) (Fraley & Waller, 1998), may account for the alleged temporal instability of attachment

FIGURE 1
Four Hypothetical Ways in Which the Construct of Depression Might Be Broken Down Into Its Latent Structure (with scrolls representing dimensions and boxes representing taxa)



NOTE: Panel A shows depression as one dimension. Panel B shows depression as two latent classes. Panel C shows depression as one dimension plus a latent class. Panel D shows depression as one dimension plus a latent class with three subtypes.

styles (Baldwin & Fehr, 1995). That is, the standard error of difference scores will cause some individuals to be classified differently over time. Moreover, forcing latent dimensions into taxa will greatly increase error by throwing away meaningful variation in scores. Because a majority of published studies on adult attachment have superimposed a typological measurement scheme on the data, this literature may be in need of both reanalysis and reconceptualization (Fraley & Waller, 1998). Similarly, arbitrary categorization of data guided by communicative convenience or preference for a particular analytic strategy (e.g., ANOVA rather than regression) has likely weakened the strength and even distorted the form of statistical relations in many other domains of psychological research.

Whereas the classification of latent dimensions into groups results in a considerable loss of information, proper classification of latent taxa has been suggested (Meehl, 1992) and twice demonstrated (Gangestad & Snyder, 1985; Strube, 1989) to yield stronger relationships between taxa and other variables than measurement using dimensional scaling. This is because, for pure taxa, any variance in observed scores around the true scores of the taxa must be measurement error. Thus, applying a dimensional measurement model can increase error when taxa exist. The general conditions under which categorical classifications outpredict dimensional scales in the presence of latent taxa remain an important open question (see Grove, 1991b).

Finally, there is an additional advantage to classification models that is seldom addressed in the assessment literature: They assist users in locating an efficient cutting score for classifying cases into taxa. Even in the presence of latent taxa, dimensional scaling models typically yield unimodal distributions of manifest scores. Without any natural breaks in such a distribution, it is quite challenging to determine an appropriate cutting score for separating individuals into groups, and the efficiency of classification drops off rapidly with suboptimal choices. Classification models, on the other hand, yield a strongly bimodal distribution of manifest scores for latent taxa. With such a distribution, one can clearly identify an efficient cutting score by locating the low point toward the center of the distribution. Moreover, with so few cases in nearby regions of the distribution, the efficiency of classification is highly robust to the selection of suboptimal cutting scores.

In sum, there are a number of practical advantages associated with a structure-based approach to psychological assessment. Therefore, we turn now to the first step of such an approach: determining the latent structure of the construct of interest.

METHOD FOR DETERMINING LATENT STRUCTURE

Because measurement models presume a latent structure that is not directly observable at the manifest level, it is critical that latent structure be evaluated using methods expressly designed for this purpose. Among the presently available techniques, we believe that Meehl's (1995, 1999) taxometric method is the most promising. Below, we briefly outline the logic of this taxometric method, provide an overview of several procedures that constitute the method, and compare the method to the available alternative approaches.

Logic of the Taxometric Method

Meehl (1973, 1995, 1999) and his colleagues (Golden & Meehl, 1979; Grove & Meehl, 1993; Meehl & Golden, 1982; Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998) have pioneered the development of a family of taxometric procedures based on the principles of coherent cut kinetics. Most procedures within the method search for orderly statistical relations between one or more variables along sliding intervals, or cuts, of another. Each procedure uses manifest indicators to search for a qualitative boundary between two latent taxa, traditionally referred to as the "taxon" and "complement." Meehl's taxometric method relies on the convergence of evidence obtained from multiple, quasi-independent analytic procedures—rather than on traditional null hypothesis significance tests—to provide clues to latent structure. Each procedure serves as a consistency check for the results provided by the others, with confidence in a structural solution increasing as each additional consistency test is passed.

Although procedures in the taxometric method directly test only a two-group latent class model, investigators can resolve more complex latent structures by combining these procedures with psychometric analyses (e.g., evaluating unidimensionality, homogeneity, and internal consistency) and applying them in an iterative fashion. For example, consider the latent structure depicted in Panel D of Figure 1. With the proper selection of indicator variables, an initial taxometric analysis would indicate that there was a qualitative boundary between the taxon (major depression) and the complement (normal to subclinical depression). Subsequent taxometric analyses within the complement class would fail to reveal any additional taxa; psychometric analyses would reveal the reliable residual variation of a single dimension. However, a series of subsequent taxometric analyses within the taxon, using new sets of indicators specific to the conjectured subtypes, would uncover three subtypes of major depression, and psychometrics would show no reliable residual variation within them. Taken together, these steps represent an idealization of a careful program of systematic research essential for the comprehensive understanding and appropriate assessment of any psychological construct.

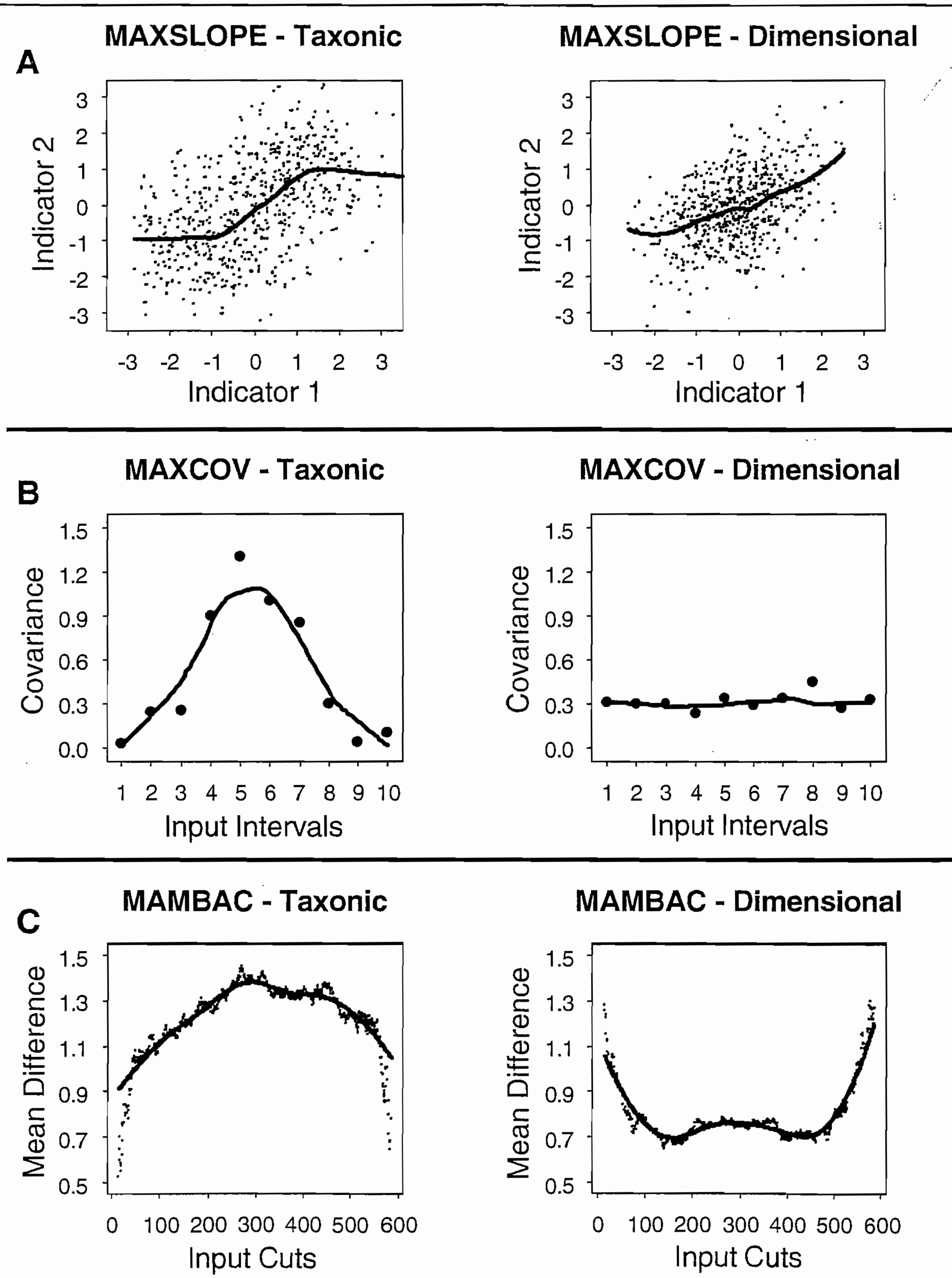
Procedures in the Taxometric Method

Although many taxometric procedures have been developed and validated, only a few of the conceptually simplest techniques will be presented here in the interest of conserving space. Whereas each of the procedures described below can be used alone to provide a structural solution, they are more appropriately used in tandem, with each procedure serving as a consistency test that checks the conclusions of the others. We focus on a conceptual presentation of the method so that it can be compared to better known alternatives, and we illustrate each taxometric procedure using three continuously distributed indicator variables—competitiveness, time urgency, and hostility—that have been suggested to distinguish individuals with Type A personality from those with Type B personality (Friedman & Rosenman, 1974), which research suggests is taxonic (Strube, 1989). Readers interested in more detailed treatments of Meehl's taxometrics, including descriptions of powerful multivariate procedures in the method, should consult Meehl and Golden (1982), J. Ruscio and Ruscio (2001), Waller and Meehl (1998), and the references cited below.

Maximum slope (MAXSLOPE). MAXSLOPE (Grove & Meehl, 1993) examines the slope of a local regression across a scatterplot of two indicators of the conjectured latent taxa. For example, suppose that two manifest indicators, such as time urgency and hostility, were plotted for a sample containing 50% Type A and 50% Type B individuals. Because these indicators are unrelated to one another within personality types, the local regression will be fairly flat in both regions dominated by one particular type—the upper right (composed mostly of Type A individuals) and the lower left (composed mostly of Type B individuals). There will be a positive slope toward the center of the scatterplot due to the mixture of personality types in that region. This slope will reach a maximum where the groups intersect. Hence, taxonic latent structure yields a steplike or S-shaped curve. For dimensional latent structure, there would be a fairly constant positive slope across the entire scatterplot, yielding a comparably straight line. Panel A of Figure 2 presents sample MAXSLOPE plots for both latent structures.

Maximum covariance (MAXCOV). MAXCOV (Meehl, 1973; Meehl & Yonce, 1996) is also based on the statistical behavior of indicators in the vicinity of group mixture.[1] Suppose that a sample of Type A and Type B individuals is

FIGURE 2
Sample Curves



NOTE: MAXSLOPE = maximum slope; MAXCOV = maximum covariance; MAMBAC = mean above minus below a cut. Taxonic (n = 600, base rate = .50, 2.00σ separation) and dimensional (n = 600, $\bar{r}_{ij}$ = .50) latent structures analyzed using three different taxometric procedures. Each graph contains a smoothed line generated by the locally weighted scatterplot smoother method.

divided into successive intervals according to their level of competitiveness (referred to as the input indicator) and that the covariance of the two remaining indicators—time urgency and hostility (referred to as output indicators)—is calculated within each interval. Intervals demarcating relatively low levels of competitiveness will contain mostly Type B individuals, whereas those demarcating relatively high levels of competitiveness will contain mostly Type A individuals. Thus, at either extreme of the input scale, the covariance between time urgency and hostility will approach zero. Covariance values will be higher within more centrally located input intervals that correspond to moderate competitiveness, reaching a maximum in the input interval containing an equal mixture of Type A and Type B individuals. Hence, in the MAXCOV procedure, taxonic latent structure yields a peaked curve. For dimensional structure, relatively constant positive covariances would be observed across all intervals of the input indicator, yielding a comparably flat line. Panel B of Figure 2 presents sample MAXCOV curves.

Mean above minus below a cut (MAMBAC). MAMBAC (Meehl & Yonce, 1994) is based on the fact that if latent taxa exist, there will be an optimal cutting score on any valid indicator for classifying individuals into these taxa. Suppose that cases are sorted along an input indicator, such as competitiveness, and that the efficiency of all possible cutting scores on this indicator is examined. To do this, means are computed on an output indicator, such as hostility, separately for cases falling above and below each cut. A MAMBAC curve is constructed by plotting the difference between hostility means above and below each cut on the competitiveness indicator. Latent taxa generate a curve that is peaked near the cutting score that best distinguishes the classes (e.g., with Type A individuals falling above the cut and Type B individuals falling below the cut), whereas latent dimensions generate comparably dish-shaped curves. Panel C of Figure 2 presents sample MAMBAC curves.

Two additional features of taxometric procedures are worthy of note. First, each procedure can be conducted using available indicators in all possible input and output combinations, permitting examination of the consistency of results across combinations. For example, if four indicators are available, each of these taxometric procedures can be performed 12 times, and a panel of graphs can be plotted for interpretation.[2] Second, each procedure can be used to estimate latent parameters such as the base rate of taxon membership (e.g., the proportion of Type A individuals) in the sample under investigation. These estimates can then be compared for consistency within and between procedures as further tests of the existence of taxa.

Comparison With Alternative Procedures

There exist other procedures for examining latent structure, most notably distributional analyses (e.g., inspection for bimodality or negative kurtosis, admixture analysis, commingling analysis), cluster analysis, and approaches that model the relationship between manifest and latent variables (latent class analysis, latent profile analysis, latent trait analysis, and factor analysis). A number of important limitations, however, render each of these procedure less effective than Meehl's taxometric method for empirically distinguishing latent taxa from dimensions.

Distributional analysis. There is a variety of ways in which a manifest distribution can be examined for clues to latent structure. One method is to look for bimodal or multimodal distributions (e.g., Harding, 1949), which are suggestive of latent taxa. However, even in the clearest case (two equal-sized groups), the individual distributions must differ by at least two within-group standard deviations before a visible dip emerges toward the center of the joint distribution and two modes become apparent (Murphy, 1964). Groups that are separated by lesser amounts might instead form a unimodal distribution that is flattened relative to the normal curve, yielding a negative kurtosis. Other methods, such as admixture or commingling analyses (e.g., MacLean, Morton, Elston, & Yee, 1976), use trial and error to determine the parameters of hypothetical subgroup distributions that, when combined, would generate the observed distribution.

The primary difficulty with each of these procedures is that as noted earlier, manifest structure need not—and often does not—correspond to latent structure. For example, a scale containing items of equal difficulty and steep discrimination will tend to yield a manifest bimodal distribution, regardless of the latent structure of the construct being assessed. By contrast, a scale containing items of widely varying difficulties will tend to yield a unimodal distribution regardless of latent structure (Grayson, 1987). Many other factors can also alter the relationship between latent and manifest structure, thereby undermining the results of any procedure that simply analyzes a manifest distribution (see Grayson, 1987, and Murphy, 1964, for extended discussions of these limitations). Finally, these approaches do not provide an independent means of checking the structural conclusions that they produce, as do the consistency tests of the taxometric method.

Cluster analysis. The procedures in this large analytical family seek to determine whether cases tend to cluster together in a multidimensional hyperspace (e.g., Sneath & Sokal, 1973; Sokal & Sneath, 1963). There are a tremendous number of clustering algorithms available, all sharing two common characteristics. First, some measure of similarity (or distance) is chosen to quantify the relations between all cases in a sample. Second, some mathematical rule is applied to parse these similarity values into clusters.

Several factors limit the ability of cluster analysis to distinguish taxonic from dimensional latent structure. For example, there is often no reliable way to determine the appropriate number of clusters (Grove, 1991a). This problem is compounded by an even greater concern: Most algorithms will always uncover clusters in the data, even if the latent structure is dimensional (see Grove & Andreasen, 1989; Meehl, 1979, 1992; and references contained therein for more detailed treatments of these and related issues). Even simply rearranging the rows in a data set can substantially alter the clusters produced by the many algorithms in which the order of cases determines how clusters are initialized. Moreover, in contrast to the role of independently derived consistency tests in the taxometric method, researchers seldom employ multiple clustering algorithms, and the handful of algorithms that predominate in psychological research (see Blashfield, 1976, 1984) seldom yield results that are consistent with one another (Golden & Meehl, 1980). Thus, although cluster analyses may be useful for classifying cases within a validated taxonomy, there is insufficient support for their use as tools to determine latent structure.

Latent class analysis and related approaches. A final family of four conceptually related analytic techniques models the association between manifest and latent variables: latent class analysis (e.g., Green, 1951; Lazarsfeld & Henry, 1968), latent profile analysis, latent trait analysis (e.g., IRT) (Embretson, 1996; Hambleton et al., 1991; Lord, 1980), and factor analysis (e.g., Gorsuch, 1983; Thurstone, 1935, 1947). These procedures differ according to the structure of the manifest variables that they analyze and the presumed structure of the inferred latent variable(s). Factor analysis, for example, is typically used to reduce a large number of continuously distributed items to a smaller number of latent factors that are nearly always presumed to be dimensional in nature. Latent class analysis is a categorical analogue of factor analysis, reducing a large number of manifest categories to a smaller set of latent categories. Latent profile analysis and latent trait analysis are, in a sense, hybrid procedures: The former uses manifest continua to infer latent categories, whereas the latter uses manifest categories to infer latent continua.

Although each of these procedures can provide valuable information when used for either exploratory (data reduction) or confirmatory (testing a conjectured latent structure) purposes, none is ordinarily employed to test the competing hypotheses of taxonic and dimensional latent structure. For example, Waller and Meehl (1998) noted that despite passages in Thurstone's (1935, 1947) classic treatises on factor analysis dealing with the possibility of categorical factors, it is usually presumed that factors represent latent dimensions.[3] Moreover, none of these methods makes use of multiple consistency tests to help identify faulty conclusions. Thus, like cluster analysis, these four procedures may be of greater value once latent structure has been established as either taxonic or dimensional in nature.

Conclusions

Each of the procedures described above has characteristics that limit its ability to distinguish taxa from dimensions at the latent level. The real test of any method, however, lies in empirical evaluations of its efficacy. A considerable body of research using Monte Carlo simulations (Cleland & Haslam, 1996; Cleland, Rothschild, & Haslam, 2000; Haslam & Cleland, 1996; Meehl, 1973; Meehl & Golden, 1982; Meehl & Yonce, 1994, 1996; J. Ruscio, 2000) and "pseudo problems" (e.g., evaluating known latent structures such as that of biological sex using empirical data) (Gangestad & Snyder, 1985; Korfine & Lenzenweger, 1995; Meehl & Golden, 1982; Trull, Widiger, & Guthrie, 1990) has demonstrated the ability of taxometric procedures to accurately distinguish taxonic from dimensional latent structure. Despite decades of research, none of the alternative methods developed for evaluating latent structure has achieved this level of success in Monte Carlo or pseudoproblem trials (cf. Meehl & Golden, 1982).

CLASSIFICATION USING BAYESIAN PROBABILITIES OF TAXON MEMBERSHIP

Once the latent structure of a construct has been established, the next step is to use this knowledge to determine which measurement model is most appropriate for the construct. The most widely used measurement models in psychological assessment, particularly the sophisticated IRT models of recent vintage, are premised on the existence of latent dimensions. However, if a construct is taxonic, it makes little sense to plot item characteristic curves along the values of a latent dimension. Instead, methods are needed to classify individuals into taxa, a task for which classification models such as Bayes's theorem are eminently well suited.

Armed with an estimate of the taxon base rate in one's sample, as well as the valid and false positive rates of available indicators, one can calculate the probability that an individual belongs to a taxon given his or her response pattern—$Pr(t|RP)$—using the following formula (Waller & Meehl, 1998, p. 29):

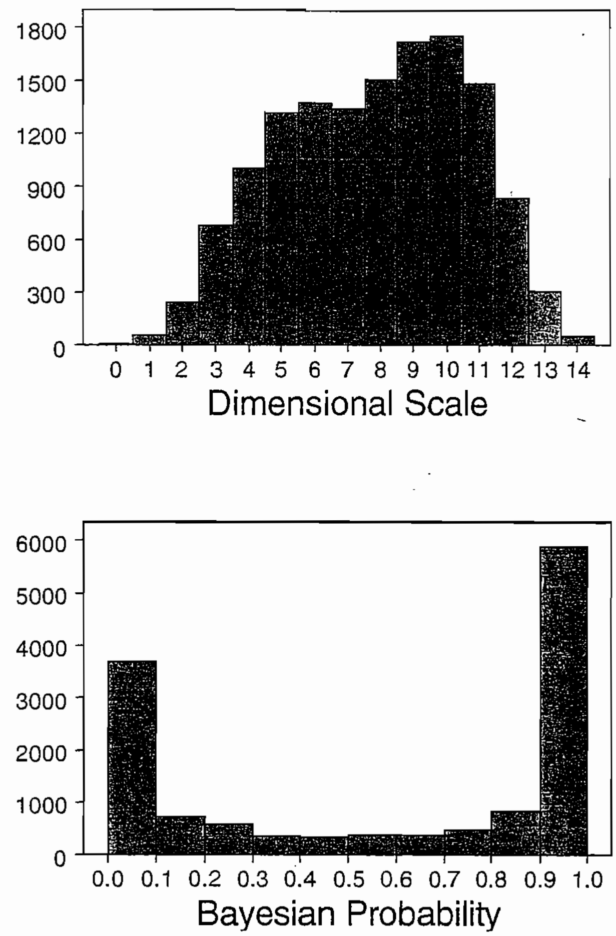
$$Pr(t|RP) = \frac{P \prod_{i=1}^{v} pt_i^{\theta} qt_i^{1-\theta}}{P \prod_{i=1}^{v} pt_i^{\theta} qt_i^{1-\theta} + Q \prod_{i=1}^{v} pc_i^{\theta} qc_i^{1-\theta}} \quad (1)$$

where Π is the cumulative product operator, P is the base rate of taxon membership in the relevant population, $Q = 1 - P$, v is the number of indicators, pt_i is the valid positive rate achieved by each indicator, $qt_i = 1 - pt_i$, pc_i is the false positive rate achieved by each indicator, $qc_i = 1 - pc_i$, and $\theta = 1$ for a positively keyed response on an indicator, 0 otherwise.

To illustrate the power of this Bayesian model, consider its application to the pseudo problem of classifying the sexes. We use this as our example because the latent structure of biological sex is indisputable: It consists of two latent taxa, men and women. At the same time, the availability of a large data set with multiple indicators of biological sex allowed us to compare the efficacy of IRT and Bayesian models for classification. Using data from the Hathaway Data Bank (see J. Ruscio & Ruscio, 2000, or Waller, 1999, for descriptions of this database), 14 Minnesota Multiphasic Personality Inventory (MMPI) items from the Masculinity-Femininity Scale (Mf) having high corrected item-total correlations and varying difficulty levels were used to classify 13,684 adults—8,056 women (keyed as the taxon) and 5,628 men (keyed as the complement)—according to their sex. The MMPI Mf items were summed to yield a 15-point (0 to 14) dimensional scale on which each individual received a score. In addition, each individual's probability of taxon membership was calculated according to Bayes's theorem using the formula above. Thus, the traditional approach of dimensional scaling was compared to a classification model using the same set of indicator variables (MMPI items) in handling a construct with taxonic latent structure.

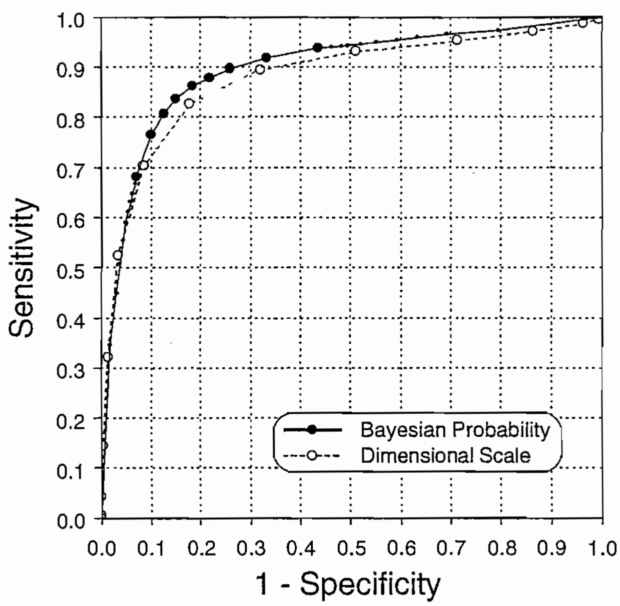
As can be seen in Figure 3, dimensional scale scores were unimodally distributed, with latent taxonicity almost completely obscured at the manifest level. The distribution of Bayesian probabilities, on the other hand, displayed a striking bimodality. To examine the efficiency with which individuals' sex could be classified by various cutting scores along the distributions, receiver operating characteristic (ROC) curves were plotted (see Figure 4). Bayesian probabilities achieved slightly greater accuracy (area under ROC curve = .897, confidence interval [CI] [95%] = .891 to .902) than did the dimensional scale (area = .879, CI [95%] = .873 to .885). Expressed as hit rates, the optimal cutting score along the distribution of Bayesian probabilities correctly classified 84.3% of all cases, whereas the optimal cut along the dimensional scale correctly classified 82.5% of all cases.

FIGURE 3
Frequency Distributions for the Dimensional Scale and Bayesian Probabilities



Although Bayesian classification yielded slightly better accuracy, the primary advantage of this approach was the extent to which it facilitated the location of an efficient cutting score to separate the latent classes. The pileup of cases at intermediate values along the dimensional scale made it difficult to choose an efficient cutting score. Moreover, this scale offered little tolerance for a suboptimal choice, as is evidenced by the wide spacing of successive cuts (open circles) on the ROC curve. In sharp contrast, it made relatively little difference where the bimodal distribution of Bayesian probabilities was cut. Cuts made anywhere from .10 to .90 (large dark circles) resulted in closely adjacent points on the ROC curve. In fact, all cutting scores between .30 and .70 on the Bayesian distribution achieved greater hit rates than did the optimal cut along the dimensional scale. This clearly illustrates that the selection of a cutting score is greatly simplified by use of a classification model when latent taxa exist.

FIGURE 4
ROC Curves for the Dimensional Scale and Bayesian Probabilities



NOTE: ROC = receiver operating characteristic. Open circles (dashed lines) represent the accuracy achieved through all cutting scores along the dimensional scale. Solid circles (solid lines) represent the accuracy achieved through cutting the distribution of Bayesian probabilities, with the nine cutting scores of .10 through .90 plotted as large points and cutting scores in increments of .01 out to the extremes of 0 and 1 plotted as small points.

The above demonstration indicates that when taxa are present, calculating Bayesian probabilities of taxon membership affords the simultaneous advantages of distributional continuity and bimodality. Continuity is useful in the event that situational demands call for the optimization of an index other than the overall hit rate of classification, allowing the selection ratio to be altered as desired to trade sensitivity for specificity or vice versa (see Meehl & Rosen, 1955). In this case, the Bayesian model provided much finer discriminations than did the summed scale scores, which yielded only 15 unique scores (0, 1, 2, . . . , 14). At the same time, bimodality simplifies the selection of an efficient cutting score and protects against a suboptimal choice.

Thus, we contend that the conventional preference for dimensional measurement in psychological research and the oft-heard claim that dimensions "retain more information" may be overly simplistic. An accurate classification of cases is much harder to achieve when items are combined using a dimensional scaling technique than when they are combined using a categorical measurement model. In the absence of empirical evidence regarding the latent structure of a given construct, it remains an open question whether a dimensional scaling or a classification approach would afford greater utility for psychological assessment, making the evaluation of latent structure particularly important.

IMPLEMENTATION OF A STRUCTURE-BASED APPROACH TO PSYCHOLOGICAL ASSESSMENT

Because latent structure is so important, we envision the rigorous development of psychological assessment devices beginning with a careful examination of the latent structure of each construct to be assessed. This would include delineation of all taxa (types and subtypes) through iterative applications of the taxometric method and evaluation of all dimensions through more conventional psychometric methods. We encourage readers to combine measurement models as suggested by empirical analysis of their constructs' latent structures, incorporating dimensional scaling and classification models as appropriate into a comprehensive assessment package. Whenever a distinction between taxa must be made, the relevant Bayesian probabilities can be calculated and used for classification. Whenever a continuum is encountered, a dimensional scaling model can be used to estimate individuals' scores along the continuum. This approach avoids the pitfalls stemming from the mismatch of latent structures and measurement models.

In recent years, technological developments have facilitated the computerized administration and scoring of psychological tests. Especially noteworthy is the rapidly expanding area of computerized adaptive testing (CAT), a highly desirable assessment method for reasons of brevity, reduced fatigue, elimination of hand-scoring errors, and immediacy of results (Embretson & Herschberger, 1999; Wainer et al., 1990). As has often been noted (e.g., Embretson & Herschberger, 1999), IRT models for dimensional scaling lend themselves well to implementation via CAT. Using CAT to implement an IRT model allows all scores on the latent trait to be estimated with equal precision through the administration of a custom-tailored subset of available items to each individual. However, although they are frequently paired in the literature (e.g., Embretson, 1996), IRT models and CAT interfaces are separable: Many measurement models can be implemented using the general strategies of CAT. Unfortunately, in keeping with psychologists' pervasive presumption of latent dimensionality, the use of CAT for classification is seldom discussed in the psychological assessment literature.[4]

Despite any apparent dissimilarity, there is a straightforward conceptual analogy between Bayesian classification and IRT models. To calculate an individual's score along a latent trait, an IRT model begins with an initial estimate of the trait score and refines it through responses to a set of items with known item characteristic curves. Bayesian classification begins by using the taxon base rate as an initial estimate of the probability of taxon membership and updates it through responses to a set of items with known valid and false positive rates. When IRT models are implemented using CAT, programmed algorithms guide item selection according to criteria such as content coverage, often involving the administration of a minimal number of items and/or the achievement of a certain standard error of measurement. Using a CAT interface, Bayesian classification could also proceed by selecting items according to content coverage, administering items of minimal redundancy until a threshold of high or low probability of taxon membership is crossed. At this point, the individual case would be classified into the taxon or complement, respectively. Thus, although our structure-based approach to assessment can be implemented using traditional paper-and-pencil methods, as can IRT or Bayesian models alone, CAT can be used to perform both scaling and classification functions. Regardless of the mode of implementation that is judged most feasible in any given assessment context, we suggest that measurement models should be chosen to best match the latent structure of the psychological construct being assessed.

NOTES

1. This procedure is conceptually quite similar to maximum slope (MAXSLOPE), which has been recommended as a "MAXCOV [maximum covariance] surrogate" when only two indicators are available (P. E. Meehl, personal communication, October 26, 1998). Another similar procedure is maximum eigenvalue (MAXEIG) (Waller & Meehl, 1998), a multivariate extension of the MAXCOV procedure.

2. MAXSLOPE and mean above minus below a cut (MAMBAC) can be performed twice using each pairwise combination of indicators by swapping each pair of indicators on the x and y axis for MAXSLOPE and switching the input and output for MAMBAC. For $k \geq 2$ indicators, one can calculate $k(k-1)$ MAXSLOPE or MAMBAC curves. MAXCOV can be performed thrice using each three-way combination of indicators by treating each member of the triplet as the input in turn. For $k \geq 3$ indicators, one can calculate $k(k-1)(k-2)/2$ MAXCOV curves.

3. Waller and Meehl (1998) have developed L-Mode, a taxometric procedure that uses elements of factor analysis to distinguish taxonic from dimensional latent structure. The procedure works by examining the number of latent modes in the distribution of true scores on the first principal factor derived from a factor analysis of all available indicator variables. Unimodality is suggestive of dimensionality, whereas bimodality is suggestive of latent taxa.

4. For example, Embretson (2000) described 10 methodological frontiers in current research on psychological testing, all of which involved refinements of computerized adaptive testing and/or item response theory (IRT) models. Only 1 of the 10 refinements touched on qualitative distinctions between individuals: IRT models are being developed to incorporate categorical data as input. Nonetheless, even these models presume that the underlying construct is structured as a latent dimension.

REFERENCES

Ainsworth, M.D.S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Lawrence Erlbaum.

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

Baldwin, M. W., & Fehr, B. (1995). On the instability of attachment style ratings. *Personal Relationships*, *2*, 247-261.

Blashfield, R. K. (1976). Mixture model tests of cluster analysis: Accuracy of four hierarchical agglomerative methods. *Psychological Bulletin*, *83*, 377-388.

Blashfield, R. K. (1984). *The classification of psychopathology: Neo-Kraeplinian and quantitative approaches*. New York: Plenum.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*, 316-336.

Cleland, C., & Haslam, N. (1996). Robustness of taxometric analysis with skewed indicators: I. A Monte Carlo study of the MAMBAC procedure. *Psychological Reports*, *79*, 243-248.

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

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. (2000, August). *Psychometric methods in the second century of testing*. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

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

Flett, G. L., Vrendenburg, K., & Krames, L. (1997). The continuity of depression in clinical and nonclinical samples. *Psychological Bulletin*, *121*, 395-416.

Foa, E. B., & Foa, U. G. (1982). Differentiating depression and anxiety: Is it possible? Is it useful? *Psychopharmacology Bulletin*, *18*, 62-68.

Fraley, R. C., & Waller, N. G. (1998). Adult attachment patterns: A test of the typological model. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 77-114). New York: Guilford.

Friedman, M., & Rosenman, R. H. (1974). *Type A behavior and your heart*. New York: Knopf.

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

Golden, R. R., & Meehl, P. E. (1979). Detection of the schizoid taxon with MMPI indicators. *Journal of Abnormal Psychology*, *88*, 217-233.

Golden, R. R., & Meehl, P. E. (1980). Detection of biological sex: An empirical test of cluster methods. *Multivariate Behavioral Research*, *15*, 475-496.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence 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. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1, pp. 313-329). Minneapolis: University of Minnesota Press.

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

Grove, W. M., & Andreasen, N. C. (1989). Quantitative and qualitative distinctions between psychiatric disorders. In L. N. Robins & J. E. Barrett (Eds.), *The validity of psychiatric diagnoses* (pp. 127-139). New York: Raven Press.

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

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

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

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

Harding, J. P. (1949). The use of probability paper for the graphical analysis of polymodal frequency distributions. *Journal of the Marine Biological Association, 28*, 141-153.

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. (1996). Robustness of taxometric analysis with skewed indicators: II. A Monte Carlo study of the MAXCOV procedure. *Psychological Reports, 79*, 1035-1039.

Korfine, L., & Lenzenweger, M. F. (1995). The taxonicity of schizotypy: A replication. *Journal of Abnormal Psychology, 104*, 26-31.

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

Lenzenweger, M. F. (1999). Deeper into the schizotypy taxon: On the robust nature of maximum covariance analysis. *Journal of Abnormal Psychology, 108*, 182-187.

Lenzenweger, M. F., & Korfine, L. (1992). Confirming the latent structure and base rate of schizotypy: A taxometric analysis. *Journal of Abnormal Psychology, 101*, 567-571.

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

MacLean, C. J., Morton, N. E., Elston, R. C., & Yee, S. (1976). Skewness in commingled distributions. *Biometrics, 32*, 695-699.

Maser, J. D., & Cloninger, C. R. (Eds.). (1990). *Comorbidity of mood and anxiety disorders*. Washington, DC: American Psychiatric Press.

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

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

Meehl, P. E. (1979). A funny thing happened to us on the way to the latent entities. *Journal of Personality Assessment, 43*, 564-581.

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., & 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., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194-216.

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.

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

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcements. *Psychological Monographs, 80*(Whole No. 609).

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. (in press). The latent structure of analogue depression: Should the BDI be used to classify groups? *Psychology Assessment*.

Ruscio, A. M., Ruscio, J., & Keane, T. M. (in press). The latent structure of posttraumatic stress disorder: A taxometric investigation of reactions to extreme stress. *Journal of Abnormal Psychology*.

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., & 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. (2001). *Distinguishing types from continua: A nontechnical introduction to the taxometric method*. Manuscript submitted for publication.

Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment, 7*, 300-308.

Sneath, P.H.A., & Sokal, R. R. (1973). *Numerical taxonomy*. San Francisco: Freeman.

Sokal, R. R., & Sneath, P.H.A. (1963). *Principles of numerical taxonomy*. San Francisco: Freeman.

Strube, M. J. (1989). Evidence for the *type* in Type A behavior: A taxometric analysis. *Journal of Personality and Social Psychology, 56*, 972-987.

Thurstone, L. L. (1935). *The vectors of mind*. Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

Trull, T. J., Widiger, T. A., & Guthrie, P. (1990). Categorical versus dimensional status of borderline personality disorder. *Journal of Abnormal Psychology, 99*, 40-48.

Tyrka, A. R., Cannon, T. D., Haslam, N., Mednick, S. A., Schulsinger, F., Schulsinger, H., et al. (1995). The latent structure of schizotypy: I. Premorbid indicators of a taxon of individuals at risk for schizophrenia-spectrum disorders. *Journal of Abnormal Psychology, 104*, 173-183.

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: Lawrence Erlbaum.

Waller, N. G. (1999). Searching for structure in the MMPI. In S. Embretson & S. Herschberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 185-217). Mahwah, NJ: Lawrence Erlbaum.

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

Waller, N. G., Putnam, F. W., & Carlson, E. B. (1996). Types of dissociation and dissociative types: A taxometric analysis of dissociative experiences. *Psychological Methods, 1*, 300-321.

Waller, N. G., & Ross, C. A. (1997). The prevalence and biometric structure of pathological dissociation in the general population: Taxometric and behavior genetic findings. *Journal of Abnormal Psychology, 106*, 499-510.

John Ruscio, Ph.D., is an assistant professor of psychology at Elizabethtown College. His primary areas of interest include the taxometric method and its applications, clinical judgment, statistical decision making, and conceptual and practical issues in diagnosis and classification.

Ayelet Meron Ruscio, M.S., is a doctoral candidate in clinical psychology at the Department of Psychology at Pennsylvania State University. Her interests include the nature and origins of normal and pathological anxiety, as well as general issues in the assessment, classification, and differential diagnosis of mental disorders.