Journal of Abnormal Psychology
2004, Vol. 113, No. 1, 145–154

0021-843X/04/$12.00 DOI: 10.1037/0021-843X.113.1.145

SHORT REPORTS

Using Taxometric Analysis to Distinguish a Small Latent Taxon From a Latent Dimension With Positively Skewed Indicators: The Case of Involuntary Defeat Syndrome

John Ruscio
Elizabethtown College

Ayelet Meron Ruscio
Pennsylvania State University

Terence M. Keane
Veterans Affairs Boston Healthcare System and Boston University School of Medicine

Joining the debate on the structure of depression, S. R. H. Beach and N. Amir (2003) analyzed college students' responses to 6 Beck Depression Inventory (BDI) items with predominantly somatic content and concluded that they identified a small latent taxon corresponding to involuntary defeat syndrome. An exact replication of these analyses yielded virtually identical taxometric results, but parallel analyses of simulated taxonic and dimensional comparison data matching the intercorrelations and skewed distributions of the BDI items showed the results to be more consistent with dimensional than with taxonic latent structure. Analyses in a clinical sample with nonskewed indicators further supported a dimensional interpretation. The authors discuss methodological strategies for conducting and interpreting taxometric analyses under the adverse conditions commonly encountered in psychopathology research, including skewed indicators and small putative taxa.

Beach and Amir (2003) recently examined the latent structure of a depressive construct known as involuntary defeat syndrome (IDS), a state of homeostatic disruption and physical demobilization hypothesized to identify the depressed individual as nonthreatening to dominant others in competitive situations (Gilbert, 1992, 2000). Arguing that somatic symptoms better represent this depressive construct than cognitive and affective symptoms, Beach and Amir conducted taxometric analyses (see Meehl, 1995) of college students' responses to six somatic items on the Beck Depression Inventory (BDI; Beck, Rush, Shaw, & Emery, 1979) and concluded that they had identified a small latent taxon.

Beach and Amir's (2003) structural investigation of IDS was performed with data whose characteristics were typical of taxometric investigations of psychopathological constructs. These include a relatively small putative taxon, substantial indicator skew, and an indicator response scale that is not truly continuous. Although such data do not necessarily prohibit informative taxometric analysis, recent findings suggest that they may, under certain conditions, yield misleading results. In particular, positively skewed indicators of a latent dimension can produce rising taxometric curves that may be easily misinterpreted as evidence of a small taxon (A. M. Ruscio & Ruscio, 2002). This places an onus on clinical scientists, particularly those investigating rare clinical phenomena in nonclinical samples, to demonstrate that their data are able to differentiate taxonic from dimensional latent structure and that their conclusions reflect the true structure of the target construct rather than a methodological artifact.

Fortunately, there are data-analytic safeguards that one can take to protect against erroneous conclusions in taxometric research. In this article, we reexamine Beach and Amir's (2003) analyses as a springboard for the consideration of strategies for performing and interpreting taxometric analyses under the adverse conditions typical of psychopathology research.

Determining the Suitability of Data for Taxometric Analysis

Before undertaking taxometric analysis, it is incumbent upon an investigator to demonstrate that the available data are suitable for analysis. Suitability may be evaluated conceptually, with consideration for careful construct definition, sample appropriateness, and indicator selection. For example, it is important that the breadth and specificity of the indicators fully and uniquely represent the target construct and that the sample contain enough putative taxon members (see J. Ruscio & Ruscio, 2004). From this perspective, Beach and Amir's (2003) data had several limitations.

John Ruscio, Department of Psychology, Elizabethtown College; Ayelet Meron Ruscio, Department of Psychology, Pennsylvania State University; Terence M. Keane, National Center for Posttraumatic Stress Disorder—Behavioral Science Division, Veterans Affairs Boston Healthcare System, and Department of Psychiatry, Boston University School of Medicine.

This research was supported by National Research Service Award MH12675 to Ayelet Meron Ruscio from the National Institute of Mental Health.

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

In particular, it is unclear whether the six self-report questionnaire items adequately represented the IDS construct and whether the debilitating physiological symptoms of IDS could be adequately studied in an unselected college sample. Indeed, because this somatic form of depression appears similar to the severe and relatively rare melancholic subtype of major depressive disorder (Haslam & Kim, 2002) and because the point prevalence of major depressive disorder itself is quite low among college students, one would expect the point prevalence of IDS to be much lower than the 14% taxon base rate reported by Beach and Amir for their college student sample.

In addition to these conceptual considerations, data suitability should be evaluated empirically. For example, taxometric analyses require that indicators separate putative groups with sufficient validity and be correlated within these groups at tolerably low levels to afford an informative structural test. Although reporting interindicator correlations in the full sample and in extreme groups is a good first step, it does not indicate whether these parameters are sufficiently strong to counteract potentially problematic factors such as a small putative taxon and highly skewed indicators. Indicator skew was present in Beach and Amir's (2003) sample, in which participants rarely reported high levels of the IDS items (as evidenced by an average skew of 2.43). This was a cause for concern because prior research has shown that the pronounced indicator skew typical of BDI data in college samples can exert a profound influence on the shapes of taxometric curves, making dimensional results appear similar to those produced by a small taxon (A. M. Ruscio & Ruscio, 2002). Beach and Amir attempted to deal with indicator skew by dropping one particularly skewed item during reanalysis and by removing individuals with a score of 0 from another reanalysis. Unfortunately, skew remained problematically high (2.11) even after the most-skewed item was removed from the indicator set. Moreover, excluding low-scoring cases did not significantly alter taxometric curve shapes because it failed to eliminate the influence of indicator skew in the interpretively important region of the curves.

Beach and Amir (2003) faced challenges common to investigations of psychopathological constructs in nonclinical samples. However, taxometric analyses can be performed in ways that reduce the interpretational ambiguity caused by positively skewed, discontinuous indicators and a small putative taxon. In particular, the most useful analyses are often those that maximize the number of data points on the graphs. For example, Beach and Amir generated maximum covariance (MAXCOV; Meehl & Yonce, 1996) curves with only seven data points (e.g., their Figures 2, 3, and 4), which may have been insufficient to distinguish the genuine, right-end peak produced by a small latent taxon from the rising, cusped curves produced by positively skewed indicators of a latent dimension. For this reason, it is often more informative to generate maximum eigenvalue (MAXEIG; Waller & Meehl, 1998) curves with a far larger number of data points.

When performing MAXEIG analyses, the number of "windows" (subsamples of cases whose members overlap by a fixed proportion from one window to the next) may exceed the number of distinct scores on the input indicator (the variable which forms the *x* axis of a MAXEIG graph and is used to partition cases into an ordered series of windows). For example, using a 7-point indicator consisting of two summed BDI items, one can construct just seven subsamples corresponding to the scores of 0 to 6, or one can construct many more windows. Of course, this guarantees that at least some of the divisions between subsamples will fall between equal-scoring cases, meaning that the assignment of these cases to adjacent windows will be arbitrary. Fortunately, the noise that this may add to a MAXEIG curve can be reduced through the use of internal replications in which the analysis is repeated many times, randomly resorting cases with equal scores each time and using the average results to plot the curve (J. Ruscio & Ruscio, in press). Even though individuals in different windows will have the same scores on the input indicator, the mean scores for the different windows will remain ordered. In this way, the use of a large number of windows can produce a smooth, well-defined MAXEIG curve that should be easier to interpret than a curve resulting from a smaller number of data points, such as those derived from nonoverlapping intervals in a MAXCOV analysis.

Another advantage of MAXEIG is that it can be used to implement the *inchworm consistency test* (Waller & Meehl, 1998), a technique specifically designed to test for a small taxon by systematically increasing the number of windows across a series of MAXEIG analyses. As the number of windows increases, skewed indicators of a latent dimension continue to produce an ambiguously rising curve that ends in a cusp, whereas a genuine taxonic peak should become more sharply defined. This is because the association among indicators reaches a maximum within the window that contains an equal mixture of taxon and complement members (Waller & Meehl, 1998). In the lowest scoring windows, complement members greatly outnumber taxon members, yielding weak indicator associations. Across windows representing higher scoring cases, the proportion of taxon members increases yielding higher indicator associations and a rising curve. With a sufficiently large number of windows—and therefore a sufficiently small subset of cases within each window—a point may be reached when the number of taxon members equals the number of complement members within a window, producing a peak in the MAXEIG curve. If enough numerous (and small) windows are used, taxon members may even outnumber complement members in the uppermost windows, causing the association between indicators (and hence the curve) to decline again. The clarity of this emergent taxonic peak depends on how large a number of windows can be used before the results degrade because of the increased sampling error within each window. To obtain clear results out to the largest possible number of windows, the internal replication technique described above may be especially useful when performing the inchworm consistency test.

Finally, a cornerstone of the taxometric method—and a key defense against misinterpretation—is the use of consistency testing to build confidence in a structural solution. Beach and Amir (2003) evaluated the consistency of their results by conducting two nonredundant taxometric procedures (mean above minus below a cut [MAMBAC] and MAXCOV/MAXEIG[1]) and by dividing their sample into four subsamples to allow replications. Unfortunately, all of the data were drawn from one population (college students), one measure of depression (the BDI), and a constrained set of indicators (six BDI items). This afforded few truly risky consistency checks, making it difficult to identify and correct any erro-

[1] For reasons that are elaborated elsewhere (e.g., J. Ruscio & Ruscio, in press), we view MAXCOV and MAXEIG as highly redundant taxometric procedures that are better conceptualized as variants of a common analytic technique than as consistency tests for one another.

neous interpretations. In addition, although Beach and Amir did conduct one MAXEIG analysis, they used a relatively small number of windows and did not perform the inchworm consistency test, making it difficult to determine whether their cusped curves were truly taxonic or an artifact of indicator skew.

Simulating Comparison Data

Even rigorous consistency testing may not provide sufficient protection against incorrect structural inferences. This is especially true when taxometric curves are interpreted relative only to the curves obtained in Monte Carlo studies, as the "messy" features of research data (e.g., high indicator skew, nontrivial within-group indicator correlations) rarely conform to the idealized parameters used in simulation studies. As noted earlier, cusped MAMBAC and MAXCOV/MAXEIG curves (and correspondingly extreme estimates of the taxon base rate) can emerge very consistently in the absence of taxa if the indicators are substantially skewed. Conversely, data that are simply unsuitable for taxometric analysis can yield results that appear consistently dimensional. For example, when indicators are not sufficiently valid to distinguish the taxon from the complement, MAMBAC and MAXCOV/MAXEIG curves will be nonpeaked even if groups truly exist at the latent level. Because some data parameters may exert a systematic influence on multiple taxometric procedures and consistency tests, the consistency of results may be a necessary, but not sufficient, condition for confidence in the accuracy of a structural inference. Thus, without a clear demonstration that a taxometric analysis can distinguish taxonic from dimensional latent structures given the specific parameters of the research data, one is left to wonder whether a meaningful structural test has actually been performed.

Fortunately, recent advances in data simulation allow researchers to empirically determine whether a given set of data is suitable for taxometric analysis. Rather than using general rules of thumb to judge the adequacy of data and to develop an analysis plan, researchers can simulate taxonic and dimensional comparison data sets whose distributional and correlational parameters are held constant to examine the power with which planned analyses can distinguish these latent structures (J. Ruscio, Ruscio, & Meron, 2003).[2] Bartholomew (1987) has shown that any variance–covariance matrix can be reproduced equally well using a structural model with m latent factors or $m + 1$ latent classes. In other words, any pattern of observed correlations may be reproduced using a dimensional or a taxonic structural model. This does not mean that the structure underlying a specific set of data is arbitrary, nor that the two structural models are equally valid for any particular construct. Rather, it means that efforts to determine the true latent structure must involve more than an examination of full-sample indicator correlations. In particular, a taxometric investigation can be enhanced by reproducing the characteristics of the research data through both taxonic and dimensional models and then examining whether the two models yield different results in taxometric analyses. If taxonic and dimensional data simulated to match the research data yield clearly distinguishable results for a given taxometric procedure, the implication is that the research data, when submitted to the same procedure, should be capable of distinguishing taxonic from dimensional structure.

To generate simulated data for this "suitability test," we developed a technique that uses a structural model with one latent factor to simulate dimensional comparison data and a structural model with two latent classes to simulate taxonic data. Our technique is consistent with existing simulation strategies (e.g., Waller, Underhill, & Kaiser, 1999) but is unique in its iterative approach, which allows users to duplicate the score distribution of each indicator while reproducing the observed matrix of indicator correlations in an unbiased manner (J. Ruscio et al., 2003). Using this technique, one can generate any number of simulated dimensional comparison data sets that reproduce the observed indicator correlation matrix through shared loadings onto a single latent factor, as well as any number of simulated taxonic comparison data sets that reproduce the observed indicator correlation matrix through a combination of group mixture (*separation* along each indicator) and within-group indicator correlations (*nuisance covariance*). All of these data sets are indistinguishable from one another in their full-sample indicator distributions and (within the constraints of sampling error) indicator correlations, with only the latent structure that gives rise to these correlations varying systematically across data sets.

The procedure for generating dimensional comparison data (a) begins by generating vectors of random normal deviates of the same sample size as the research data to represent each indicator, (b) reproduces the indicator correlation matrix through loadings on a single latent factor, (c) transforms each indicator's score distribution to precisely match that of its corresponding indicator in the research data, (d) assesses the extent to which the indicator correlations have been altered (typically reduced) by the distributional transformations, (e) updates the target correlation matrix accordingly, and (f) repeats a–e until the correlations among the simulated indicators reproduce those in the research data as well as possible. This iterative technique yields a simulated dimensional data set that matches the score distribution of each indicator and reproduces the indicator correlations as well as sampling error allows.

Applying this technique separately to members of the putative taxon and complement groups and then merging the results, one can generate a simulated taxonic data set that reproduces the indicator distributions and correlations both within and between groups. The simulation of taxonic comparison data requires a criterion with which to distinguish putative taxon and complement members. For example, one can use diagnostic status as a fallible criterion or calculate a total score across all available indicators and treat a specified proportion of the highest scoring individuals as taxon members (e.g., to simulate a taxon with a base rate of .50, assign the higher scoring half of the sample to the taxon and the lower scoring half of the sample to the complement for purposes of data simulation). To the extent that the putative taxon base rate is uncertain, it may be useful to simulate multiple taxonic data sets with base rates spanning a plausible range of values to determine whether taxa of various sizes could be detected.

Each of the simulated comparison data sets is subjected to the series of taxometric analyses intended for the research data to determine which procedures, if any, are capable of distinguishing taxonic from dimensional structure, given the distributional and correlational characteristics of the research data (J. Ruscio et al., 2003). If parallel analyses of taxonic and dimensional comparison

[2] Programs for simulating taxonic and dimensional comparison data can be downloaded from John Ruscio's Web site at www.etown.edu/psychology/faculty/ruscio.htm

data yield discernibly different results, one can be more confident that the planned analytic approach will afford a useful test between competing structures using the research data. In this way, researchers can simultaneously evaluate the suitability of their unique set of research data as well as their chosen method of implementing specific taxometric procedures. Moreover, so long as the research data are withheld from analysis until the analytic plan passes the suitability test using simulated data, researchers can explore a wide range of analytic options (e.g., alternative techniques for indicator construction and procedural implementation) without running the risk of systematically biasing the taxometric results.

We propose that until the data and analysis plan are shown to provide a genuine test between taxonic and dimensional structure, no structural conclusion should be drawn. In other words, the failure of any analysis to pass this suitability test suggests that it will not yield informative results and should therefore not be applied to the research data. Instead, it is necessary to obtain more valid indicators or a more appropriate sample to implement the taxometric procedure in more effective ways or to rely on other procedures that are demonstrated to be suitable for the available data. Thus, parallel analyses of simulated data are not a substitute for consistency testing, but they do suggest how strongly investigators should weigh the results obtained from each test of the research data.

In addition to helping researchers evaluate the suitability of their data for taxometric analysis, the analysis of simulated taxonic and dimensional comparison data can also facilitate the interpretation of taxometric results (J. Ruscio et al., 2003). Because the simulated data sets match the unique characteristics of the research data and are analyzed in precisely the same way, their results can provide a more useful comparative benchmark than the guidelines provided by relatively idealized Monte Carlo studies, whose indicators are usually normally distributed along truly continuous scales (rather than skewed along discrete values) and are correlated little, if at all, within groups. In addition, extant Monte Carlo studies do not adequately explore the joint influence of potentially problematic factors (e.g., skewed indicators, a small putative taxon, borderline indicator validity, and nontrivial nuisance covariance). Hence, the greater the deviation of a unique set of research data from the idealizations of Monte Carlo samples, the more useful parallel analyses of comparison data may be as an interpretive aid.

Replicating the IDS Analyses in College and Clinical Samples

To illustrate the utility of simulated comparison data for evaluating the suitability of data and facilitating the interpretation of results, we conducted taxometric analyses in two samples using the IDS indicators chosen by Beach and Amir (2003). Our first sample was drawn from a population highly similar to Beach and Amir's, namely undergraduate students at a large university (see A. M. Ruscio & Ruscio, 2002, for details). Our second sample consisted of male veterans who received a psychological evaluation at the National Center for Posttraumatic Stress Disorder in the Veterans Affairs Boston Healthcare System. Although this sample was exclusively male and may have differed from community samples or other clinical samples on such factors as potentially higher rates of comorbid alcohol- and substance-related disorders, it had the advantage of including a substantial number of individuals with clinically significant depression, thereby providing an opportunity to conduct analyses with relatively nonskewed indicators (see J. Ruscio & Ruscio, 2000, Study 1, for details). A sample of 2,293 college students and 882 veterans had complete data on the six BDI items (Items 12, 16, 18, 19, 20, and 21) selected by Beach and Amir as IDS indicators. For each analysis of both samples, we simulated 10 taxonic and 10 dimensional comparison data sets in the manner described above. In the college sample, taxonic data sets were generated using a base rate of .14 (the putative base rate of IDS suggested by Beach & Amir, 2003); in the veteran sample, three different base rates were used: .63 (the rate of diagnosed major depressive disorder), .50, and .37. The latter two values were suggested by MAMBAC and MAXEIG analyses of the veteran data (see *Analyses in the Veteran Sample*), and we used multiple base rates to simulate taxonic data sets to determine whether a putative taxon could have been detected across this broad range of plausible values.

Analyses in the College Student Sample

Following Beach and Amir (2003), we used the MAXEIG and MAMBAC procedures to analyze each data set. To conserve space and afford the most direct comparison with Beach and Amir's results, we present only the averaged curves resulting from analyses of paired-item indicators[3]; all base-rate estimates were derived from these averaged curves. First, MAXEIG analyses were conducted by removing one variable to serve as the input indicator and using the remaining two variables as output indicators for each curve. This yielded three MAXEIG analyses per data set, each repeated with 20, 40, and 60 overlapping windows to implement the inchworm consistency test (see Figure 1, top). The research data yielded averaged MAXEIG curves that slanted upward regardless of the number of windows, producing decreasing base-rate estimates (.15, .12, and .11) as the number of windows increased. The first curve, with 20 windows, was virtually identical to that of Beach and Amir, who used 20 windows (see their Figure 5).[4] However, these curves were more similar to the results of the dimensional comparison data than the taxonic comparison data. As was the case for the research data, the dimensional curves continued to rise unabated and base-rate estimates declined across curves with increasing numbers of windows (.17, .15, and .12). In contrast, taxonic data caused a well-defined taxonic peak to emerge as

[3] According to N. Amir (personal communication, September 26, 2002), BDI Items 12 and 16, 18 and 19, and 20 and 21 were summed in pairs by Beach and Amir (2003). Full graph panels for all of our analyses, including those using single-item indicators, are available on request.

[4] The MAXEIG curve in Beach and Amir (2003) appears to be steeper toward the right than our MAXEIG curve because the authors plotted the data points in an unconventional manner. The *x* value for a MAXEIG point should represent the average score of all cases contained in that window (Waller & Meehl, 1998). Instead, Beach and Amir used the ordinal number of each window (1 through 20) as the *x* values. This artificially forces the points to be spaced equally along the *x* axis and accentuates the increase in eigenvalues among the final few windows. If this technique were applied to our MAXEIG curve, it would appear much steeper as well: Points toward the left would be spread further apart, flattening that region, and points toward the right would be drawn closer together, yielding a steep rise in values. Nonetheless, taxonic and dimensional curves would be affected similarly, and the difference between well-defined taxonic peaks and the mere upward slope of dimensional curves would remain despite an artificial increase in their apparent steepness.

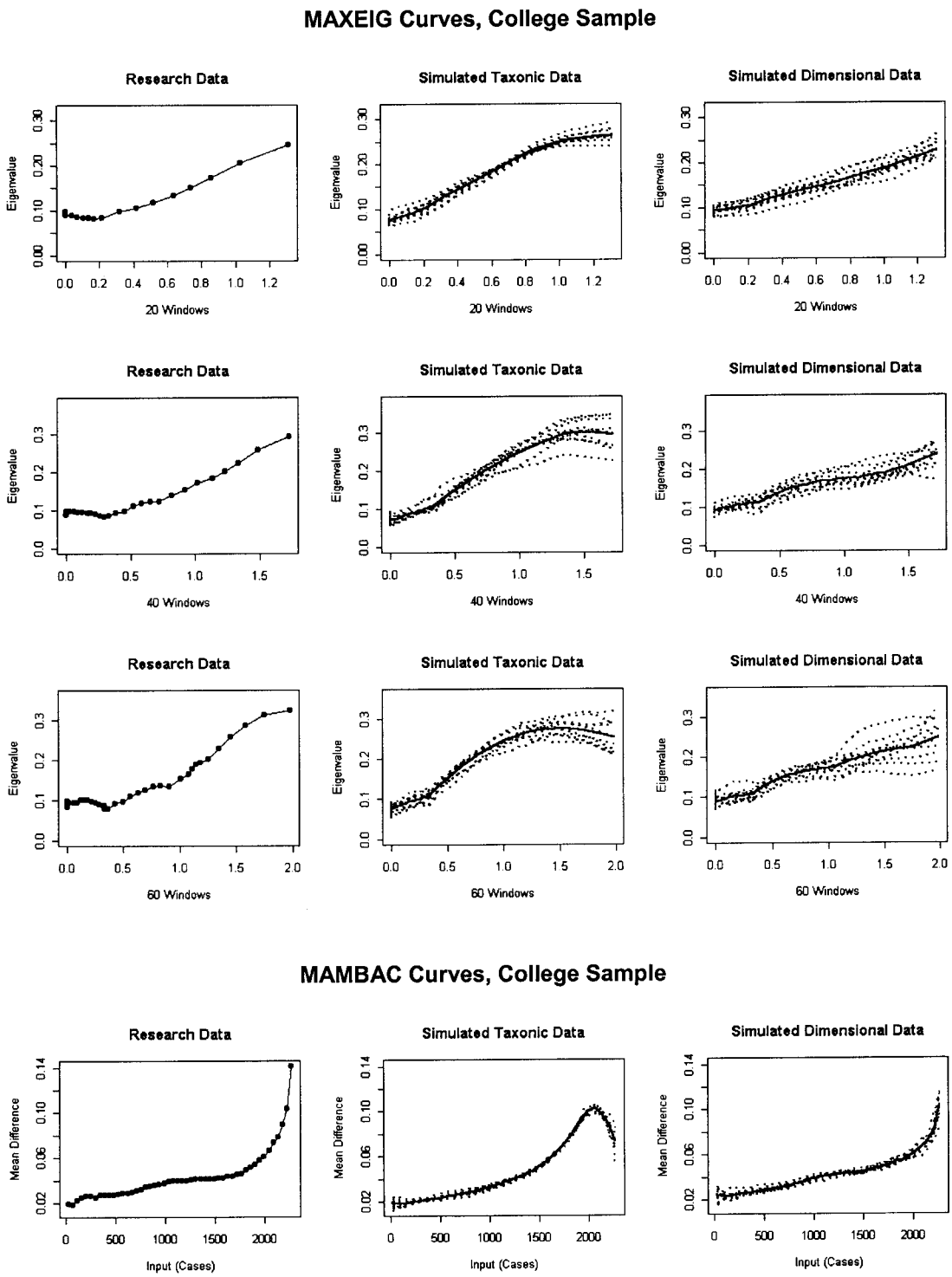

Figure 1. Taxometric analyses in the college student sample. In the first three rows, the averaged MAXEIG (maximum eigenvalue) curves are shown for analyses with 20, 40, and 60 overlapping windows (90% overlap). Values along the *x* axis denote cases' standardized scores on the input indicator. In the fourth row, the averaged MAMBAC (mean above minus below a cut) curves are shown. Values along the *x* axis denote cases, which have been sorted according to their scores on the input indicator. For each MAMBAC analysis, 50 evenly spaced cuts were placed between cases, beginning and ending 25 cases from either extreme. Both MAXEIG and MAMBAC analyses were internally replicated 10 times by randomly shuffling equal-scoring cases on the input indicator, recalculating eigenvalues or mean differences each time, and using the average values for each curve to help stabilize its shape. Graphs for the simulated comparison data show the curves for each of 10 data sets generated using each latent structure (dotted lines) plus the average of these 10 curves (solid line).

the number of windows increased, resulting in stable base-rate estimates across curves (.17, .18, and .19).[5]

We performed additional MAXEIG analyses after excluding all cases with a total score of 0 on the three indicators (constituting approximately one third of the sample), a strategy recommended by Beach and Amir (2003) to reduce skew. Although the skew values of the original indicators (1.26, 2.39, and 2.37) dropped a bit with the removal of low-scoring cases, substantial positive skew remained (0.84, 1.72, and 1.74). More important, there was little change in the shape of the MAXEIG curves after low-scoring cases were excluded, especially toward the right end of the curve that is critical for interpretation when a small taxon is posited. However, the targeted removal of cases did facilitate interpretation by changing taxon base-rate estimates in a predictable way (J. Ruscio, 2000). For example, the research data yielded a taxon base-rate estimate of .11 in the full sample. If the lowest scoring 500 cases in the sample were removed—most of whom would presumably be complement members—one would expect the base rate of the taxon to increase, specifically, to approach .14: (.11 × 2,293)/1,793 = .14. Or, if all cases with total scores of 0 were removed (n = 818), one would expect the estimate to increase further, approaching .17: (.11 × 2,293)/1,475 = .17. In contrast, when we performed these analyses, the base-rate estimate first dropped to .09 with the removal of the 500 lowest scoring cases, then rose only to .13 with the removal of all cases whose total score was 0. Similar results were observed for analyses of simulated dimensional data: The full-sample base-rate estimate (.12) dropped to .10 with the removal of 500 cases and rose only to .12 with the removal of all zero-scoring cases. In contrast, analyses of simulated taxonic data yielded predictable increases, rising from the full-sample base-rate estimate of .19 to .22 with the removal of 500 cases (expected value = .24) and still higher to .27 with the removal of all cases with total scores of 0 (expected value = .28). Thus, although the removal of low-scoring cases does not substantially reduce the influence of indicator skew on curve shape, it does allow the taxon base-rate estimates yielded by MAXEIG to be usefully compared within and across data sets to inform a structural conclusion. In the present data, these comparisons pointed to a latent dimension.

Next, MAMBAC was conducted by removing one paired-item variable to serve as the output indicator and summing the remaining two paired-item variables to serve as the input indicator for each curve. Hence, three curves were generated and averaged for each data set (see Figure 1, bottom). The research data yielded a rising MAMBAC curve and base-rate estimate (.12) highly similar to those obtained by Beach and Amir (2003; see their Figure 6). Whereas they interpreted this as evidence of a latent taxon, these results mirror those of our dimensional comparison data, in which a similar rising curve yielded a base-rate estimate of .15 and were quite different from those of the taxonic comparison data, in which a peak was well-defined and yielded a base-rate estimate of .25.

We also conducted MAMBAC analyses after excluding all cases with a total score of 0 on the three indicators. Once again, there was little change in the shape of the MAMBAC curves after low-scoring cases were excluded. Whereas every point on the curve can be used to estimate the taxon base rate in a MAXEIG or MAXCOV analysis (J. Ruscio, 2003), only the two endpoints of the curve are used to generate this estimate in a MAMBAC analysis (Meehl & Yonce, 1994). This means that any technique that artificially reduces mean differences at the leftmost cutting score (such as removing a substantial proportion of low-scoring cases) would be expected to yield lower MAMBAC estimates of the taxon base rate, regardless of latent structure. Thus, the case-removal consistency test is unlikely to provide informative results when used with MAMBAC. In sum, our rising MAMBAC curves (obtained with and without low-scoring cases) and those obtained by Beach and Amir (2003) appear likely to be the result of positively skewed indicators of a latent dimension, not a small latent taxon.

Analyses in the Veteran Sample

MAXEIG was conducted in the same manner as in the college sample, but because the putative taxon was fairly large and the indicators were not very skewed (skew = −0.16, 0.78, and 0.31), the inchworm consistency test was not needed (see Figure 2, top). The research data appeared suitable for analysis and yielded results more similar to those of the dimensional comparison data than the taxonic comparison data generated using any of the three taxon base rates. The similarity of base-rate estimates for the research data (.37) and the dimensional comparison data (.36) lent further support to a dimensional interpretation.

MAMBAC was performed next (see Figure 2, bottom). Because the indicators were not very skewed, MAMBAC curves did not

[5] Although the MAXEIG results of the college student sample appeared more consistent with a dimensional solution than a taxonic solution, the corresponding curves yielded by the simulated taxonic and dimensional comparison data sets were not as easy to distinguish as the curves for all other analyses, making a dimensional interpretation of these results perhaps more debatable than the others. Thus, rather than simply relying on visual inspection to reach a structural inference in this case, we also sought objective corroboration by quantitatively measuring the fit between the curves of the simulated comparison data and those of the research data. To do this, we used an index calculated as the root mean square residual (RMSR) of the y values on each of the averaged curves (J. Ruscio, Haslam, & Ruscio, 2003; J. Ruscio & Ruscio, 2003). This index is computed once to evaluate the fit of the simulated taxonic data to the research data and once to evaluate the fit of the simulated dimensional data to the research data. The Fit_{RMSR} index is calculated as follows:

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

where $y_{res.data}$ refers to a data point on the averaged curve for the research data, $y_{sim.data}$ refers to the corresponding data point on the averaged curve for simulated taxonic or dimensional data, and N refers to the number of points on each curve. Lower values of Fit_{RMSR} reflect better fit, with perfect fit represented by a value of 0. This index is not interpreted in absolute terms, but in terms of the comparative fit across structural models when parameters of the analysis and the data are held constant. To the extent that Fit_{RMSR} differs across the simulated taxonic and dimensional data, the evidence favors the structure that yields the superior (lower) fit value.

For the MAXEIG analyses in the college sample, the Fit_{RMSR} value for the dimensional comparison data was roughly half the size of the value for the taxonic data, regardless of the number of windows used (.014 vs. .032, .021 vs. .040, and .025 vs. .040 at 20, 40, and 60 windows, respectively). Thus, objective as well as subjective evaluation of the curves pointed to a dimensional solution. Given the readily interpretable curves yielded by all subsequent analyses, we omitted the remaining fit values to conserve space (all fit values are available on request). However, in each analysis, the fit indices favored dimensional structure over taxonic structure to an equal or greater extent than that observed here.

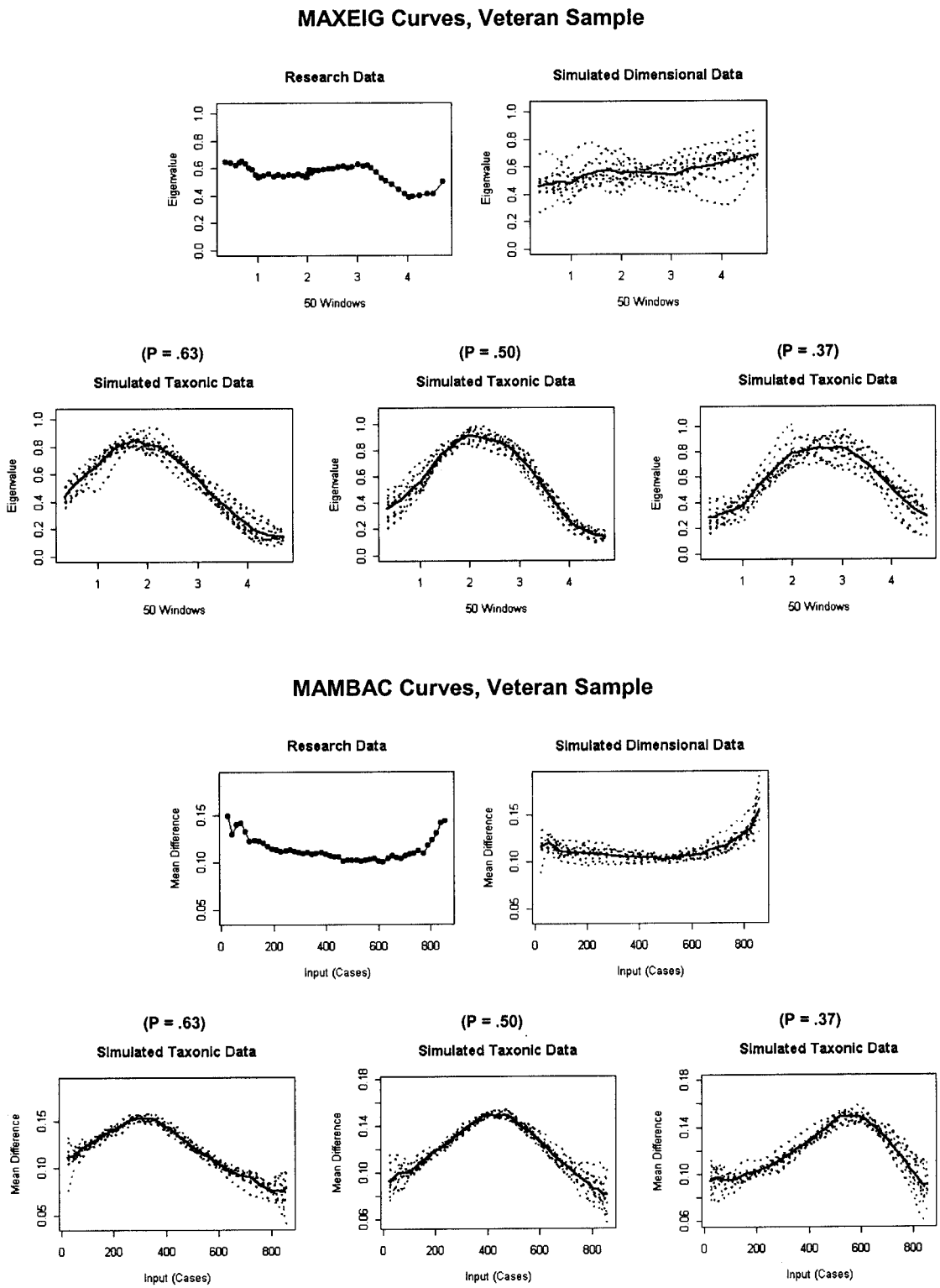

Figure 2. Taxometric analyses in the veteran sample. In the top two rows, the averaged MAXEIG (maximum eigenvalue) curves are shown for analyses with 50 overlapping windows (90% overlap); the first row contains curves for the research and simulated dimensional data sets, and the second row contains curves for taxonic data sets simulated using three different base rates. Values along the *x* axis denote cases' standardized scores on the input indicator. In the bottom two rows, the averaged MAMBAC (mean above minus below a cut) curves are shown; the third row contains curves for the research and simulated dimensional data sets, and the fourth row

(*Figure 2 caption continues on next page*)

slope upward as they did in the college sample. Moreover, as might be expected, taxonic and dimensional structures were more clearly differentiated in a clinical sample with a larger absolute number and base rate of putative taxon members. Taxonic comparison data simulated using each of three taxon base rates yielded peaked MAMBAC curves, whereas the dimensional comparison data yielded U-shaped curves characteristic of dimensional structure. Consistent with previous analyses, the MAMBAC curves of the research data bore a striking resemblance to those of the dimensional comparison data, as did their estimates of the taxon base rate (.50 vs. .48). Thus, taxometric analyses corroborated the dimensional structure of IDS in a clinical population that avoided the obfuscating influence of high indicator skew.

Implications for Taxometric Research in Psychopathology

The present study replicated the analyses of a recent taxometric investigation (Beach & Amir, 2003) that examined the latent structure of a clinical construct—IDS—in an unselected college sample using a small number of items from a single measure of depression. What made Beach and Amir's (2003) study illustrative is that, like other psychopathology studies conducted with analogue samples, it involved unfavorable data conditions such as a small putative taxon and indicators that possessed high positive skew. These conditions have previously been found to produce rising taxometric curves that may cause dimensional data to be mistakenly interpreted as taxonic (A. M. Ruscio & Ruscio, 2002). In fact, the present results closely mirrored those obtained by Beach and Amir, but parallel analysis of simulated comparison data clearly revealed these results to be more indicative of dimensional than taxonic structure. Further replication within a clinical sample characterized by a high rate of depressive pathology similarly refuted Beach and Amir's original taxonic interpretation of IDS.

This example highlights the problems that may ensue when the taxometric curves of less-than-ideal data—such as those examining relatively rare clinical phenomena in nonclinical samples—are interpreted relative to the curves from idealized Monte Carlo studies. At the same time, it also suggests that careful selection of taxometric procedures appropriate to the data, in conjunction with a comparative benchmark matching the properties of the research data, can do much to guard against misinterpretation. To this end, we have illustrated a versatile technique involving the iterative simulation and parallel analysis of taxonic and dimensional comparison data that can help to evaluate the suitability of one's data for the planned analyses and facilitate the interpretation of otherwise ambiguous or misleading results.

Because unsuitable data may yield not only ambiguous but misleading taxometric results, it is essential to demonstrate at the outset of a taxometric investigation that the research data are suitable for analysis. This means that no conclusions should be drawn unless simulated comparison data have been shown to differentiate taxonic from dimensional latent structure. Such a suitability test may be failed for a number of reasons, including too few putative taxon members, too much indicator skew, insufficiently valid indicators, excessive nuisance covariance, or an unsound analysis plan that does not take full advantage of multivariate taxometric procedures and the most appropriate consistency tests. Continued refinement of the indicators and analysis plan may lead to passing a subsequent suitability test, but there is no guarantee that the available data will ultimately be appropriate for taxometric analysis. In fact, it is possible that some conclusions reached in published taxometric studies were based on unsuitable data.

A second benefit of parallel analyses of simulated comparison data is that their results can serve as a valuable interpretive aid. This benefit was recently demonstrated when Rothschild, Cleland, Haslam, and Zimmerman (2003) used comparison data simulated with our iterative technique in their study of borderline personality disorder. Whereas an initial series of analyses yielded rising taxometric curves that were interpreted as evidence of a small taxon, subsequent analyses performed on comparison data showed that the results were more consistent with dimensional structure. These results add to a growing body of evidence suggesting that (a) factors such as pronounced indicator skew and small putative taxa can affect the shapes of taxometric curves in ways that may be mistaken as taxonic, but (b) informed taxometric analysis using simulated comparison data can allow investigators to successfully distinguish taxonic from dimensional constructs even under some of these adverse conditions. This raises the possibility that some previously published taxometric investigations may benefit from reanalysis using simulated comparison data to determine whether their conclusions of a small latent taxon were warranted. For example, a number of studies (e.g., Strong, Green, & Schinka, 2000; Waller & Ross, 1997; Woodward, Lenzenweger, Kagan, Snidman, & Arcus, 2000) have drawn inferences of relatively small latent taxa on the basis of results that included rising MAMBAC or MAXCOV/MAXEIG curves. In each case, insufficient information was provided to determine whether positively skewed indicators of a latent dimension may have been responsible for the observed results.

In addition to the use of simulated comparison data, we recommend that researchers positing small taxa use overlapping windows (as in MAXEIG) rather than nonoverlapping intervals (as in MAXCOV) to increase the number of data points on the curve so that a small taxon is more likely to be detected. We also believe that the inchworm consistency test should be a required analytic tool in any taxometric investigation that involves a small putative taxon. Among all taxometric procedures and consistency tests, this

contains curves for taxonic data sets simulated using three different base rates. Values along the *x* axis denote cases that have been sorted according to their scores on the input indicator. For each MAMBAC analysis, 50 evenly spaced cuts were placed between cases, beginning and ending 25 cases from either extreme. Both MAXEIG and MAMBAC analyses were internally replicated 10 times by randomly shuffling equal-scoring cases on the input indicator, recalculating eigenvalues or mean differences each time, and using the average values for each curve to help stabilize its shape. Graphs for the simulated comparison data show the curves for each of 10 data sets generated using each latent structure (dotted lines) plus the average of these 10 curves (solid line).

technique stands out as best suited for distinguishing a small latent taxon from a latent dimension with positively skewed indicators.

In contrast, there are three practices that we believe should be used with caution in taxometric studies, especially those involving a small putative taxon. First, although the coherence of taxon base-rate estimates across taxometric analyses can serve as a useful consistency test, it is important to remember that such estimates may appear coherent for reasons other than taxonic latent structure. The rising taxometric curves produced by positively skewed indicators tend to yield low and highly consistent estimates of the taxon base rate, whether the latent structure is taxonic or dimensional. Thus, whereas a high standard deviation for a set of base rate estimates may provide relatively strong evidence of a latent dimension, a low standard deviation may be considerably more ambiguous. Additional ambiguity is introduced by the fact that Monte Carlo research has not yet established a threshold below which the standard deviation of taxon base-rate estimates can confidently be used to infer taxonic latent structure, nor has it evaluated the general applicability of any particular threshold across a range of data parameters, taxometric procedures, and analytic implementation decisions. Thus, structural interpretations based on the consistency of taxon base-rate estimates remain highly subjective. One way to facilitate these interpretations might be to contextualize them within the comparative framework provided by parallel analyses of simulated data. To the extent that simulated taxonic data yield a low standard deviation for base-rate estimates while simulated dimensional data yield a discernibly higher standard deviation for the same taxometric procedure, the base-rate consistency test may be more confidently used to draw structural inferences about the research data, and the standard deviation of estimates yielded by the research data may be compared with those of the simulated data to facilitate accurate interpretation. Thus, just as the curves yielded by parallel analyses of simulated comparison data can serve as a useful interpretive benchmark, so can estimates of latent parameters derived from these analyses.

Second, removing cases from a sample does not appear to substantially reduce indicator skew or its distorting effects on taxometric graphs. This is because low-scoring cases contribute to data points in a region of the graph that is not critical to interpretation, leaving the critical region largely unaffected.[6] However, at least for MAXCOV/MAXEIG analyses, targeted case removal can serve as a consistency test by focusing on whether the change in taxon base-rate estimates following case removal is consistent with what would be expected for taxonic structure (J. Ruscio, 2000).

Third, we recommend against dividing a large sample into subsamples to replicate analyses unless the putative taxon base rate is substantial. Although the goal of replication is admirable, spreading a small number of taxon members across subsamples can drastically reduce the odds of their successful detection in any subsample, let alone their consistent detection across all subsamples. The consistency tests within the taxometric method—especially when performed in conjunction with parallel analyses of simulated comparison data—afford ample opportunity to check the coherence of results without creating subsamples (J. Ruscio & Ruscio, in press). On the other hand, when supplementary data from other populations are available (especially populations in which the number of putative taxon members is likely to be higher, as in our veteran sample), conceptual replication may be valuable.

Given the relative ease with which large samples of data can be collected in college settings, it is likely that clinical scientists will continue to conduct taxometric investigations with college student data. The present results suggest that such investigations have the potential to be informative when the data first pass a suitability test and are then submitted to carefully designed analyses alongside simulated comparison data. As these findings are replicated in appropriate clinical samples, confidence in their structural solutions, and in their relevance to clinical populations of interest, will be further enhanced.

[6] Elsewhere J. Ruscio and Ruscio (2004) have demonstrated that the addition of complement members to a sample does not easily obscure a taxonic peak. Thus, the increase in the putative taxon's base rate achieved by dropping low-scoring cases provides an illusory gain, as the absolute number of taxon members in the sample appears to be at least as important for the successful detection of taxonic structure. For this reason, we consistently refer to a "small taxon" rather than a "low base-rate taxon."

References

Bartholomew, D. J. (1987). *Latent variable models and factor analysis.* New York: Oxford University Press.

Beach, S. R. H., & Amir, N. (2003). Is depression taxonic, dimensional, or both? *Journal of Abnormal Psychology, 112,* 228–236.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression.* New York: Guilford Press.

Gilbert, P. (1992). *Depression: The evolution of powerlessness.* New York: Guilford Press.

Gilbert, P. (2000). Varieties of submissive behavior as forms of social defense: Their evolution and role in depression. In L. Sloman & P. Gilbert (Eds.), *Subordination and defeat* (pp. 3–45). Mahwah, NJ: Erlbaum.

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

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

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.

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

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. (2003). *Extending the maximum eigenvalue (MAXEIG) taxometric procedure.* 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. (2004). Clarifying boundary issues in psycho-

pathology: The role of taxometrics in a comprehensive program of structural research. *Journal of Abnormal Psychology, 113,* 24–38.

Ruscio, J., & Ruscio, A. M. (in press). A conceptual and methodological checklist for conducting taxometric investigation. *Behavior Therapy.*

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

Strong, D. R., Green, R. L., & Schinka, J. A. (2000). A taxometric analysis of MMPI-2 infrequency scales [F and F(p)] in clinical settings. *Psychological Assessment, 12,* 166–173.

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

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.

Waller, N. G., Underhill, J. M., & Kaiser, H. A. (1999). A method for generating simulating plasmodes and artificial test clusters with user-defined shape, size, and orientation. *Multivariate Behavioral Research, 34,* 123–142.

Woodward, S. A., Lenzenweger, M. F., Kagan, J., Snidman, N., & Arcus, D. (2000). Taxonic structure of infant reactivity: Evidence from a taxometric perspective. *Psychological Science, 11,* 296–301.

Received February 3, 2003
Revision received July 14, 2003
Accepted July 14, 2003 ■