Psychological Assessment
2009, Vol. 21, No. 1, 99–111

© 2009 American Psychological Association
1040-3590/09/$12.00 DOI: 10.1037/a0015010

To Sum or Not to Sum: Taxometric Analysis With Ordered Categorical Assessment Items

Glenn D. Walters
Federal Correctional Institution–Schuylkill

John Ruscio
The College of New Jersey

Meehl's taxometric method has been shown to differentiate between categorical and dimensional data, but there are many ways to implement taxometric procedures. When analyzing the ordered categorical data typically provided by assessment instruments, summing items to form input indicators has been a popular practice for more than 20 years. A Monte Carlo study compared the accuracy of taxometric analyses implemented in the traditional way (without summing items) and taxometric analyses implemented with the summed-input method. These analyses generated no support for the summed-input method, which substantially reduced discriminating power for 2 of the 3 procedures studied. Accuracy was highest when 5 or more indicators and 4 or more ordered categories were used. Findings from the simulation study were then used to help interpret the results for taxometric analyses of antisocial personality disorder criteria with real research data. In this example, the traditional method yielded clearer results than the summed-input method. Implications for the use and further study of the taxometric method in assessment research are discussed.

Keywords: taxometrics, MAMBAC, MAXCOV, MAXEIG, implementation

The goal of taxometric analysis is to assist researchers in determining whether the latent structure of a theoretical construct is categorical (taxonic) or continuous (dimensional). To accomplish this, researchers can make use of procedures derived from Meehl's (1992, 1995, 2004) coherent cut kinetics framework, such as mean above minus below a cut (MAMBAC; Meehl & Yonce, 1994), maximum covariance (MAXCOV; Meehl & Yonce, 1996), and maximum eigenvalue (MAXEIG; Waller & Meehl, 1998). To say that there has been an explosion of interest in Meehl's taxometric method in recent years is no exaggeration. Although Meehl began working on the taxometric problem of distinguishing between categorical and continuous latent structures in the early 1960s, more than half the 160-plus papers published on taxometrics in peer-reviewed journals since 1978 have been published in the last 5 years, as have more than half the 200-plus papers that have cited the *American Psychologist* article that brought the taxometric method to the attention of a large audience (Meehl, 1995).[1] In recent years, many taxometric studies have examined constructs and measures of special interest to researchers and practitioners specializing in psychological assessment, including anxiety sensitivity (Bernstein et al., 2007), disgust sensitivity (Olatunji & Broman-Fulks, 2007), the Minnesota Multiphasic Personality Inventory–2 (MMPI-2) *F(p)* scale (Strong, Glassmire, Frederick, & Greene, 2006), and the Psychopathy Checklist: Screening Version (Walters, Gray, et al., 2007).

Meehl's original intent was to use taxometrics to test an important taxonic aspect of his theory of schizophrenia, and he was primarily interested in applying the taxometric method to well-articulated theoretical constructs as a means of determining whether they possessed a categorical latent structure (Meehl & Golden, 1982). However, Meehl also acknowledged that dimensions can be just as important as taxa and that the taxometric method need not be guided by a theoretical model to be useful (Meehl, 1995, 2004). The recognition that taxometrics can be used to distinguish between two structural models, each of which holds important implications, helped stimulate research on the taxometric method and supports the competing structural hypotheses inferential framework in which taxonic and dimensional structure are construed as competing alternative hypotheses (Ruscio, Haslam, & Ruscio, 2006). Making this distinction holds important implications for the development and use of assessment instruments (Ruscio & Ruscio, 2002). For example, the goal when assessing a taxonic construct is to assign individuals to groups, whereas the goal when assessing a dimensional construct is to locate individuals' positions along one or more latent traits. Classifying individuals into groups can be done most effectively when items are developed and selected to maximize their discriminating power near the location of the boundary separating the groups. In contrast, locating individuals' scores along continua requires the use of some items that discriminate at each point in the full range of trait levels. Because one cannot simultaneously use items whose discriminating powers are clustered together and widely dispersed, empirically determining the latent structure of the target construct can help to guide assessment.

Glenn D. Walters, Psychology Services, Federal Correctional Institution–Schuylkill, Minersville, Pennsylvania; John Ruscio, Department of Psychology, The College of New Jersey.

The assertions and opinions contained herein are the private views of the authors and should not be construed as official or as reflecting the views of the Federal Bureau of Prisons or the U.S. Department of Justice.

Correspondence concerning this article should be addressed to Glenn D. Walters, Psychology Services, Federal Correctional Institute–Schuylkill, PO Box 700, Minersville, PA 17954-0700. E-mail: gwalters@bop.gov

[1] On February 20, 2008, a search of the PsycINFO database for the terms *taxometric* and *taxometrics* in the title, abstract, or keywords identified 163 peer-reviewed journal articles, 92 (56%) of which were published between 2003 and 2007. On the same date, a search of the ISI Web of Science database identified 203 articles that had cited Meehl (1995), 110 (54%) of which were published between 2003 and 2007.

Each taxometric procedure has been introduced with its own conventional method for assigning indicators (observed variables) to input and output roles, and the use of composite input indicators is an alternative technique. The MAMBAC (Meehl & Yonce, 1994) procedure involves calculation of differences on one variable (the output indicator) for cases scoring above and below a series of cuts and thresholds along another variable (the input indicator). The prototypical MAMBAC results for taxonic data show a peak in the mean differences near an optimal threshold along the input indicator, whereas the prototypical results for dimensional data are concave along the input indicator (Meehl & Yonce, 1994). To perform MAMBAC, one can arrange k items into all possible input–output pairs, yielding $k(k - 1)$ analyses. This will be referred to as the traditional MAMBAC method, or MAMBAC-T, because it is the method originally adopted by the procedure's creator, Paul Meehl. Alternatively, one can assign a single item as the output indicator and combine the remaining items into a composite input indicator by summing them, yielding k analyses; this will be referred to as the summed-input method, or MAMBAC-S.

The MAXCOV (Meehl & Yonce, 1996) procedure involves the calculation of covariances between two variables (the output indicators) within a series of ordered subsamples along another variable (the input indicator). The prototypical MAXCOV results for taxonic data show a peak in the covariances near a subsample containing a fairly even mixture of members from two groups, whereas the prototypical results for dimensional data are flat (Meehl & Yonce, 1996). To perform MAXCOV, one can arrange k items into all possible input–output–output triplets, yielding $k(k - 1)(k - 2)/2$ analyses (MAXCOV-T), or set aside two items to serve as outputs and sum the remaining $k - 2$ to form a summed-input indicator (MAXCOV-S), which yields $k(k - 1)/2$ analyses. The MAXEIG (Waller & Meehl, 1998) procedure is very similar to MAXCOV in that ordered subsamples are formed along an input indicator and the association between output indicators is calculated within each subsample. Rather than calculate covariances for a pair of output indicators, as is done for MAXCOV, the MAXEIG procedure calculates eigenvalues for two or more output indicators. Specifically, the first (largest) eigenvalue is calculated with a covariance matrix, which is the variance–covariance matrix modified such that variances are replaced with zeros along the diagonal to leave only off-diagonal covariances. With MAXEIG, one can arrange k items into a single input indicator and $k - 1$ output indicators, yielding k analyses (MAXEIG-T), or set aside two items to serve as outputs and sum the remaining $k - 2$ to form a summed-input indicator (MAXEIG-S), which yields $k(k - 1)/2$ analyses.

For the implications of choosing traditional versus summed-input methods, consider how many analyses are performed for $k = 5$ indicators: 20 for MAMBAC-T versus 5 for MAMBAC-S, 30 for MAXCOV-T versus 10 for MAXCOV-S, and 5 for MAXEIG-T versus 10 for MAXEIG-S. In sum, this is 55 analyses for traditional input indicator methods versus 25 analyses for summed-input indicators. Of course, the total number of analyses is less important than their ability to determine whether a taxonic or dimensional structural model better fits the data. That is the question addressed in the present research.

Published taxometric studies have obtained inconsistent findings with the traditional and summed-input methods for MAMBAC, MAXCOV, and MAXEIG. For example, with data drawn from the Psychological Inventory of Criminal Thinking Styles (Walters, 1995) in a sample of male and female prison inmates and college students (Walters & McCoy, 2007), MAMBAC-T and MAMBAC-S results agreed in only two of five instances. Both procedures showed evidence of dimensionality in the offender and male subsamples, but MAMBAC-S suggested dimensionality in the total sample and student subsample, whereas MAMBAC-T yielded a mixture of ambiguous and taxonic results. In the female subsample, MAMBAC-T produced taxonic results, in contrast to the ambiguous results produced by MAMBAC-S. There was weak consistency in the results obtained with MAXCOV and MAXEIG implemented in traditional and summed-input ways, too. These inconsistent findings raise questions about which implementation to trust.

Without a gold standard against which to judge the differential results stemming from these implementations, it is difficult to determine whether one method should be given greater weight or whether some methods are biased in favor of producing taxonic results and others are biased in favor of producing dimensional results. Summing items to construct input indicators has been recommended when the indicators vary across a small number of ordered categories and/or the number of indicators is high (e.g., Gangestad & Snyder, 1985; Ruscio et al., 2006). The indicators used in several recent taxometric investigations, like the ones conducted on the revised Psychopathy Checklist: Screening Version (Hare, 2003) facet scores (Edens, Marcus, Lilienfeld, & Poythress, 2006; Guay, Ruscio, Knight, & Hare, 2007; Walters, Gray, et al., 2007) or the ones conducted on the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychiatric Association, 2000) criteria for major depressive disorder (Ruscio, Zimmerman, McGlinchey, Chelminski, & Young, 2007; Solomon, Ruscio, Seeley, & Lewinsohn, 2006), are composed of a relatively small number of ordered categories rather than being continuously distributed. In each of these studies, investigators have included analyses of summed-input indicators.

Many assessment instruments contain items with ordered categorical response scales. For example, items on each version of the Minnesota Multiphasic Personality Inventory are dichotomous, and items on each version of the Beck Depression Inventory allow four possible responses. When performing taxometric analyses with ordered categorical data, one especially important question of how to implement the most commonly used procedures involves choosing how to assign the available indicator variables to the required roles of input and output indicators. Input indicators are used to rank-order cases and locate thresholds so that statistics can be calculated (using the output indicators) across a series of ordered subsamples. The need to rank-order cases reliably suggests that these procedures may perform poorly when indicators vary across a fairly small number of ordered categories (e.g., dichotomous items). One way to handle this potential problem is to set aside the output indicators required for analysis and combine all remaining indicators (e.g., by summing) to form a composite input indicator (Gangestad & Snyder, 1985). Ruscio et al. (2006) suggested a number of potential benefits and costs associated with this approach. By creating a larger range of values, one can obtain a more reliable rank ordering of cases, and this may be especially helpful when ordered categorical variables are being used. Another potential advantage is that when all available indicators are sufficiently valid, the inclusion of all indicators in each analysis may increase the power with which taxonic and dimensional structures can be differentiated.

Because the important decision of whether to use the traditional or the summed-input method has not been studied, empirically based guidelines are unavailable. Moreover, relatively little is known about the performance of the MAMBAC, MAXCOV, and MAXEIG procedures with varying numbers of items and ordered categories in their response scales. We designed a Monte Carlo study to address these issues. We generated data sets with differing numbers of indicators and differing numbers of ordered categories and performed MAMBAC, MAXCOV, and MAXEIG analyses using traditional and summed-input methods. The hypothesis for the first study was that the summed-input method would be more useful than the traditional method when the number of ordered categories is small, especially when the number of indicators is large. To the extent that one or both of these conditions were not met, we anticipated that there would be little, if any, advantage to the summed-input method. With few ordered categories, it becomes difficult to rank-order cases reliably along an input indicator, so summed-input indicators may be helpful. For summed-input indicators to provide much benefit, there must be a nontrivial number of variables to sum, or else the reliability with which cases can be rank ordered may not improve much. Because each target data set was analyzed with three taxometric procedures, the study afforded an opportunity to examine their relative performance. From the outset, however, we emphasize that our primary goal was to test the relative performance of the two input indicator methods and only secondarily the relative performance of the three taxometric procedures.

In a second study, we used the Monte Carlo results to help interpret the results of taxometric analyses using real data collected on prison inmates to examine the construct of antisocial personality disorder (ASPD). Whereas the true latent structure was known for target data sets in the Monte Carlo study, the latent structure of ASPD was not known with certainty. Because most prior research suggests that a dimensional model better fits the data than a taxonic model (e.g., Edens et al., 2006; Guay et al., 2007; Marcus, Lilienfeld, Edens, & Poythress, 2006; Walters, 2007; Walters, Diamond, Magaletta, Geyer, & Duncan, 2007; Walters, Gray, et al., 2007; Walters & McCoy, 2007), we expected dimensional results but remained open to the possibility that they might support either of these structural models.

Study 1

Method

Design and data generation. Using an iterative technique developed by Ruscio and Kaczetow (2008), we generated a total of 4,800 (2,400 dimensional, 2,400 taxonic) data sets for a Monte Carlo analysis of the traditional and summed-input versions of MAMBAC, MAXCOV, and MAXEIG using ordered categorical data. Latent structure (dimensional, taxonic), number of ordered categories (C = 2, 3, 4, 5, 6, 7, 10, 20), and number of indicators (k = 3, 4, 5, 6, 7, 8) were systematically varied, and all other data parameters (sample size, taxon base rate, indicator validity, within-group correlation, asymmetry, tail weight, taxon:complement variance ratio, indicator correlation) were randomly sampled. The design therefore includes 96 cells (2 latent structures × 8 ordered categories × 6 indicators), within which 50 configurations of data parameters were randomly sampled, for a total of 4,800 (96 cells × 50 samples) target data sets.

In constructing the taxonic data sets, we randomly sampled data parameters from the following uniform distributions: sample size (N = 300–1,000), taxon base rate (p = .05–.50), indicator validity (d = 1.25–2.00), within-group correlations (r = .00–.30), asymmetry (g = .00–.30), tail weight (h = .00–.15), and taxon:complement variance ratio (VR = 1–4). The values of g and h were used to generate data from a g-and-h distribution (Hoaglin, 1985); the magnitude of g controls the asymmetry relative to a normal distribution (in which g = 0), and the magnitude of h controls the tail weight relative to a normal distribution (in which h = 0). Because only positive values of g and h were used, conditions of positive skew and heavy tail weight were studied. For the g-and-h populations used in this study, smallest skew (γ_1) and kurtosis (γ_2) values were $\gamma_1 = 0$, $\gamma_2 = 0$ for $g = 0$ and $h = 0$ (a normal distribution), and the largest values were $\gamma_1 = 2.60$, $\gamma_2 = 38.89$ for $g = .30$ and $h = .15$; other pairings of g and h correspond to γ_1 and γ_2 values within this range. To generate a taxonic data set, we used the iterative technique of Ruscio and Kaczetow (2008) to sample N cases from a g-and-h distribution with $\mu = 0$, $\sigma = 1$, and a correlation matrix in which all indicators correlated r with one another. Next, a proportion (P) of cases was randomly selected and identified as taxon members, with the remainder identified as members of the complement class. The variance ratio was achieved by multiplying scores in the taxon by VR. Then, separation between classes was achieved by adding a constant to taxon members such that the standardized mean difference equaled d.

For dimensional data, parameters were sampled at random in the same way, although several of those used for taxonic data (P, d, r, and VR) do not correspond to parameters of the dimensional (common factor) model. In their place, indicator correlations were sampled from a uniform distribution ranging from r_{xy} = .15 to r_{xy} = .65. The iterative algorithm of Ruscio and Kaczetow (2008) was used to sample N cases from a g-and-h distribution with $\mu = 0$, $\sigma = 1$, and a correlation matrix in which all indicators correlated r_{xy} with one another. Extensive checking showed that our data generation programs created taxonic and dimensional target data sets with the intended indicator correlations, distributions, and variance ratios.

For both latent structures, each randomly sampled configuration of data parameters was used to generate 48 target data sets by crossing the eight levels of C with the six levels of k. As described above, data were originally continuous, and each indicator's distribution was cut into ordered categories with equally spaced thresholds spanning the range of observed values. To eliminate the possibility that outliers would result in categories with very low, or even zero, frequencies at or near the extremes, we created the continuous g-and-h distributions using uniform quantiles spanning the range from .005 to .995. Figure 1 illustrates the results of this technique for a taxonic indicator that was originally continuous and then cut to form indicators with successively smaller numbers of ordered categories.

Taxometric analyses. The MAMBAC, MAXCOV, and MAXEIG procedures were performed, each with the traditional and summed-input indicator methods. Although summed-input MAXEIG is rarely used, the same cannot be said of summed-input MAMBAC and summed-input MAXCOV, particularly when dichotomous and trichotomous indicators are analyzed (Gangestad & Snyder, 1985; Schmidt, Kotov, & Joiner, 2004). MAMBAC was

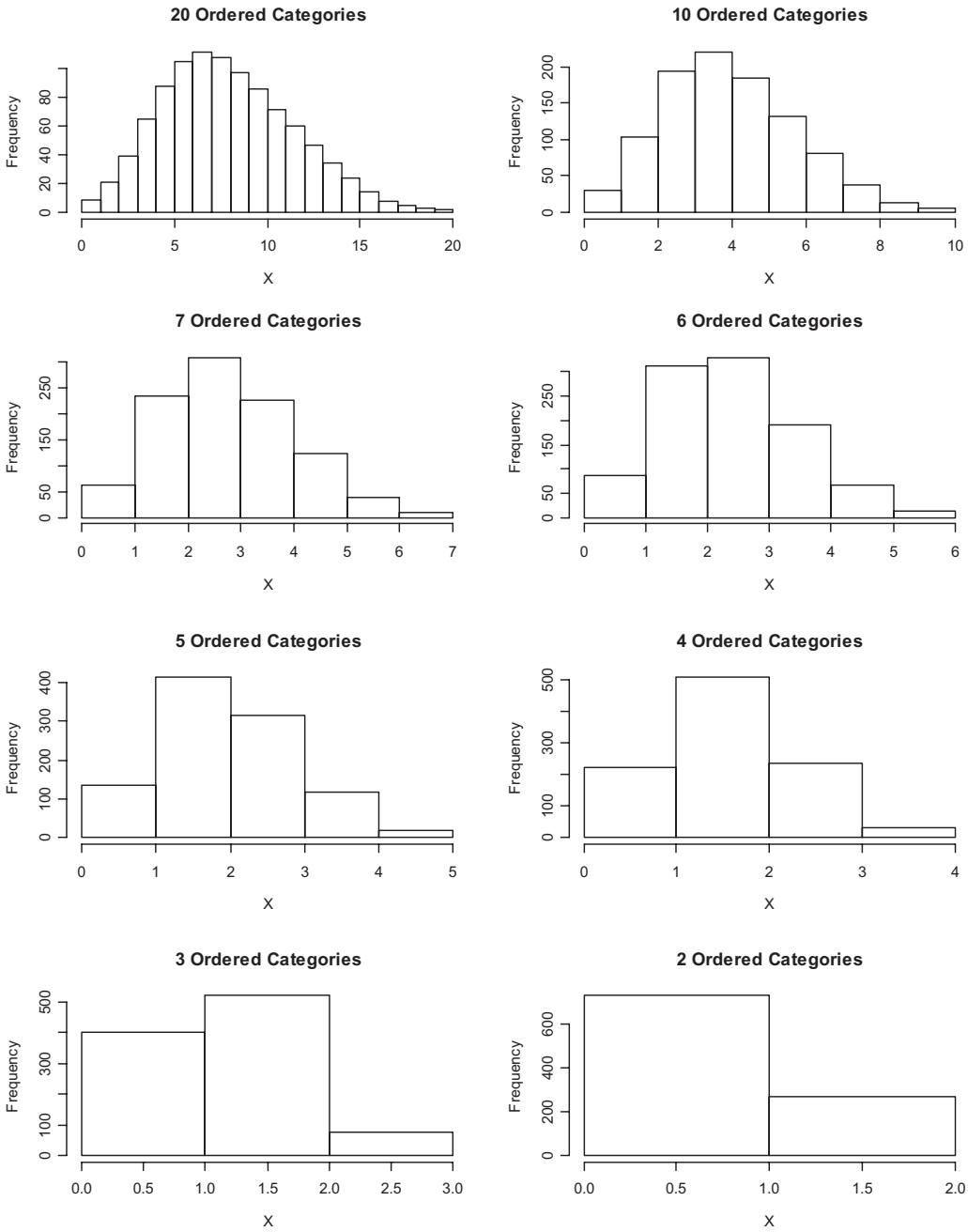

Figure 1. An indicator that was positively skewed within taxon (base rate $p = .25$) and complement, but nonetheless distinguished them with substantial validity ($d = 1.86$), cut into decreasing numbers of ordered categories.

performed with 50 equally spaced cuts beginning and ending 25 cases from either extreme, and MAXCOV and MAXEIG were performed with 50 windows that overlapped 90% with adjacent windows.[2] It should be noted that MAXCOV and MAXEIG are highly redundant procedures that produce nearly identical results (Ruscio, Walters, Marcus, & Kaczetow, 2008). As a result of this redundancy, most taxometric studies do not include both procedures, although both were included in the present study to be thorough. A Monte Carlo study of 100,000 samples revealed

[2] There are no data comparing the performance of taxometric procedures with different numbers of cutting scores or overlapping windows. Several large Monte Carlo studies (e.g., Ruscio et al., 2007, 2008) have used 50 cuts for MAMBAC and 50 windows for MAXCOV–MAXEIG, and holding these values constant across other data conditions (e.g., sample size) provides conservative tests of the utility of taxometric analysis.

nearly identical results for the two procedures. Because all data varied across ordered categories, 10 internal replications were used for each taxometric procedure to stabilize the curves and minimize the obfuscating effect of placing thresholds between tied scores (Ruscio et al., 2006).

Curves were averaged for each MAMBAC, MAXCOV, and MAXEIG analysis and compared with the averaged curves for 10 samples apiece of taxonic and dimensional comparison data. All comparison data were generated with an algorithm described in Ruscio, Ruscio, and Meron (2007) and refined by Ruscio and Kaczetow (2008) that uses a bootstrap technique to reproduce the unique distributional and correlational characteristics of the data. Taxonic comparison data were generated with the base rate classification method, which assigns individuals with the highest indicator total scores to the taxon according to the mean base rate estimate obtained from the analyses of the target data set (Ruscio, in press). Because some of the ordered categorical data spanned very narrow ranges of values (e.g., dichotomous data when $C = 2$), on occasion the reproduction of within-group indicator correlations could not proceed because there was no variance on one or more indicators within a group. When this occurred, a small amount of random variance ($M = 0$, $SD = .0001$) was added to any indicator with zero variance. This does not introduce biases into the results (the expected value for all *r*s for indicators with this random variance added would be 0), but it does enable correlations to be calculated and reproduced.

To quantify the relative fit of the averaged curve for the target data set to that of each type of comparison data, the comparison curve fit index (CCFI) was calculated (Ruscio et al., 2006, 2007). The CCFI can range from 0 to 1, with lower values more indicative of dimensional structure, higher values more indicative of taxonic structure, and .50 representing the most ambiguous results. A number of studies have shown this objective index to be at least as valid as any alternative approach to interpreting taxometric curves (Ruscio, 2007; Ruscio & Marcus, 2007; Ruscio, Meron, & Ruscio, 2007).

Outcome measures. For some analyses, the continuous distribution of CCFI values was examined. Concordance between the actual latent structure of a data set (taxonic vs. dimensional) and the CCFI was measured with the area under the curve (AUC) statistic calculated from a receiver operating characteristic analysis. The AUC is independent of base rates and ranges from .00 to 1.00, with 1.00 indicating perfect accuracy, .00 perfect inaccuracy, and .50 chance-level accuracy.

For other analyses, two dependent variables were constructed to distinguish correct versus incorrect CCFI values. The dichotomous dependent variable measured whether the outcome was scored as correct (1) or incorrect (0) with a CCFI threshold of .50; for taxonic data, CCFI $>$ .50 was scored as correct, and CCFI $<$ .50 was scored as incorrect, and the opposite was true for dimensional data. Because this dichotomous variable counted as correct or incorrect many CCFI values that investigators perhaps should consider to be ambiguous, a trichotomous dependent variable measured whether the outcome was scored as correct (1), ambiguous (.5), or incorrect (0) with CCFI thresholds of .40 and .60. The direction of correct results was the same as for the dichotomous dependent variable, and results were scored as ambiguous when .40 $<$ CCFI $<$ .60. (No CCFI values of precisely .40, .50, or .60 were observed.)

Repeated measures analysis of variance (ANOVA) was used to test for main effects or interactions of method (traditional vs. summed input), *C*, and *k*, with repeated measures on method. We did not calculate *p* values for two reasons. First, our dichotomous and trichotomous dependent variables would have led to violations of parametric ANOVA assumptions (e.g., normality and homoscedasticity). Second, because of the large sample size for these analyses, even statistically significant results may be of trivial magnitude. Thus, we calculated ω^2 as a measure of effect size for each main effect or interaction (Olejnik & Algina, 2003), and we constructed 95% confidence intervals by the percentile bootstrap method (Efron & Tibshirani, 1993) with $B = 5{,}000$ for each analysis. Thus, we were able to focus our attention on a measure of effect size using a method for constructing confidence intervals that does not require the usual parametric assumptions.

Results

The results for the six repeated measures ANOVAs performed for each procedure (MAMBAC, MAXCOV, and MAXEIG) and each dependent variable (dichotomous and trichotomous accuracy measures) are summarized in Table 1 as 95% confidence intervals of ω^2 effect sizes. Rather than consider statistical significance, we used a threshold of $\omega^2 \leq .01$ to classify effects as trivial in

Table 1
Ninety-five Percent Confidence Intervals for ω^2 Calculated From Analysis of Variance Results

	Dichotomous DV			Trichotomous DV		
Effect	MAMBAC	MAXCOV	MAXEIG	MAMBAC	MAXCOV	MAXEIG
C	.069, .125	.011, .051	.013, .053	.073, .138	.011, .055	.015, .062
k	.000, .010	.002, .025	.002, .024	.000, .013	.003, .033	.001, .025
$C \times k$	.013, .041	.020, .059	.020, .056	.016, .048	.025, .071	.026, .072
M	.000, .000	.007, .013	.016, .025	.000, .000	.013, .020	.045, .060
$M \times C$	.000, .000	.000, .003	.000, .001	.000, .000	.001, .006	.000, .004
$M \times k$	.000, .000	.002, .006	.002, .006	.000, .000	.004, .009	.008, .014
$M \times C \times k$	.000, .000	.000, .002	.000, .004	.000, .000	.000, .002	.000, .004

Note. Ninety-five percent confidence intervals for ω^2 were estimated with the percentile bootstrap method with $B = 5{,}000$ for each analysis. Dichotomous DV (dependent variable) = outcome scored as correct (1) or incorrect (0); trichotomous DV = outcome scored as correct (1), ambiguous (.5), or incorrect (0); MAMBAC = mean above minus below a cut; MAXCOV = maximum covariance; MAXEIG = maximum eigenvalue; *C* = number of ordered categories (2, 3, 4, 5, 6, 7, 10, or 20); *k* = number of indicators (3, 4, 5, 6, 7, or 8); *M* = method (traditional or summed-input indicators).

magnitude. For MAMBAC, there were no main or interactive effects with respect to method. MAMBAC-T achieved a mean accuracy of .818 ($SD = .386$) and MAMBAC-S a mean accuracy of .819 ($SD = .385$) on the dichotomous dependent variable, and MAMBAC-T achieved a mean accuracy of .789 ($SD = .332$) and MAMBAC-S a mean accuracy of .791 ($SD = .329$) on the trichotomous dependent variable. For MAXCOV and MAXEIG, there was a main effect favoring the traditional over the summed-input method. MAXCOV-T produced significantly more accurate results than MAXCOV-S on both the dichotomous ($M = .896$, $SD = .305$ vs. $M = .819$, $SD = .385$) and trichotomous ($M = .868$, $SD = .267$ vs. $M = .783$, $SD = .315$) dependent variables. MAXEIG-T likewise achieved significantly more accurate results than MAXEIG-S whether outcome was measured dichotomously ($M = .899$, $SD = .301$ vs. $M = .788$, $SD = .409$) or trichotomously ($M = .869$, $SD = .264$ vs. $M = .721$, $SD = .320$). Of the 12 interaction effects involving method in the ANOVAs for MAXCOV and MAXEIG, only one 95% confidence interval exceeded .01, and even that upper limit rounded to .01. We conclude that method did not interact with the number of ordered categories (C) or indicators (k) to a nontrivial extent.

A consistent finding across all three procedures was that accuracy was weakest when C was low, specifically with $C < 4$. Main effects for k were weaker, and trivial for MAMBAC, but in general accuracy increased with larger values of k. For all three procedures, the interaction between C and k showed that variance in accuracy across levels of C decreased with larger values of k. Figure 2 plots the accuracy of MAMBAC, MAXCOV, and MAXEIG (collapsing across input indicator methods) for levels of C and k. For each procedure, it appears that accuracy levels converge with $C \geq 4$ and $k \geq 5$ but that the former restriction is considerably more important than the latter.

Receiver operating characteristic analyses of the continuous CCFI values were performed, with AUC indexing accuracy. The results, which can be found in Table 2, show that MAXCOV-T and MAXEIG-T significantly outperformed (nonoverlapping confidence intervals) MAXCOV-S and MAXEIG-S, respectively, but that MAMBAC-T and MAMBAC-S differed only slightly. Follow-up analyses show the extent to which accuracy differed with relatively few ordered categories ($C < 3$) versus more ordered categories ($C \geq 4$), with few indicators ($k < 5$) versus more indicators ($k \geq 5$), and with $C \geq 4$ and $k \geq 5$. Consistent with the graphs in Figure 2, it is apparent that having at least four ordered categories exerted a more substantial influence on accuracy than having at least five indicators.

The final way that we expressed and examined accuracy levels was by plotting the probability of reaching a correct structural inference based on the CCFI across the continuum of CCFI values. Figure 3 shows these results for each taxometric procedure across all 4,800 target data sets. Unlike previous analyses, this examination of the data shows that larger CCFI values (suggestive of taxonic structure) were more likely to be correct than smaller CCFI values (suggestive of dimensional structure). However, a follow-up analysis revealed that the inaccuracy for small CCFI values was attributable in large part to the inclusion of data sets with $C < 4$ ordered categories. As Figure 3 also shows, replotting the probability of correct structural inferences for the 3,600 target data sets with $C \geq 4$ revealed impressive accuracy for both small and large CCFI values.

Discussion

For MAMBAC, it seems to make little difference whether items are summed to form input indicators. For MAXCOV and MAXEIG, the traditional methods of allocating items to the required input and output indicators consistently outperformed summed-input indicators. The lack of interaction effects between implementation method and the number of ordered categories or the number of indicators fails to support the use of summed-input indicators for MAXCOV or MAXEIG analyses even under circumstances that would seem to favor the summed-input method (e.g., with a large number of dichotomous items). Because we expect that some readers may find this surprising, we emphasize that performance for all procedures degraded considerably with fewer than four ordered categories. Though the results do suggest that it would be preferable to analyze ordered categorical data with the traditional rather than the summed-input method, the results also suggest caution in submitting items with fewer than four ordered categories to taxometric analysis at all.

Study 2

Whereas Study 1 involved artificial data sets whose latent structure was known, our second study involved empirical data whose latent structure was unknown. There is preliminary evidence from recent research on ASPD, as measured by the ASPD module of the Structured Clinical Interview for *DSM-IV* Axis II Personality Disorders (SCID-II; First, Gibbon, Spitzer, Williams, & Benjamin, 1997; Marcus et al., 2006) and the antisocial features scale of the Personality Assessment Inventory (Morey, 2007; Walters, Diamond, et al., 2007), as well as from research on related constructs like psychopathy (Edens et al., 2006; Guay et al., 2007; Walters, Gray, et al., 2007) and criminal lifestyle (Walters, 2007; Walters & McCoy, 2007), to suggest that the latent structure of the construct investigated in the present study (ASPD) may be dimensional. Taxometric analyses were performed with the traditional and summed-input methods to evaluate the utility of applying the guidelines provided by Study 1 to results for real data.

Method

Participants. Three hundred and twenty-seven male inmates from a medium security federal prison served as participants in this study. These individuals had an average age of 36.26 years ($SD = 10.13$) and a mean educational level of 11.54 years ($SD = 1.55$). The racial breakdown for this sample was 16.5% White, 71.6% Black, 10.4% Hispanic, and 1.5% Asian or Native American. Three quarters (74.6%) of the sample listed their marital status as single, with the remaining 18.7%, 5.8%, and 0.9% of participants listing their marital status as married, divorced, and widowed, respectively. A drug crime was the modal confining offense (42.5%) for participants in this study, with 19.0% serving time for a firearm violation, 13.1% for robbery, 9.2% for a violent offense, 6.1% for a property crime, and 10.1% for a miscellaneous offense.

Measure. The first author administered the ASPD module of the SCID-II (First et al., 1997) to all participants. A random sample of 10% of participants ($n = 33$) was independently interviewed by a second mental health worker within 1 week of the original interview for the purpose of evaluating interrater reliability. An

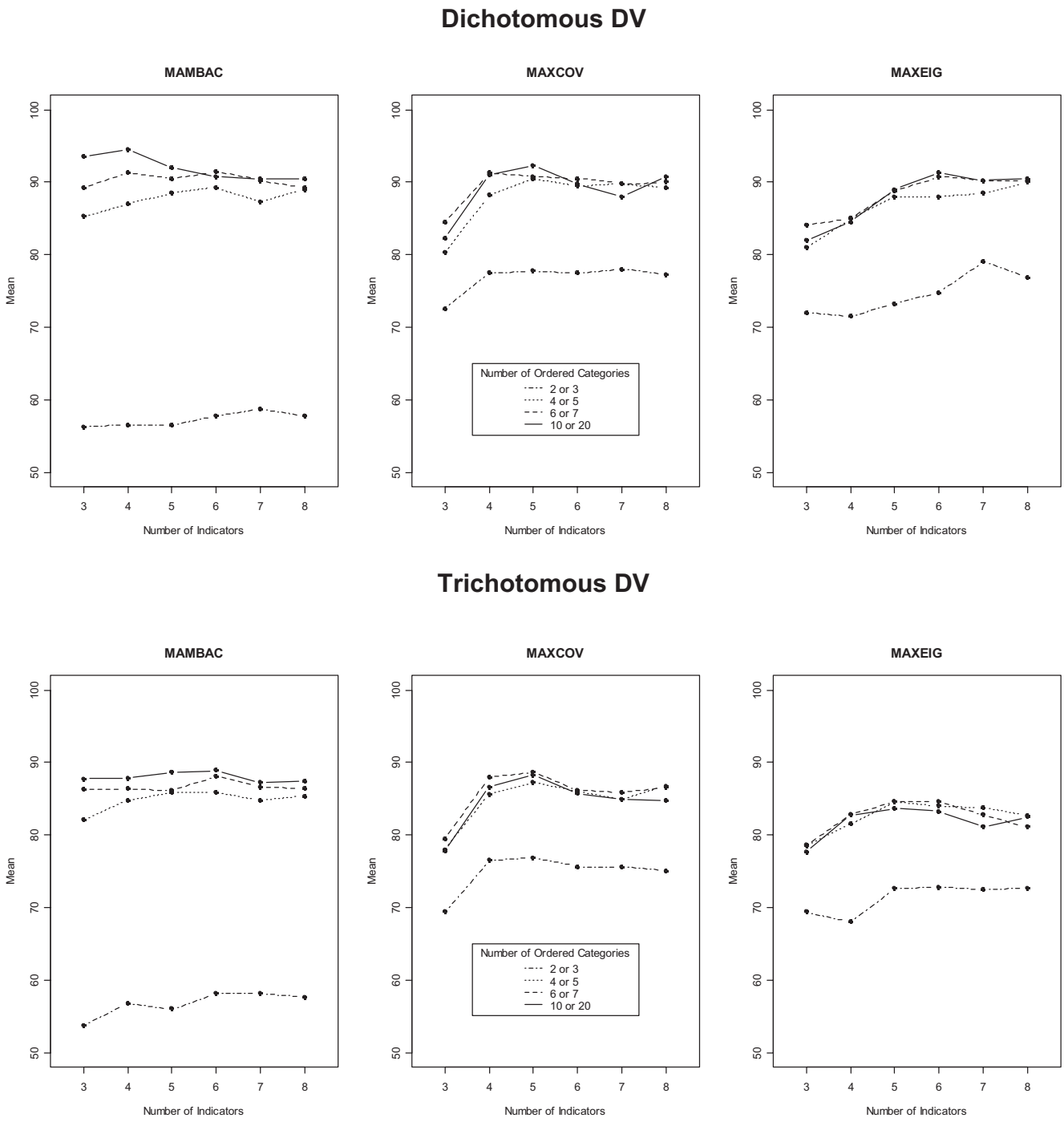

Figure 2. Mean scores for each taxometric procedure, collapsing traditional and summed-input indicator methods, across numbers of ordered categories (collapsed in pairs for clarity of presentation) and numbers of indicators. Dichotomous dependent variable (DV) = outcome scored as correct (1) or incorrect (0); trichotomous DV = outcome scored as correct (1), ambiguous (.5), or incorrect (0). MAMBAC = mean above minus below a cut; MAXCOV = maximum covariance; MAXEIG = maximum eigenvalue.

intraclass correlation coefficient of .91 was obtained for the total ASPD symptom count, and $\kappa = .74$ for a diagnosis of ASPD. The four factor-based indicators identified in an earlier taxometric investigation of the ASPD by Marcus et al. (2006) were used as indicators in the present study: Adult Antisocial (six items), Childhood Non-Assaultive Crimes (six items), Physical Violence (six items), and Childhood Severe Antisocial Behaviors (five items). Study 1 suggests that taxometric results are most likely to be

Table 2
Receiver Operating Characteristic Results for Comparison Curve Fit Index Values in the Total Sample and Subsamples

Total sample	*N*	MAMBAC-T	MAMBAC-S	MAXCOV-T	MAXCOV-S	MAXEIG-T	MAXEIG-S
	4,800	.877 (.867, .888)	.886 (.876, .896)	.954 (.948, .960)	.908 (.900, .916)	.954 (.949, .960)	.872 (.862, .881)
$C \geq 4$	3,600	.974 (.969, .978)	.976 (.971, .980)	.983 (.979, .986)	.930 (.922, .938)	.982 (.979, .986)	.896 (.885, .906)
$C < 4$	1,200	.599 (.567, .631)	.623 (.591, .654)	.855 (.834, .877)	.847 (.825, .869)	.860 (.839, .881)	.791 (.766, .816)
$k \geq 5$	3,200	.894 (.882, .906)	.901 (.889, .912)	.968 (.962, .974)	.909 (.899, .919)	.972 (.966, .978)	.869 (.857, .882)
$k < 5$	1,600	.851 (.831, .871)	.863 (.843, .882)	.917 (.904, .931)	.906 (.892, .921)	.912 (.898, .927)	.883 (.867, .899)
$C \geq 4$ and $k \geq 5$	2,400	.985 (.981, .990)	.986 (.983, .990)	.995 (.993, .997)	.929 (.919, .939)	.995 (.993, .997)	.891 (.878, .904)

Note. Table entries are areas under the receiver operating characteristic curve (calculated with distribution-free geometric formula), with asymptotic 95% confidence intervals in parentheses. *N* = number of data sets; MAMBAC-T = traditional mean above minus below a cut; MAMBAC-S = summed-input mean above minus below a cut; MAXCOV-T = traditional maximum covariance; MAXCOV-S = summed-input maximum covariance; MAXEIG-T = traditional maximum eigenvalue; MAXEIG-S = summed-input maximum eigenvalue; *C* = number of ordered categories; *k* = number of indicators.

informative when there are at least five indicators, each with at least four ordered categories. The present data were near the borderline established by these criteria: There were only four indicators, but each had at least four ordered categories (three had seven ordered categories, scores of 0–6, and one had four ordered categories, scores of 0–3).

Procedure. Taxometric analyses were performed with Ruscio's (2008) taxometric software. The MAMBAC, MAXCOV, and MAXEIG procedures were run five times apiece, and the CCFI results were averaged for each procedure. As in Study 1, 50 equally spaced cuts were made with MAMBAC, 50 overlapping windows were used with MAXCOV, and 10 internal replications were used in all analyses. There were two principal procedural differences, however. First, because of the smaller number of analyses involved in Study 2 than in Study 1, populations of taxonic and dimensional comparison data were used to obtain 100 rather than 10 bootstrap samples. Second, because MAXCOV-S and MAXEIG-S were redundant with only four items, MAXCOV-S was computed with 15 nonoverlapping intervals instead of 50 overlapping windows to produce nonredundant results (Ruscio et al., 2006). As in Study 1, the CCFI served as the measure of relative fit for taxonic and dimensional structural models.

Results

Table 3 lists the means, standard deviations, skew, kurtosis, and estimated validity of the four ASPD indicators; we calculated validity as the standardized mean difference (Cohen's *d*) across cases assigned to putative taxon and complement groups using the base rate classification method. Three of the four indicators displayed a moderately high degree of nonnormality (skew and kurtosis), and two of the validity estimates fell slightly below Meehl's (1995) recommended cutoff ($d = 1.25$); the average estimated validity of $d = 1.68$ was well above Meehl's cutoff. Estimated within-group correlations were not problematic for these data: The mean interindicator correlation was substantially larger in the full sample ($r = .45$) than within subsamples of cases assigned to the putative taxon ($r = .25$) or complement ($r = .09$) under the base rate classification method.

The CCFI results for the taxometric analyses are summarized in Table 4. With a single threshold (dimensional if CCFI $< .50$; taxonic if CCFI $> .50$), the range of CCFI values and the mean CCFI value for five of the six series of analyses were more consistent with dimensional than taxonic latent structure; the only exception was for the MAXEIG-S procedure, in which case the range and mean were both slightly above .50. With dual thresholds to set aside ambiguous results (i.e., $.40 <$ CCFI $< .60$), all nonambiguous CCFI values were more consistent with dimensional than taxonic structure. All MAXEIG-S results were ambiguous, as were some of the MAXCOV-S and MAMBAC-T results. Illustrative results for each procedure are reproduced in Figure 4.

Discussion

The results of Study 1 showed that for MAXCOV and MAXEIG, the traditional method should be less likely to yield ambiguous results, whereas the distinction between MAMBAC-T and MAMBAC-S should make less of a difference. Although the latent structure of ASPD is far from established, there is tentative support for a dimensional interpretation. A researcher who followed the advice offered by Study 1 and performed only MAMBAC (either way), MAXCOV-T, and MAXEIG-T would have obtained clearer and more consistent findings, with relatively few ambiguous or potentially misleading results. In supporting dimensional latent structure, the MAMBAC-S, MAXCOV-T, and MAXEIG-T results tend to confirm the outcome of previous taxometric research on ASPD (Haslam, 2007; Marcus et al., 2006; Walters, Diamond, et al., 2007). Consistent with the findings in Study 1, we conclude that there is no evidence to support preference for MAXCOV-S or MAXEIG-S rather than MAXCOV-T or MAXEIG-T under any conditions. However, it is possible that MAMBAC-S may have a slight advantage over MAMBAC-T when indicators are skewed, leptokurtic, and vary across few ordered categories. This is consistent with the slightly, but not statistically significantly, greater accuracy for MAMBAC-S than MAMBAC-T in Study 1. The small sample size and moderately high skew in several indicators aside, CCFI consistency is a key indicator of accuracy regardless of sample size or indicator skew (Ruscio et al., 2008).

General Discussion

MAMBAC differentiated taxonic and dimensional data about equally well with either the traditional or the summed-input indicator method, but MAXCOV and MAXEIG yielded significantly more valid structural inferences with the traditional method. Because these main effects were not qualified by interactions with the number of indicators or number of ordered categories, these find-

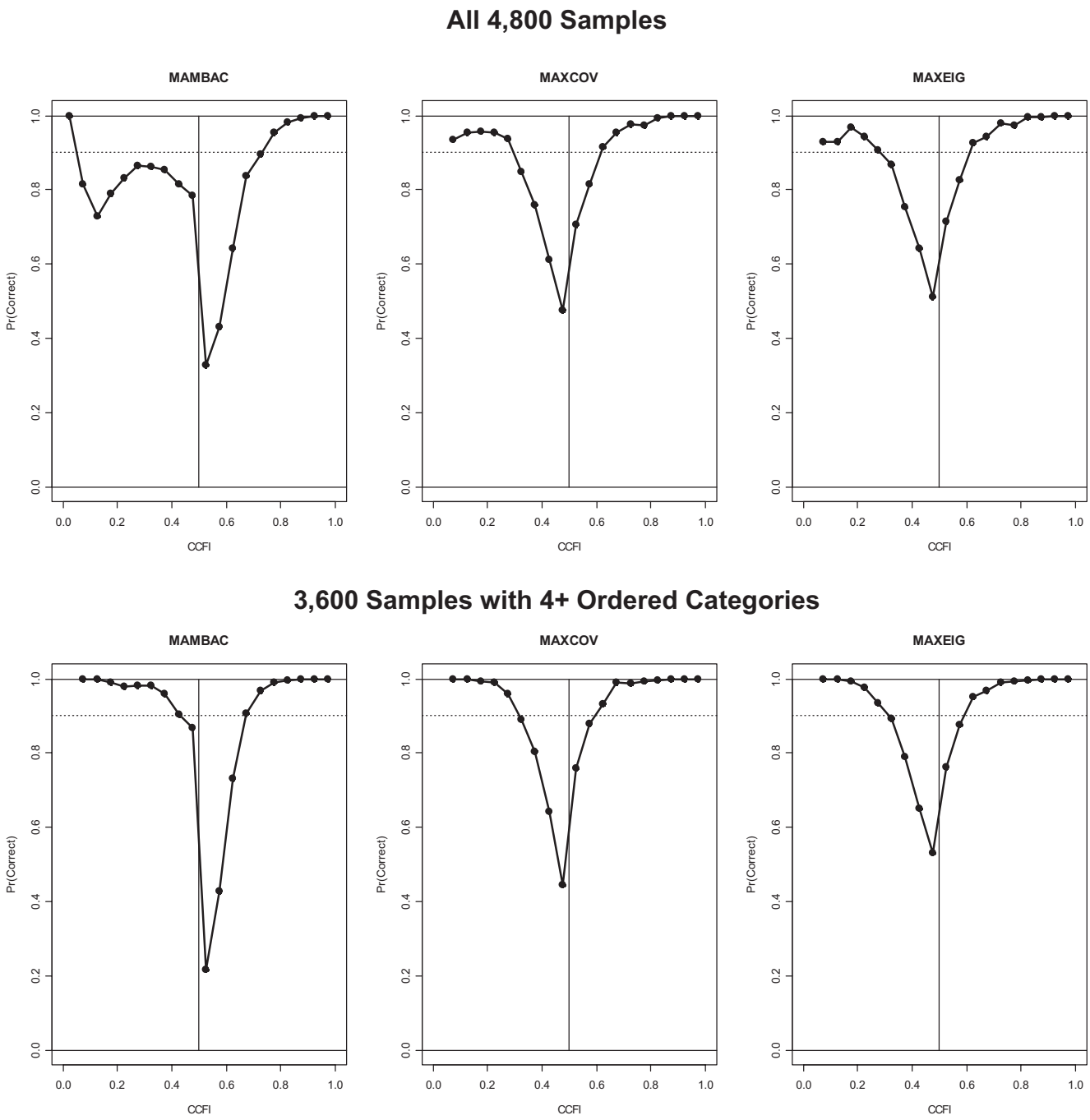

Figure 3. Probability (Pr) of comparison curve fit index (CCFI) being correct for each taxometric procedure, collapsing traditional and summed-input indicator methods. MAMBAC = mean above minus below a cut; MAXCOV = maximum covariance; MAXEIG = maximum eigenvalue.

ings provide no support for using the summed-input method. For MAXCOV and MAXEIG, the nontrivial increase in the accuracy with which taxonic and dimensional data were identified with the traditional rather than the summed-input indicator method suggests that the traditional method should be used.

This advice counters the convention for taxometric analyses of items with few ordered categories, such as the dichotomous data that prompted Gangestad and Snyder (1985) to develop the summed-input method. However, it should be noted that Gangestad and Synder introduced this innovative technique at a time when taxometric analyses were rare and few studies had examined the method itself. At that time, it was customary to locate thresholds (for MAMBAC) or form-ordered subsamples (for MAXCOV; MAXEIG did not exist) at unique scores on the input indicator. As

Table 3
Descriptive Statistics and Validity Estimates for Four Indicators From the Antisocial Personality Disorder (ASPD) Module of the Structured Clinical Interview for DSM-IV *Axis II Personality Disorders (SCID-II)*

ASPD indicator	*M*	*SD*	Skew[a]	Kurtosis[b]	Validity (Cohen's *d*)[c]
Adult Antisocial	12.83	2.89	−0.19	−0.72	1.17
Childhood Non-Assaultive Crimes	8.06	2.91	1.51	1.48	2.80
Physical Violence	8.34	2.17	1.80	3.52	1.72
Childhood Severe Antisocial Behaviors	3.31	0.86	3.26	12.16	1.05

Note. $N = 327$.
[a] The standard error of measurement for skew in this sample was .14. [b] The standard error of measurement for kurtosis in this sample was .27. [c] Estimated with a SCID-II ASPD diagnosis positive for antisocial personality disorder.

a result, some adaptation of the traditional input indicator method was required. For example, with dichotomous data, only a single threshold can be placed between the two unique scores (e.g., 0 and 1) for MAMBAC analysis, and only two subsamples can be formed with the two unique scores for MAXCOV (or MAXEIG) analysis. No conclusions can be drawn from a MAMBAC curve with a single data point or a MAXCOV–MAXEIG curve with two data points. Thus, if one insists that thresholds be placed between unique scores or subsamples be formed on the basis of unique scores, a number of dichotomous items would have to be summed to form input indicators.

Study 1 provides no support for the summed-input method. So how can one perform taxometric analyses with the ordered categorical data typical of assessment instruments used in psychology and related disciplines? Even when there are few unique scores, large numbers of thresholds or subsamples can be used. Although it may seem peculiar to use many more cuts (in MAMBAC) or subsamples (in MAXCOV or MAXEIG) than there are unique values along an input indicator, Ruscio et al. (2006, pp. 111–115, 153–154) discussed why this can be effective and how to take advantage of internal replications to clarify the results. Because we anticipate that some readers still may be wary of submitting data with as few as four ordered categories to taxometric analyses using the traditional method, we present the results for two illustrative data sets—one taxonic and one dimensional, each with $N = 1{,}000$, $C = 4$, and $k = 5$. All other data parameters of these samples were prototypical of the conditions in Study 1, specifically the midpoints of each range (for the taxonic data, $p = .275$, $d = 1.625$, $r = .15$, $g = .15$, $h = .075$, $VR = 2.5$; for the dimensional data, $r_{xy} = .40$, $g = .15$, and $h = .075$). MAMBAC, MAXCOV, and MAXEIG were performed with the traditional method, and the results are presented in Figure 5. Note that each two-graph panel superimposes the results for the target data set (points connected by dark lines) on those for the standard error bounds for 100 samples apiece of taxonic and dimensional comparison data. CCFI values for the taxonic data were .651 (MAMBAC), .777 (MAXCOV), and .745 (MAXEIG), and those for the dimensional data were .241 (MAMBAC), .189 (MAXCOV), and .209 (MAXEIG). Even though the curves shown in Figure 5 do not follow the expectations for prototypical taxonic or dimensional data, each of these CCFI values unambiguously and correctly differentiated between taxonic and dimensional structure. In the absence of the curves for comparison data and the CCFI, the results for these target data sets would be exceptionally difficult to interpret. This underscores the value of calculating an objective fit index based on comparison data, which controls for idiosyncratic aspects of a data set such as indicator distributions that vary asymmetrically across ordered categories rather than normally along a continuum.

In addition to the weak method effect for MAXCOV and MAXEIG, the absence of such an effect for MAMBAC, and the absence of interactions between method and other factors, there were main effects and interactions for the number of ordered categories and the number of indicators. Accuracy was much better for $C \geq 4$ than for $C < 4$. To a lesser extent, accuracy was better for $k \geq 5$ than for $k < 5$, and accuracy converged across levels of C with larger values of k. This would seem to suggest that researchers need to exercise caution when interpreting the results of taxometric studies based on dichotomous or trichotomous indicators. At the same time, there appears to be relatively little improvement in performance between as few as 4 and as many as 20 ordered categories, which approximates a continuous distribution fairly well. It is not uncommon for an assessment instrument to contain a large number of items, and the present results suggest some advice for how to work with such data to construct indicators for taxometric analyses. If individual items vary across relatively few ordered categories—especially if they are only dichotomous or trichotomous response scales—one should consider forming composites from items tapping related aspects of the target construct. For example, if 40 dichotomous items can be assigned in roughly equal numbers to five indicators that are conceptually and empirically nonredundant, such indicators would vary across a sufficiently large number of scores to afford more informative taxometric results than analyses of the original dichotomous items. Forming composites prior to analysis, rather than analyzing a subset of the original items, incorporates more of the available data into the analyses. Provided that items are combined wisely, the composites should be more reliable and valid than individual items, which in turn should improve the performance of each taxometric procedure.

Table 4
Base Rate, Validity, and Comparison Curve Fit Index (CCFI) Results for Traditional and Summed-Input MAMBAC, MAXCOV, and MAXEIG

Procedure	Taxon base rate estimates		CCFI
	M	*SD*	
MAMBAC traditional	.21–.26 (.24)	.12–.16 (.14)	.384–.505 (.438)
MAMBAC summed-input	.31–.32 (.31)	.08–.09 (.08)	.299–.386 (.351)
MAXCOV traditional	.21–.25 (.22)	.13–.21 (.15)	.273–.350 (.309)
MAXCOV summed-input	.29–.36 (.34)	.31–.36 (.35)	.341–.404 (.368)
MAXEIG traditional	.14–.15 (.14)	.03–.04 (.03)	.359–.392 (.371)
MAXEIG summed-input	.20–.21 (.21)	.23–.23 (.23)	.509–.550 (.523)

Note. Numbers in each column separated by a dash are the range, and number in parentheses is the mean across five analyses; $N = 327$. MAMBAC = mean above minus below a cut; MAXCOV = maximum covariance; MAXEIG = maximum eigenvalue; M = mean of the taxon base rate across curves; SD = standard deviation of the taxon base rate across curves.

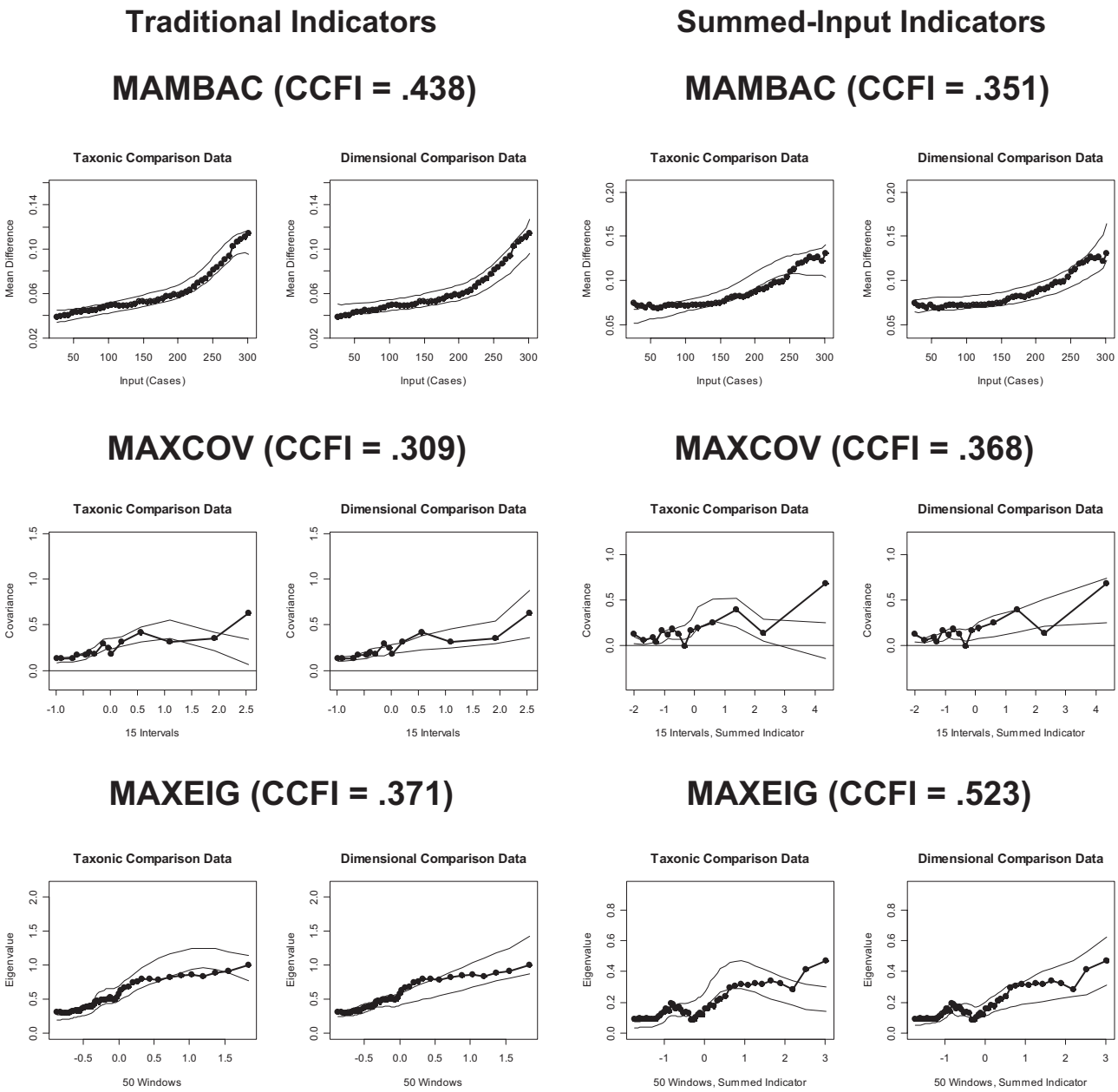

Figure 4. Taxometric analyses of the four antisocial personality disorder indicators. Within each two-graph panel, results for the target data set (points connected by dark lines) are superimposed on those for the standard error bounds for 100 samples apiece of taxonic and dimensional comparison data. Illustrative graphs are shown along with the mean comparison curve fit index (CCFI) values from each series of analyses (as reported in Table 4). MAMBAC = mean above minus below a cut; MAXCOV = maximum covariance; MAXEIG = maximum eigenvalue.

The current investigation examined one implementation decision faced in a taxometric study—whether to sum items to form input indicators—and found that summed-input indicators offered no advantage over the traditional method. There remain many important implementation decisions that require empirical attention. The number of cuts in MAMBAC, and the number of intervals or windows in MAXCOV or MAXEIG, have usually been decided by convention and personal preference. It would be helpful to know what effect increasing or decreasing the number of cutting scores or subsamples has on the overall accuracy of a taxometric procedure and whether there are data conditions that moderate this effect. A related decision for MAXCOV and MAXEIG is whether to use intervals, which do not overlap, as is the tradition with MAXCOV, or windows, which do overlap, as is the tradition with MAXEIG. There is nothing preventing researchers from using windows with MAXCOV or intervals with MAXEIG. Ruscio et al. (2006) explained why windows may be more useful than intervals most or all of the time, but a study that

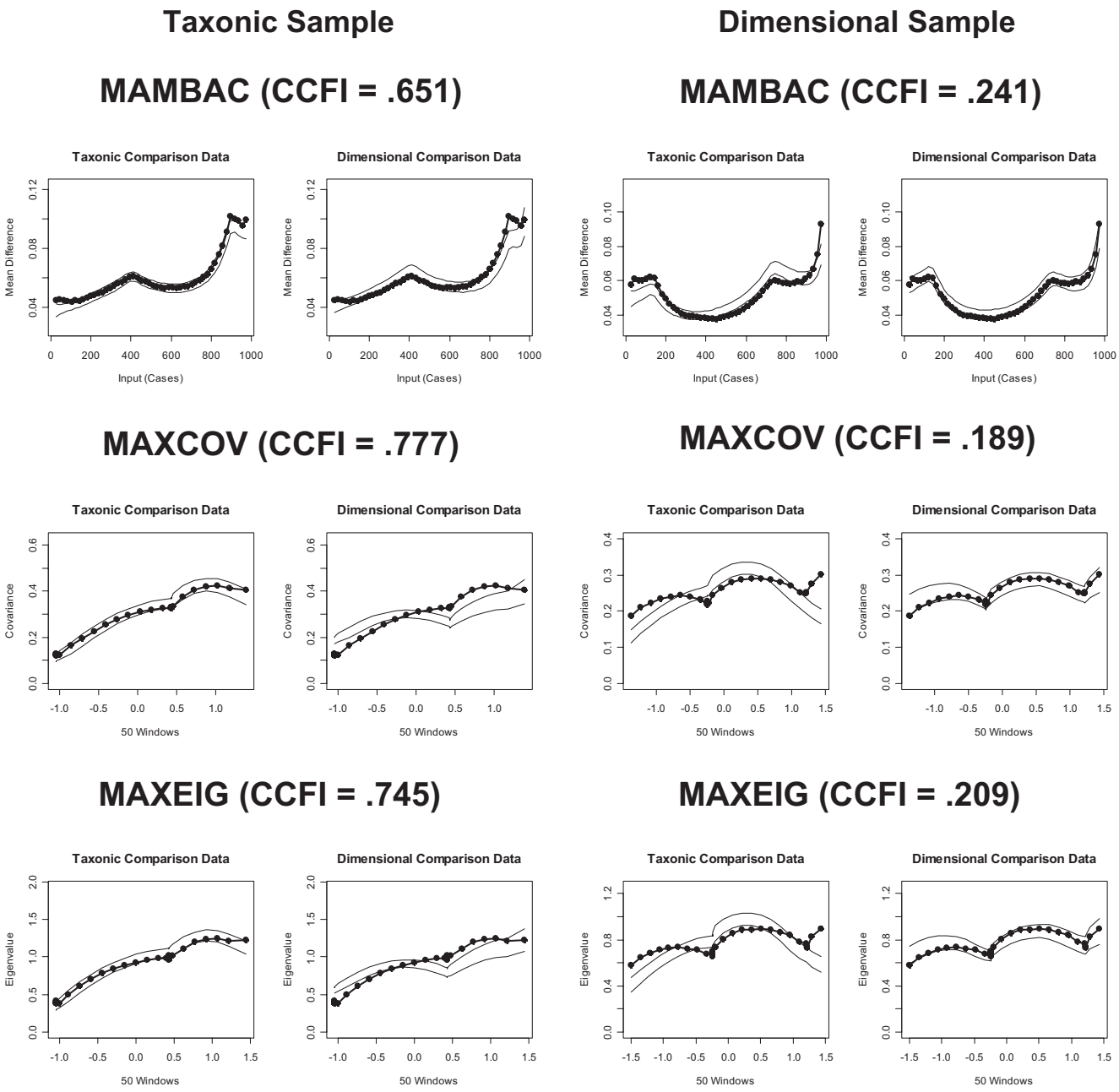

Figure 5. Taxometric analyses of illustrative taxonic and dimensional data with $C = 4$ and $k = 5$. Each procedure was implemented in the traditional manner (without using summed-input indicators). Within each two-graph panel, results for the target data set (points connected by dark lines) are superimposed on those for the standard error bounds for 100 samples apiece of taxonic and dimensional comparison data. MAMBAC = mean above minus below a cut; MAXCOV = maximum covariance; MAXEIG = maximum eigenvalue; CCFI = comparison curve fit index.

crosses procedures (MAXCOV, MAXEIG), subsample types (intervals, windows), and number of subsamples would be a welcome addition to the literature.

These are only some of the potentially important implementation decisions that should be addressed empirically to harness the full power of the taxometric method. Ruscio et al. (2006) provided a more thorough discussion of implementation decisions for each procedure and offered some suggestions for making smart choices. Understanding the latent structure of psychological constructs is essential in developing effective assessment strategies and has led to a sharp rise in the number of articles published in psychological journals in which the taxometric method has been used to investigate the latent structure of constructs that various psychological procedures are designed to measure. Because the taxometric

method is being used with increased frequency by specialists in psychological assessment, research investigating the methodology itself should prove useful in the eventual development, refinement, and application of assessment instruments.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Bernstein, A., Zvolensky, M. J., Norton, P. J., Schmidt, N. B., Taylor, S., Forsyth, J. P., et al. (2007). Taxometric and factor analytic models of anxiety sensitivity: Integrating approaches to latent structural research. *Psychological Assessment, 19,* 74–87.

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

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* San Francisco: Chapman & Hall.

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B., & Benjamin, L. (1997). *Structured Clinical Interview for DSM-IV Axis II personality disorders (SCID-II*). Washington, DC: American Psychiatric Press.

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

Guay, J.-P., Ruscio, J., Knight, R. A., & Hare, R. D. (2007). A taxometric analysis of the latent structure of psychopathy: Evidence for dimensionality. *Journal of Abnormal Psychology, 116,* 701–716.

Hare, R. D. (2003). *The Hare Psychopathy Checklist—Revised manual* (2nd ed.). Toronto, Canada: Multi-Health Systems.

Haslam, N. (2007). The latent structure of mental disorders: A taxometric update on the categorical vs. dimensional debate. *Current Psychiatry Reviews, 3,* 172–177.

Hoaglin, D. C. (1985). Summarizing shape numerically: The *g*-and-*h* distributions. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Exploring data, tables, trends, and shapes* (pp. 461–511). New York: Wiley.

Marcus, D. K., Lilienfeld, S. O., Edens, J. F., & Poythress, N. G. (2006). Is antisocial personality disorder continuous or categorical? A taxometric analysis. *Psychological Medicine, 36,* 1571–1581.

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,* 266–275.

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

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

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

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

Morey, L. C. (2007). *The Personality Assessment Inventory (PAI): Professional manual* (2nd ed.). Lutz, FL: Psychological Assessment Resources.

Olatunji, B. O., & Broman-Fulks, J. J. (2007). A taxometric study of the latent structure of disgust sensitivity: Converging evidence for dimensionality. *Psychological Assessment, 19,* 437–448.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8,* 434–447.

Ruscio, J. (2007). Taxometric analysis: An empirically-grounded approach to implementing the model. *Criminal Justice and Behavior, 34,* 1588–1622.

Ruscio, J. (2008). *Taxometric programs for the R computing environment: User's manual* [Computer software and manual]. Retrieved June 12, 2008, from http://www.taxometricmethod.com

Ruscio, J. (in press). Assigning cases to groups using taxometric results: An empirical comparison of classification techniques. *Assessment.*

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

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

Ruscio, J., & Marcus, D. K. (2007). Detecting small taxa using simulated comparison data: A reanalysis of Beach, Amir, and Bau's (2005) data. *Psychological Assessment, 19,* 241–246.

Ruscio, J., & Ruscio, A. M. (2002). A structure-based approach to psychological assessment: Matching measurement models to latent structure. *Assessment, 9,* 4–16.

Ruscio, J., Ruscio, A. M., & Meron, M. (2007). Applying the bootstrap to taxometric analysis: Generating empirical sampling distributions to help interpret results. *Multivariate Behavioral Research, 42,* 349–386.

Ruscio, J., Walters, G. D., Marcus, D. K., & Kaczetow, W. (2008). *Comparing the relative fit of categorical and dimensional latent variable models using consistency tests.* Manuscript submitted for publication.

Ruscio, J., Zimmerman, M., McGlinchey, J. B., Chelminski, I., & Young, D. (2007). Diagnosing major depressive disorder XI: A taxometric investigation of the structure underlying DSM-IV symptoms. *Journal of Nervous and Mental Disease, 195,* 10–19.

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

Solomon, A., Ruscio, J., Seeley, J. R., & Lewinsohn, P. M. (2006). A taxometric investigation of unipolar depression in a large community sample. *Psychological Medicine, 36,* 973–985.

Strong, D. R., Glassmire, D. M., Frederick, R. I., & Greene, R. L. (2006). Evaluating the latent structure of the MMPI-2 *F(p)* scale in a forensic sample: A taxometric analysis. *Psychological Assessment, 18,* 250–261.

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

Walters, G. D. (1995). The Psychological Inventory of Criminal Thinking Styles: Part I. Reliability and preliminary validity. *Criminal Justice and Behavior, 22,* 307–325.

Walters, G. D. (2007). The latent structure of the criminal lifestyle: A taxometric analysis of the Lifestyle Criminality Screening Form and Psychological Inventory of Criminal Thinking Styles. *Criminal Justice and Behavior, 34,* 1623–1637.

Walters, G. D., Diamond, P. M., Magaletta, P. R., Geyer, M. D., & Duncan, S. A. (2007). Taxometric analysis of the antisocial features scale of the Personality Assessment Inventory in federal prison inmates. *Assessment, 14,* 351–360.

Walters, G. D., Gray, N. S., Jackson, R. L., Sewell, K. W., Rogers, R., Taylor, J., & Snowden, R. J. (2007). A taxometric analysis of the Psychopathy Checklist: Screening Version (PCL:SV): Further evidence of dimensionality. *Psychological Assessment, 19,* 330–339.

Walters, G. D., & McCoy, K. (2007). Taxometric analysis of the Psychological Inventory of Criminal Thinking Styles in incarcerated offenders and college students. *Criminal Justice and Behavior, 34,* 781–793.

Received August 10, 2008
Revision received November 7, 2008
Accepted November 19, 2008 ■