Differentiating Categorical and Dimensional Data With Taxometric Analysis: Are Two Variables Better Than None?

John Ruscio
The College of New Jersey

Glenn D. Walters
Federal Correctional Institution–Schuylkill, Minersville, Pennsylvania

Taxometric analyses have proven helpful for distinguishing categorical and dimensional data. Many taxometric procedures require at least 3 variables for analysis. What if a construct is defined by only 2 conceptually nonredundant characteristics or a data set contains only 2 empirically nonredundant variables? In Study 1, we performed extensive simulations to determine whether informative results can be obtained when only 2 variables are available for taxometric analysis. The mean above minus below a cut (MAMBAC) and maximum slope (MAXSLOPE) procedures, used with parallel analyses of comparison data, successfully differentiated categorical and dimensional structure. With just 2 variables, it seems especially important that indicators vary across as many distinct values as possible and that investigators obtain as large a sample as possible. Additional findings address questions about the most effective way to implement taxometric analyses. In Study 2, the potential utility of 2-variable taxometric analysis is illustrated using data on proactive and reactive childhood aggression, where the results provided strong support for dimensional structure. As long as high-quality data are available, it appears that one can have confidence in the results of taxometric analyses performed with only 2 variables.

Keywords: taxometrics, MAMBAC, MAXSLOPE, comparison data, CCFI

Many data-analytic techniques are available to help determine whether a categorical (latent class, qualitative, taxonic) or dimensional (latent factor, quantitative, continuous) model better fits a set of data. Ascertaining the structure of psychological constructs has important implications for assessment, classification, diagnosis, research design, statistical analysis, and theory (MacCallum, Zhang, Preacher, & Rucker, 2002; Meehl, 1992; Ruscio, Haslam, & Ruscio, 2006). For example, one might focus on classification accuracy when validating a measure that assesses a categorical construct, whereas one might focus on correlations with other variables when validating a measure that assesses a dimensional construct. Likewise, the process of construct validation should involve a consideration of whether the scores on a measure match the latent structure of the construct. Whereas accurate cutting scores are essential for assessing categorical constructs, measures of dimensional constructs should not yield dichotomous scores.

Data-analytic methods to differentiate categorical and dimensional data include cluster analysis (Everitt, Landau, & Leese, 2001), mixture modeling (McLachlan & Peel, 2001), latent class or latent profile analysis (McCutcheon, 1987), and Meehl's (1995) taxometric method. As Ruscio, Walters, Marcus, and Kaczetow

(2010) have discussed, the latter has become popular in research on psychological assessment; their Table 1 (p. 7) summarizes 19 studies published in three leading journals in this area (*Assessment*, *Journal of Personality Assessment*, and *Psychological Assessment*). In some studies, multiple measures have been used to examine the structure of a construct (e.g., depression, psychopathy, symptom overreporting), whereas in other studies the focal point has been the structure of the construct underlying responses to a particular instrument (e.g., Beck Depression Inventory, Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; Psychopathy Checklist—Revised, Hare, 2003). That taxometric research is being published at an accelerating pace attests to the importance of empirically evaluating latent structure for psychological assessment.

Prior studies of taxometric methodology have not considered cases in which only two variables are available for analysis because many taxometric procedures, and especially those that are most popular, require at least three variables. Under certain circumstances, however, an investigator may not have three variables to work with. Because the taxometric method requires that each variable be conceptually and empirically nonredundant (Ruscio et al., 2006), there are at least three reasons why one might have only two variables available for a taxometric analysis. First, a construct might be defined by only two conceptually distinct features (e.g., a clinical diagnosis of delusional disorder, which is composed of just two criteria: nonbizarre delusion and absence of severe impairment). Second, even if a construct is defined by more than two distinct features, it may not be possible to measure some of these in ways that are empirically nonredundant, which can prompt the combination or removal of one or more empirically redundant variables (e.g., whereas one might wish to use items on the Beck Depression Inventory to construct variables representing several

features of syndromal depression, empirically these items seem to load onto two factors: affective/cognitive and somatic). Third, even if it is possible to define and measure a construct via more than two distinct features, an available data set might contain only two variables suitable for analysis (e.g., rather than item-level data from Hare's (2003) Psychopathy Checklist, one might have access only to scores on Factors 1 and 2). Though it is possible to perform a taxometric analysis using only two variables, virtually no data are available to suggest whether this is advisable and, if so, under what conditions. The primary purpose of the present study is to address this issue.

Proactive and reactive childhood aggression is an example of a construct composed of two conceptually distinct features—namely, proactive and reactive aggression. Measuring proactive and reactive childhood aggression normally begins with Dodge and Coie's (1987) Teacher Checklist (TC) rating procedure, the most frequently used measure in this research literature. Each of the two TC scales (proactive, reactive) is made up of three items, and the pool of six items lends itself conceptually and practically to just two variables for analysis. Other rating procedures have been developed to assess proactive and reactive aggression in children (e.g., Brown, Atkins, Osborne, & Milnamow, 1996; Kempes, Matthys, Maassen, van Gooze, & van Engeland, 2006), but because these measures are largely redundant with the TC (Hubbard, McAuliffe, Morrow, & Romano, 2010), they are unlikely to provide additional variables suitable for analysis. Another way to expand the set of variables would be to construct proactive and reactive scales from a self-report measure like Buss and Perry's (1992) 29-item Aggression Questionnaire (AQ) and pair them with the two TC scales. Walters, Ronen, and Rosenbaum (2010) sought to do just this but discovered that the AQ and TC variables correlated too weakly with one another to permit taxometric analysis. This leaves investigators interested in the structure of proactive and reactive aggression in children with little choice but to perform analyses with only two variables.

The availability of just two variables constrains the implementation of Meehl's (1995) taxometric method. Nearly every published taxometric study illustrates one of its invaluable features: The results from multiple data-analytic procedures are examined to determine whether they consistently support a categorical or a dimensional model. Within Meehl's (2004) approach, there are three possible outcomes: consistent evidence in favor of a categorical model (which he referred to as *taxonic*), consistent evidence in favor of a dimensional model (which he referred to as *nontaxonic*), or ambiguous evidence from which no conclusion should be drawn. Most taxometric procedures also afford consistency checks by analyzing the available data in different configurations. For example, the mean above minus below a cut (MAMBAC; Meehl & Yonce, 1994) procedure produces a plot of mean differences on one variable (the *output indicator*) for each of a large number of cutting scores on another variable (the *input indicator*). By assigning each of k variables to the input and output roles in all possible ways, one obtains $k(k − 1)$ MAMBAC curves. One can examine the shapes of these curves to judge whether they support a categorical or dimensional model. The prototypical MAMBAC curve for categorical data contains a peak near the cutting score that best differentiates members of two groups, whereas the prototypical MAMBAC curve for dimensional data

does not contain a peak; instead, it tends to be concave (see Figure 1, Panel A).

When data characteristics deviate from idealized values (e.g., when variables are skewed), this can affect taxometric curve shapes (Ruscio, Ruscio, & Keane, 2004). For this reason, it has become customary in taxometric analyses to perform parallel analyses of categorical and dimensional comparison data that reproduce important characteristics of the empirical data (e.g., sample size, number of variables, marginal distributions, correlation matrix) using known structural models. The comparison curve fit index (CCFI) quantifies the relative fit of these two models to the data at hand, and many studies have shown that the CCFI distinguishes data generated using these models well (e.g., Ruscio, 2007; Ruscio & Kaczetow, 2009; Ruscio & Marcus, 2007; Ruscio, Ruscio, & Meron, 2007; Ruscio, Walters, Marcus, & Kaczetow, 2010; Walters, McGrath, & Knight, 2010; Walters & Ruscio, 2009, 2010). CCFI values range from 0 (strongest possible support for dimensional structure) to 1 (strongest possible support for categorical structure), with .50 an ambiguous value that signals equivalent support for both structures. Panel B in Figure 1 shows the MAMBAC results for a dimensional data set with positively skewed variables. Whereas the empirical data yield a curve that might be interpreted as peaked at the right end, plotting these results along with those for categorical and dimensional comparison data shows that they are much more similar to those for the latter. The CCFI value of .284 provides an objective measure that, because it is well below .50, supports this conclusion.

The number of MAMBAC curves is reduced when fewer variables are available for analysis, but in the limiting case of two variables, most of the other widely used taxometric procedures cannot be performed at all. Maximum covariance (MAXCOV; Meehl & Yonce, 1996), maximum eigenvalue (MAXEIG; Waller & Meehl, 1998), and latent mode (L-Mode; Waller & Meehl, 1998) require at least three variables. The only other taxometric procedure that has been introduced for use with two variables is maximum slope (MAXSLOPE; Grove, 2004; Grove & Meehl, 1993). Using this procedure, one applies a locally weighted scatterplot smoother (LOWESS; Cleveland, 1979) to calculate the slope of a nonlinear regression of one variable on the other. One can plot these slopes by values of the variable placed on the *x*-axis in the original scatterplot to generate a MAXSLOPE curve. A second curve is obtained by reversing the variables' roles as *x* and *y* variables, which produces a second scatterplot and series of slopes. Standardizing the variables prior to analysis allows the resulting curves to be averaged for presentation and interpretation (see Ruscio et al., 2006, for a discussion of averaging taxometric curves and the value of standardizing variables for doing so). By presenting MAXSLOPE results as plots of slopes, users can interpret the results as they would the more familiar MAXCOV or MAXEIG curves that are staples in the taxometric literature. The prototypical MAXSLOPE curve for categorical data contains a peak, whereas the prototypical MAXSLOPE curve for dimensional data does not (see Figure 1, Panel C). Because data characteristics such as skew can influence MAXSLOPE curves, we adapted the CCFI for use with MAXSLOPE. The usual method of calculating the CCFI (see Ruscio & Kaczetow, 2009) was applied to the conditional slopes of the MAXSLOPE plots. Panel D of Figure 1 shows the MAXSLOPE results for the same dimensional data set analyzed earlier using MAMBAC. Despite what might be inter-
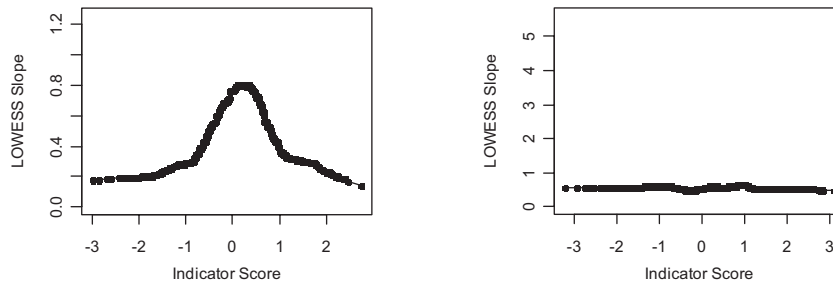
A: Prototypical MAMBAC Curves for Categorical (Left) and Dimensional (Right) Data
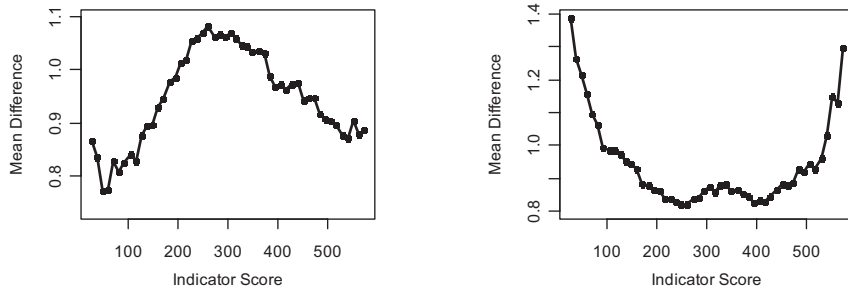
B: MAMBAC Analysis of Dimensional Data with Positively Skewed Variables
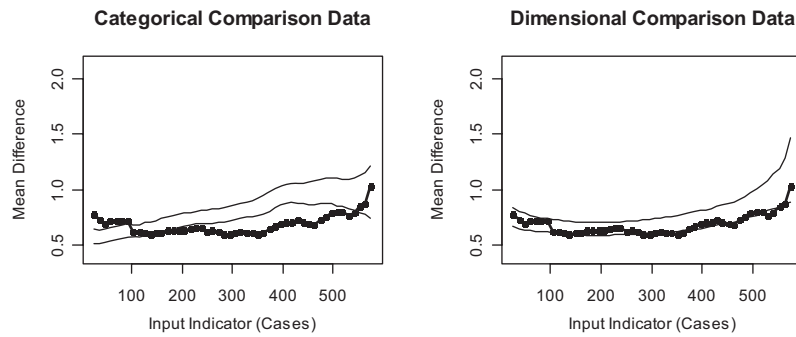
C: Prototypical MAXSLOPE Curves for Categorical (Left) and Dimensional (Right) Data

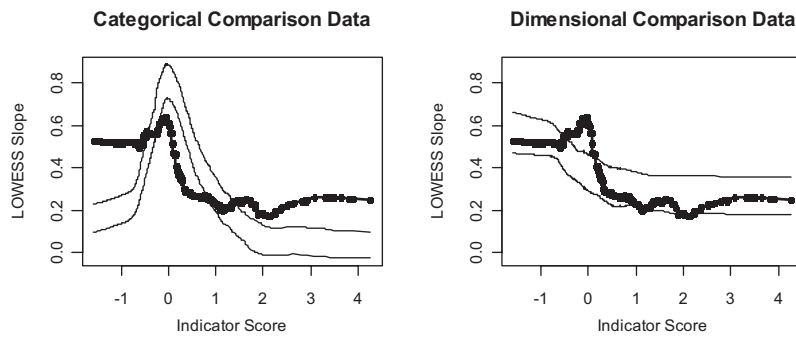**D: MAXSLOPE Analysis of Dimensional Data with Positively Skewed Variables**

Figure 1.   Results in Panels A and C are for empirical data only. Results in Panels B and D are plotted with data points and dark lines for empirical data, light lines for the range of values typical of comparison data. MAMBAC = mean above minus below a cut; MAXSLOPE = maximum slope.

preted as a peak to the left of the center of the curve, it is apparent that results are more similar to those for dimensional than for categorical comparison data. The CCFI value of .310, which is well below .50, is also supportive of dimensional structure.

In a recent study of taxometric consistency testing, Ruscio et al. (2010) found that combining the CCFI values from MAMBAC, MAXCOV, and L-Mode analyses effectively distinguished categorical and dimensional data. As noted above, however, neither MAXCOV nor L-Mode procedures can be performed when only two variables are available for taxometric analysis. How well would the combination of MAMBAC and MAXSLOPE analyses work in this case? MAMBAC has been studied fairly extensively (e.g., Ruscio, 2007; Ruscio et al., 2010), but never with only two variables. MAXSLOPE has not been studied using a comparatively rigorous design. In Study 1, we tested how well MAMBAC, MAXSLOPE, and their joint use differentiated between categorical and dimensional samples spanning a wide range of data conditions. The primary goal of this study was to determine whether taxometric analyses with only two variables are advisable, and if so what data conditions moderate the confidence one can place in the results. Two additional goals involved previously unstudied aspects of taxometric methodology.

The second goal addressed the issue of handling tied scores in taxometric analyses. For example, when one is sorting cases along the *x*-axis to locate cutting scores for a MAMBAC analysis, there may be cases with tied scores on one but not both variables. Their sorting along the *x*-axis will be arbitrary, and, in the event that this changes which cases are located on either side of a cutting score, this will affect the mean difference that is calculated and plotted in the MAMBAC curve. Ruscio et al. (2006) introduced a technique referred to as *internal replications* to minimize the obfuscating influence of tied scores by randomly resorting cases with tied scores, rerunning the analysis, and averaging the results across all such replications. In prior studies that included ordered categorical data and that therefore contained many tied scores, averaging over 10 internal replications has been the norm (e.g., Walters & Ruscio, 2009). Informal evaluations readily demonstrate the improvement in curve stability and smoothness that is obtained using 10 internal replications, but this issue has never been studied carefully. In Study 1, we examined the performance of the MAMBAC procedure with increasing numbers of internal replications.

The third goal addressed the question of whether to generate one population or more than one population of categorical comparison data when multiple taxometric procedures are performed. Ruscio et al. (2007) stressed the importance of generating categorical comparison data carefully and discussed several options for doing so. For example, one can provide a classification of cases into putative groups on the basis of prior research or theory, one can provide a base rate of taxon members on the basis of prior research or theory, or one can use an analysis of the empirical data at hand to estimate the base rate of taxon members or to classify cases into putative groups. When one provides only a taxon base rate, cases are then assigned to putative groups on the basis of their total scores on all available variables (Ruscio, 2009). Ruscio et al. (2007) argued that using an empirically estimated taxon base rate to assign cases to groups provides a conservative test of how well the taxometric analysis identifies categorical data and that researchers will often have a more valid classification to use and thereby be in a better position to identify categorical data. This

conservative practice has been standard in simulation studies of taxometric analysis so that the utility of the method is not overstated. However, there remains an important choice for how to implement this approach. One could use a single taxon base rate estimate to generate a single population of categorical comparison data and draw samples from this population for analysis using each taxometric procedure. In Study 1, we did this by obtaining taxon base rate estimates using MAMBAC and MAXSLOPE, and providing their averaged value as a single base rate estimate to generate one population of categorical comparison data. Alternatively, one could generate separate populations of comparison data using each taxometric procedure's taxon base rate estimate. We did this as well in Study 1, affording a test of which method works best.

In sum, Study 1 was designed to determine (a) whether to perform taxometric analyses with only two variables and how results vary across data conditions, (b) how much improvement is observed with increasing numbers of internal replications in MAMBAC analyses, and (c) whether to generate a single population of categorical comparison data for analysis using multiple taxometric procedures. A large-scale Monte Carlo study was conducted to address these issues. Study 2 applies the empirically based guidelines from Study 1 to an analysis of data on proactive and reactive aggression in children using Dodge and Coie's (1987) TC rating scale. It was hypothesized that, consistent with conventional wisdom (Hubbard et al., 2010), the latent structure of proactive and reactive aggression in children would be dimensional in nature.

Study 1

In this simulation study, MAMBAC and MAXSLOPE were performed across a wide range of data conditions to determine how effectively they could differentiate categorical and dimensional data.

Method

Design and data generation.   A total of 25,000 categorical and dimensional data sets (12,500 for each structure) were generated using a Monte Carlo design in which data parameters were independently randomly sampled from specified ranges. These data conditions were similar to those used in previous simulation studies of taxometric analysis (e.g., Ruscio & Kaczetow, 2009; Ruscio & Walters, 2009; Walters & Ruscio, 2010). For categorical data, which consisted of two latent classes referred to as the *taxon* (higher scoring group) and its *complement* (lower scoring group), random values were drawn for the following parameters of each target data set: sample size ($N = 300-1,000$), number of ordered categories ($C = 4, 5, 6, 7, 10,$ or 20), taxon base rate ($p = .10-.50$), indicator validity (standardized mean difference between classes of $d = 1.25-2.00$), within-group correlation ($r = .00-.30$), asymmetry ($g = .00-.30$), tail weight ($h = .00-.15$), and variance ratio ($VR = .25-4.00$; this is the ratio of variance in the taxon relative to variance in the complement).

Values of N, C, P, d, r, g, and h were drawn from uniform distributions (continuous for all but C, which was discrete) spanning the ranges listed above. The value of VR was determined by drawing a random value X from a uniform, continuous distribution

ranging from 1 to 4; with probability .50, $VR = X$, and with probability .50, $VR = 1/X$. The values of g and h were used to generate data from a g-and-h distribution (Hoaglin, 1985, p. 486). 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 (leptokurtosis) were studied. For the g-and-h populations used in this study, the smallest skew (γ_1) and kurtosis (γ_2) values were $\gamma_1 = 0$, $\gamma_2 = 0$ for $g = 0$ and $h = 0$, respectively (a normal distribution), and the largest values were $\gamma_1 = 2.60$, $\gamma_2 = 38.89$ for $g = .30$ and $h = .15$, respectively; other pairings of g and h correspond to γ_1 and γ_2 values within this range. This covers a wide range of symmetric and asymmetric distributions that should span those encountered in most empirical data (Micceri, 1989) and pose a substantial challenge to the correct identification of latent structure.

To generate a categorical data set, the iterative technique of Ruscio and Kaczetow (2008) was used to sample N cases from a g-and-h distribution with $\mu = 0$, $\sigma = 1$ in which the two variables 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 X (when $VR > 1$) or $1/X$ (when $VR < 1$). Separation between classes was achieved by adding a constant to scores for taxon members such that the standardized mean difference equaled d. Finally, each variable's distribution was cut into C-ordered categories using equally spaced thresholds spanning the range of observed values. To eliminate the possibility that outliers would result in categories with very low, or even 0, frequencies at or near the extremes, the original continuous g-and-h distributions were created using uniform quantiles spanning the range from .005 to .995.

For dimensional data, values of N through VR were drawn in the same way. However, because P, d, and r do not correspond to parameters of the dimensional (common factor) model, they were combined to yield an expected correlation using the following formula (Meehl & Yonce, 1994):

$$r_{xy} = \frac{P(1 - P)d^2 + r}{P(1 - P)d^2 + 1}$$

The iterative algorithm of Ruscio and Kaczetow (2008) was used to sample N cases from a g-and-h distribution with $\mu = 0$, $\sigma = 1$ in which the two variables correlated r_{xy} with one another, and data were cut into C-ordered categories. Because VR does not correspond to a parameter of the dimensional model, it was not used in the generation of dimensional data. Extensive checking showed that our data generation programs created categorical and dimensional target data sets with the intended characteristics.

When generating each of the 25,000 samples of target data for the present study, we initialized the random number seed to the sample number; this allows for the exact reproduction of these data for replication or further study; an R-file containing programs to reproduce our data is available on request.[1]

Data analysis. For each sample, we performed MAMBAC and MAXSLOPE analyses as described earlier. For MAMBAC, 50 evenly spaced cutting scores were located beginning and ending 25

cases from each extreme. A total of 25 internal replications were performed, and CCFI values were calculated at 4 points in the process—after 1, 5, 10, and 25 replications—to determine how much improvement was observed with additional replications. For MAXSLOPE, tied scores have no effect, and internal replications were not used.

For each target data set, we generated a single population of dimensional comparison data, along with either one or two populations of categorical comparison data. In the first case, a single population was generated using the mean taxon base rate estimate from the MAMBAC and MAXSLOPE analyses of the target data. This single population was used to draw samples of categorical comparison data for the MAMBAC and MAXSLOPE analyses. In the second condition, we generated separate populations of categorical comparison data using the taxon base rate estimate from each procedure. When performing MAMBAC, we drew samples of categorical comparison data from the population generated using this procedure's base rate estimate; the same was then done for MAXSLOPE.

To generate each population of categorical comparison data, we used the base-rate classification technique to assign cases to groups (Ruscio, 2009). Each base rate was used to classify cases into groups by rank-ordering cases according to their indicator total scores and applying a threshold corresponding to the proportion of the sample to be assigned to the taxon. To prevent the possibility that extreme base rate estimates would cause all or nearly all cases to be assigned to the same group, base rate estimates were constrained to the range of [.05, .95] for assigning cases to groups.

Once all populations of comparison data were generated, MAMBAC and MAXSLOPE analyses were performed for each of $B = 10$ samples drawn randomly from each population. Analyses were implemented in the same way as for the target data. Using the results for the sample of target data and its corresponding samples of comparison data, we calculated a CCFI value for each procedure. First, the fit value between all data points on the curves for the target data and categorical comparison data (Fit_{cat}) was calculated as the root-mean-square of the series' y values (i.e., mean differences for MAMBAC, slopes for MAXSLOPE). Next, the corresponding fit value was calculated between all data points on the curves for the target data and dimensional comparison data (Fit_{dim}). Finally, the CCFI was calculated as CCFI = $Fit_{dim}/(Fit_{dim} + Fit_{cat})$. As noted earlier, this means that the CCFI can range from 0 (strongest support for dimensional structure) to 1 (strongest support for categorical structure), with .50 an ambiguous value that indicates equivalent support for both structures. In addition to calculating the CCFI for each procedure, we calculated the mean CCFI value across procedures.

Results and Discussion

We examined accuracy levels to address each of the study goals outlined earlier. Table 1 summarizes the accuracy with which

---

[1] The 25,000 data files themselves would be too cumbersome to upload to a website or to transmit electronically as an e-mail attachment. Interested parties can run the comparatively small R programs we will provide to recreate some or all of these data files on demand.

Table 1
Percentage of Samples Correctly Classified, Study 1

Procedure and number of populations of categorical comparison data	Number of internal replications			
	1	5	10	25
Single threshold				
MAXSLOPE CCFI[a]				
One population	78.8	78.8	78.8	78.8
Two populations	74.9	74.9	74.9	74.9
MAMBAC CCFI				
One population	78.7	81.2	81.7	81.9
Two populations	78.8	81.3	81.5	81.8
Mean CCFI				
One population	81.1	81.9	82.1	82.1
Two populations	79.5	80.5	80.9	81.2
Consistent CCFIs[b]				
One population	86.1 (79.6)	86.5 (82.2)	86.6 (82.7)	86.5 (83.1)
Two populations	85.9 (74.7)	86.8 (76.3)	86.9 (76.5)	87.0 (76.6)
Dual thresholds				
MAXSLOPE CCFI[a]				
One population	82.8 (85.7)	82.8 (85.7)	82.8 (85.7)	82.8 (85.7)
Two populations	79.0 (83.3)	79.0 (83.3)	79.0 (83.3)	79.0 (83.3)
MAMBAC CCFI				
One population	84.9 (78.8)	86.0 (83.6)	86.2 (84.4)	86.3 (85.1)
Two populations	84.9 (77.6)	85.8 (83.2)	86.1 (83.9)	86.1 (84.5)
Mean CCFI				
One population	86.2 (82.3)	86.4 (84.1)	86.5 (84.7)	86.5 (85.0)
Two populations	85.2 (79.2)	86.0 (81.4)	86.0 (82.1)	86.0 (82.4)
Consistent CCFIs[b]				
One population	90.8 (62.0)	90.7 (66.8)	90.5 (67.8)	90.4 (68.4)
Two populations	90.9 (55.5)	90.7 (60.0)	90.8 (60.4)	90.7 (60.8)

Note. For the single threshold, the percentage correct was calculated for all 25,000 samples. For dual thresholds, the percentage correct was calculated for all samples for which the comparison curve fit index (CCFI) was outside the ambiguous range of .45 ≤ CCFI ≤ .55 (this percentage of nonambiguous CCFI values appears in parentheses). MAXSLOPE = maximum slope; MAMBAC = mean above minus below a cut.
[a] Because tied scores have no influence on MAXSLOPE analyses, internal replications were not used.   [b] The criterion of "consistent" required that both procedures' CCFI values were in the same direction (i.e., both > .50 or both < .50 for single threshold, both > .55 or both < .45 for dual thresholds).

CCFI values distinguished between categorical and dimensional structure across all 25,000 samples. Results are shown for MAMBAC CCFIs, MAXSLOPE CCFIs, mean CCFIs, and consistent CCFIs—where the criterion for consistency was that both procedures' CCFIs were in the same direction (i.e., both values > .50 or both values < .50)—using both a single population of categorical comparison data as well as two populations and across varying numbers of internal replications. Accuracy was scored using two criteria for the CCFI. First, the *single threshold* refers to a cutoff at CCFI = .50. Scores below .50 are evidence of dimensional structure, and scores above .50 are evidence of categorical structure. Second, the *dual thresholds* refer to cutoffs at CCFI < .45 (evidence of dimensional structure) and CCFI > .55 (evidence of categorical structure), with scores in the intermediate range from .45 to .55 set aside as ambiguous evidence.[2] When some results are set aside as ambiguous (i.e., when requiring consistency across procedures or when using dual thresholds), it is important to interpret an accuracy rate in the context of the rate of ambiguous evidence. For example, a technique that achieves 85% accuracy for 80% of all target data (with the other 20% set aside as ambiguous)

is better than a technique that achieves 85% accuracy for only 75% of all target data.

An inspection of the results shown in Table 1 reveals several noteworthy trends. First, using a single population of categorical comparison data attained greater accuracy than did using two populations. The difference was substantial for MAXSLOPE (78.8% accuracy for one population vs. 74.9% accuracy for two populations), modest for the mean CCFI (82.1% accuracy for one population vs. 81.2% accuracy for two populations), and negligible for MAMBAC (81.9% for one populations vs. 81.8% for two populations); accuracy was approximately the same for consistent

[2] Ruscio et al. (2010) discussed the selection and application of relatively narrow dual thresholds (those used here) versus relatively broad dual thresholds (e.g., CCFI < .40 and CCFI > .60). We used only the relatively narrow dual thresholds because it simplified the presentation of results, and the choice of narrow versus broad dual thresholds did not make any difference for our interpretations; all of our conclusions remain the same for each.

CCFIs with one population (86.5%) or two (87.0%), but results were ambiguous less often with one population (16.9%) than with two populations (23.4%). The figures cited above were observed using a single threshold, and the same pattern emerged using dual thresholds (see Table 1). Not only was accuracy higher for one population than for two populations, but this was achieved with fewer ambiguous results. In what follows, only results obtained using one population of categorical comparison data are considered further.

Second, accuracy increased with the number of internal replications, but there were rapidly diminishing returns. For example, MAMBAC accuracy increased from 78.7% to 81.2% with an increase from 1 to 5 replications, but only from 81.7% to 81.9% with an increase from 10 to 25 replications. The gains were more modest for the mean CCFI values, perhaps in part because there was no corresponding gain for MAXSLOPE, which is unaffected by tied scores and therefore unassisted by replications. For consistent CCFIs, accuracy increased little, but results were ambiguous less often with increasing numbers of replications. Once again, these figures were observed using a single threshold, and the same pattern emerged using dual thresholds (see Table 1). In what follows, only results obtained using 25 replications are considered further.

Third, greater accuracy was achieved with MAMBAC than with MAXSLOPE. Using a single threshold, MAMBAC was 81.9% accurate, whereas MAXSLOPE was 78.8% accurate. Using dual thresholds, MAMBAC was 86.3% accurate, and MAXSLOPE was 82.8% accurate. Even taking into account that MAMBAC yielded ambiguous results slightly more often (14.9%) than MAXSLOPE (14.3%), the difference in accuracy rates is not trivial.

Fourth, the mean CCFI value outperformed that for MAXSLOPE but was comparable to that for MAMBAC. Table 1 shows the pertinent results using both single and dual thresholds. To compare accuracy across procedures independent of threshold, the probability-based measure of effect size A was used (Ruscio, 2008). This is equivalent to the area under a receiver operating characteristic curve, and in this context it can be interpreted as the probability that a randomly chosen categorical sample would yield a larger CCFI than would a randomly selected dimensional sample. For MAMBAC, MAXSLOPE, and their mean CCFI values, A = .894, .860, and .896, respectively. The 95% confidence intervals for A overlap for MAMBAC CCFI (.890, .898) and mean CCFI (.892, .900); neither of these intervals overlaps with that for MAXSLOPE CCFI values (.855, .865).

Fifth, requiring consistent CCFI values achieved the greatest accuracy of all, but at the cost of a substantial increase in the frequency of ambiguous results. For example, using dual thresholds, the mean CCFI value attained 86.5% accuracy for 85.0% of all samples. Requiring consistent CCFI values increased accuracy to 90.4%, but only 68.4% of all samples met this more stringent criterion. This pattern of results is similar to what Ruscio et al. (2010) observed when comparing the mean CCFIs to consistent CCFIs for MAMBAC, MAXEIG, and L-Mode. Ultimately, the choice of a criterion (mean CCFIs vs. consistent CCFIs) involves a value judgment that a researcher must make: Is the gain in accuracy worth the increased risk of ambiguous results? Given the rather modest difference in accuracy, we used the mean CCFI values in subsequent analyses.

At this point, the examination of results only partially addresses the primary goal of this study. It does appear justifiable to perform taxometric analyses with only two variables, but it is not yet clear how data conditions affect the accuracy of results. Figure 2 presents the results for mean CCFI values across levels of each data condition. Dual thresholds were used to identify correct, incorrect, and ambiguous results. Within each graph, solid data points represent accuracy among results that were not set aside as ambiguous, or the percentage correct divided by the sum of the percentage correct and the percentage incorrect. As shown in Table 1, the overall accuracy level was 86.5% after setting aside 15.0% of samples yielding ambiguous results. Accuracy remained within the range of 80%–90% for most data conditions. For categorical data, accuracy was influenced most heavily by the degree of separation between groups (indicator validity), within-group correlations, and the number of ordered categories. In each case, the direction of the effect is what one would expect: Accuracy was greater for more valid variables, variables with lower correlations within groups, and variables that ranged across larger numbers of ordered categories. For dimensional data, sample size and the number of ordered categories were the most important factors, again in predictable ways.

In some instances, data conditions had different influences for categorical and dimensional data. For example, whereas larger tail weight decreased accuracy for dimensional data, the opposite was true for categorical data. This underscores the importance of examining the influence of data conditions across all samples because in an actual taxometric study one would not know whether the data are categorical or dimensional. Figure 3 presents the results collapsed across structure for all data conditions that apply to both structures: sample size, number of ordered categories, the correlation between the two variables, asymmetry, and tail weight. There were modest increases in accuracy with larger samples and more substantial increases in accuracy with more ordered categories. The other three factors appear to make little difference. Figure 4 shows that accuracy increased to a modest extent with larger samples at each number of ordered categories and that the accuracy increased substantially with more ordered categories across all sample sizes. Thus, it appears that the number of ordered categories is the most important factor that an investigator can consider when planning a study or interpreting the results.

Study 2

Despite their conventional description as types or subtypes of aggression, researchers in the field of developmental psychology treat proactive and reactive childhood aggression as dimensional rather than categorical. In summarizing several decades of research on the topic, Hubbard et al. (2010) concluded that proactive and reactive aggression "are most accurately conceptualized as continuous dimensions that exist to varying degrees in each child, rather than as categories into which children are exclusively placed" (p. 96). Although it is commonly assumed that proactive and reactive aggression are dimensions, this assumption has never been formally tested. In the present study, taxometric analyses were performed using a sample of proactive and reactive scores from the most commonly utilized measure of childhood proactive and reactive aggression, that is, Dodge and Coie's (1987) TC rating scale.
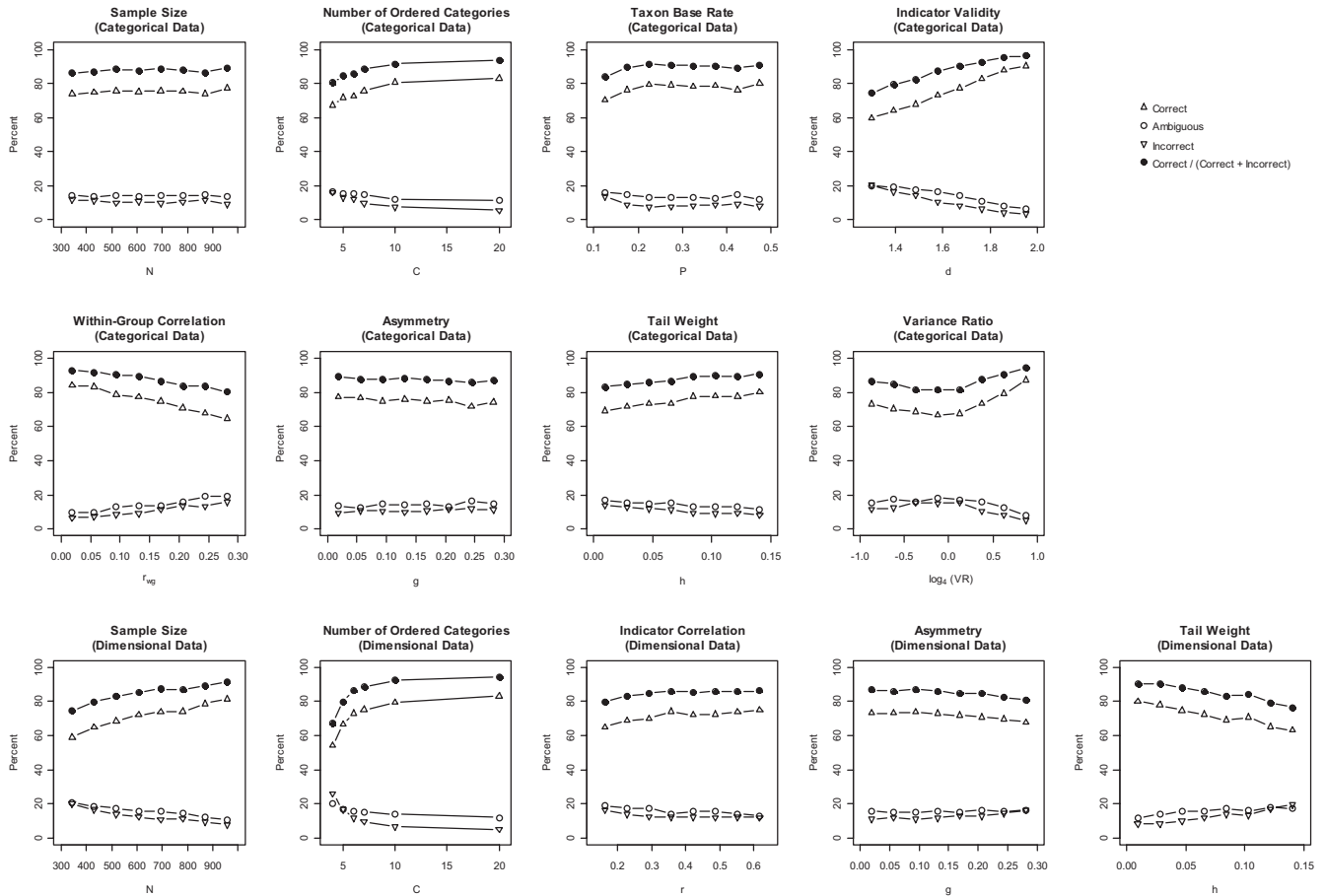
Figure 2. Accuracy of mean comparison curve fit index (CCFI) values across data conditions. Results for variance ratios (*VR*s) are presented using the base-4 logarithm of the variance ratio to create a symmetric scale ranging from –1 (corresponding to *VR* = 0.25), to 0 (corresponding to *VR* = 1.00), to +1 (corresponding to *VR* = 4.00).

Method

Participants for this study were 1,005 Israeli primary school (Grades 3–6) students previously rated by their teachers on the Dodge and Coie (1987) TC. This was a mixed-gender (789 boys, 216 girls) and culturally diverse (500 Jews, 505 Arabs) sample of Israeli schoolchildren originally selected for a national project on aggression and later used in a taxometric study on childhood aggression (Walters, Ronen, et al., 2010) that included more than the TC variables examined here. Each child and at least one parent gave their informed consent for the child to participate in the project prior to the start of the original study.

The TC is a six-item rating scale completed by a child's teacher and designed to assess the level of proactive and reactive aggression engaged in by the child. Research on the TC has produced encouraging results with respect to the measure's reliability and convergent validity but less than stellar results with respect to its discriminant validity (Hubbard et al., 2010). The TC items are normally rated on a 5-point scale, but in the current study a 3-point scale was used. Hence, the proactive and reactive scores each ranged from 3 to 9.

The two TC scales were subjected to MAMBAC and MAXSLOPE analysis using Ruscio's (2010) taxometric programs for the R computing environment. Following the results of Study 1, a taxon base rate estimate was calculated using each procedure, and these were averaged to create a single population of categorical comparison data for both procedures. MAMBAC was calculated with 25 internal replications, and CCFI values were calculated for MAMBAC and MAXSLOPE.

Results and Discussion

To assess the appropriateness of the two TC indicators for taxometric analysis, the sample was divided into putative taxon and complement groups using a total TC cut score (11) that produced a base rate (21.6%) that came closest to the average base rate of serious aggression (18.7%) in a multisample study by Broidy et al. (2003). Cases with scores of 11 or higher were assigned to the putative taxon, and scores below 11 were assigned to its complement class. The full sample interindicator correlation ($r = .58$) exceeded Meehl's (1995) recommended threshold for sufficient indicator covariation ($r \geq .30$), and the within-group
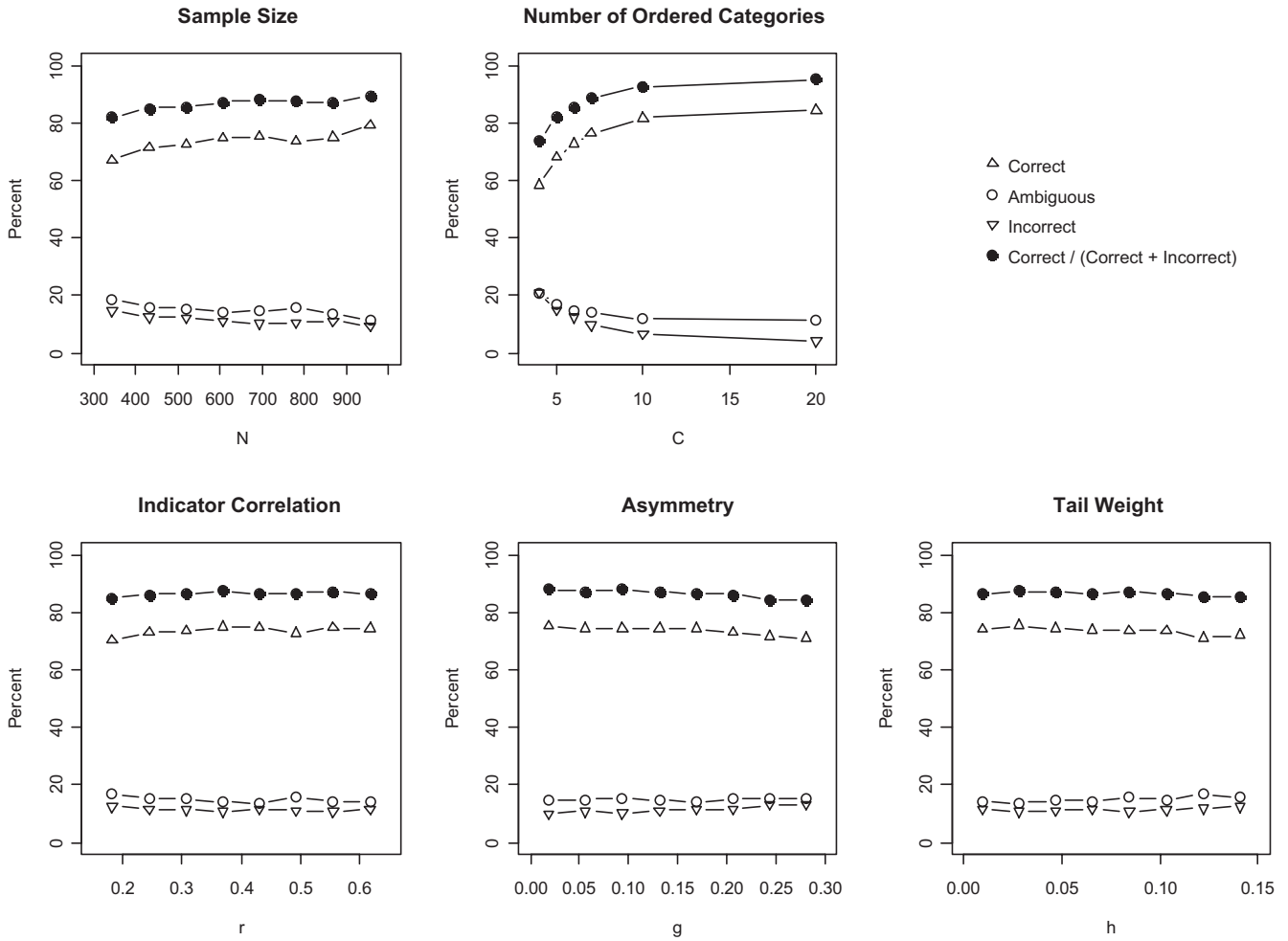
Figure 3. Accuracy of mean comparison curve fit index (CCFI) values across data conditions, collapsing across categorical and dimensional target data sets.

correlations (taxon $r = -.06$, complement $r = .19$) fell below Meehl's (1995) recommended threshold for tolerable nuisance covariance ($r < .30$). The ability of the two indicators to distinguish between the putative taxon and complement groups exceeded Meehl's (1995) recommended threshold for meaningful group differences (Cohen's $d \geq 1.25$): for proactive aggression, $d = 2.98$; for reactive aggression, $d = 2.26$.

The MAMBAC and MAXSLOPE analyses performed with the two TC indicators yielded taxon base rate estimates of .264 and .447, respectively. The mean of these two values (.356) was used to generate a single population of categorical comparison data for use with the MAMBAC and MAXSLOPE procedures. Figure 5 presents the MAMBAC (top panel) and MAXSLOPE (bottom panel) curves for the empirical data along with those for comparison data. The MAXSLOPE curves are rather jagged owing to the limited number of distinct scores on each variable, hence the small number of slopes that can be calculated. Nonetheless, for both procedures it is apparent that the curves for the empirical data were more similar to those for the dimensional than categorical comparison data. The CCFI values (.280 for MAMBAC, .294 for MAXSLOPE, mean CCFI = .287) corroborate the superior fit of

the dimensional model; these values are well outside of an intermediate range of ambiguous CCFI values near .50.

The results of this study provide preliminary support for the hypothesis that the construct underlying these indicators is dimensional rather than categorical. As such, the domain of proactive and reactive aggression in children appears to be organized quantitatively along one or more dimensions rather than qualitatively into distinct proactive and reactive aggression categories, and individual differences are a matter of degree rather than a difference in kind. The moderately high to high correlations between proactive and reactive aggression reported in the literature (see Hubbard et al., 2010) may have prompted researchers to assume that the constructs are overlapping dimensions, but the present results provide more direct empirical support for this structural assumption. The ability to perform a taxometric analysis with only two variables enabled this test of competing structural models.

General Discussion

The primary goal of this research was to determine whether taxometric analyses can be recommended when only two variables
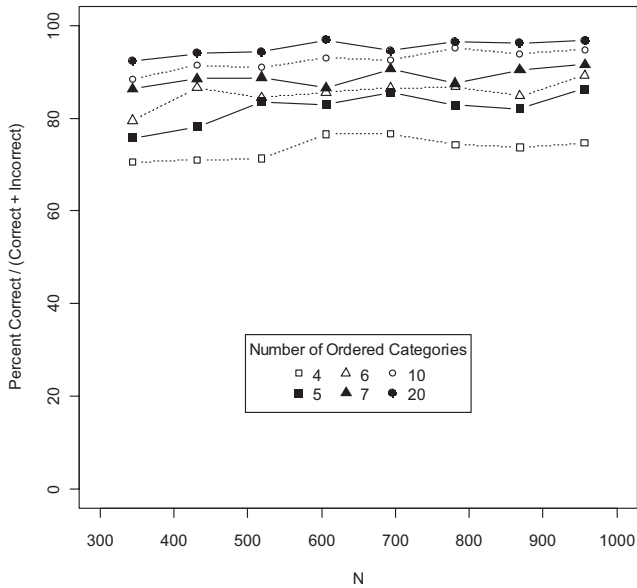
Figure 4. Accuracy of mean comparison curve fit index (CCFI) values across sample size and number of ordered categories.

are available and, if so, under what conditions. For all 25,000 samples in Study 1, the mean CCFI value for MAMBAC and MAXSLOPE achieved an accuracy level of 82.1%. As was the case in Ruscio et al.'s (2010) study of consistency testing, MAMBAC yielded the greatest accuracy level of any single procedure studied—in the present Study 1, MAMBAC was 81.9% accurate, and MAXSLOPE was 78.8% accurate—but the mean CCFI was a bit more accurate still. If one were willing to set aside as ambiguous the 15.0% of mean CCFI values in the intermediate range from .45 to .55, accuracy would rise to 86.5%. Setting aside an additional 15.6% of mean CCFI values in the intermediate range from .40 to .60 would increase accuracy to 90.1%. Requiring that CCFI values lie in the same direction boosts accuracy further (to 86.5% for a single threshold, 90.4% for narrow dual thresholds of .45 and .55, and 92.9% for broad dual thresholds of .40 and .60), but at the cost of more ambiguous results (16.9%, 31.6%, and 46.3%, respectively). Accuracy rates differed substantially across many data conditions. Of the factors that an investigator can examine without knowledge of the target construct's latent structure, sample size and the number of ordered categories were the most influential. Accuracy rose with an increase in either of these factors, but the effect was more dramatic for the number of ordered categories. This suggests that when only two variables are available for analysis, it is critical to construct and select them such that they will span many distinct values to approximate continuous distributions as well as possible.

Given these results, we believe that taxometric analysis with only two variables can be recommended, but it is more important than ever to establish that one's data are capable of yielding informative results and to take proper precautions when implementing the taxometric analyses. Ruscio et al. (2006) discussed a number of issues to consider, and Meehl's (1995) rules of thumb for data acceptable for taxometric analysis (e.g., $N > 300$, $d > 1.25$, $r_{wg} < .30$) received strong support in Ruscio et al.'s (2010)

study of consistency testing. The current study provides the first empirical evaluation of two specific implementation decisions, how to handle tied scores and how to generate populations of comparison data. For taxometric procedures affected by the presence of tied scores (e.g., MAMBAC), using at least 10 internal replications seems warranted, though there appears to be relatively little gain beyond that point. This coincides with the default value of 10 replications in Ruscio's (2010) taxometric programs. With advances in computing power, this default value may be raised in the future. In the meantime, investigators might consider using 25 or more replications if they can complete their analyses in a timely fashion given the unique processing demands of their research. When generating comparison data for parallel analysis and the calculation of the CCFI, it appears to be preferable to create a single population of categorical comparison data for all taxometric procedures rather than creating a new population using the results of each procedure. If one wishes to use the base-rate classification procedure to assign cases to putative groups, as was done in Study 1, a single taxon base rate estimate must be provided. An initial series of analyses can be performed without using any comparison data just to estimate the taxon base rate using each procedure. The mean of these estimates can then be used to generate categorical comparison data when each procedure is rerun to calculate a CCFI value. If the same taxon base rate and random number seed is provided for each procedure, Ruscio's (2010) taxometric programs will recreate the same populations of categorical (and dimensional) comparison data. By default, the seed is set to 1 for each procedure and therefore identical comparison data will be generated as long as the supplied base rate is the same.

Even though one can obtain informative taxometric results with only two variables, the conditions required to draw conclusions with much confidence are more restrictive than when three or more variables are available. Walters and Ruscio (2009) included MAMBAC and MAXCOV analyses in a simulation study spanning data conditions very similar to those found in Study 1; as noted earlier, MAXCOV uses at least three variables but is otherwise very similar to MAXSLOPE. Under the subset of data conditions that match those of the present Study 1—specifically, when $C \geq 4$—the mean CCFI from MAMBAC and MAXCOV achieved an accuracy of 81.5% with three variables, which is very close to the 82.1% accuracy rate in the present Study 1. However, accuracy increased to 93.0% with four variables, 94.7% with five variables, and 97.0% with six to eight variables. Thus, when more than two variables are available, it is not as important that one work with large samples of data that span a large number of ordered categories. It would still be wise to do so, when possible, but informative results can be obtained under a wider range of data conditions when more variables are available.

We hasten to add, however, that increasing the number of variables submitted to taxometric analysis must be done with care. There remain other data requirements. One important consideration is that each additional variable should be conceptually and empirically nonredundant with those already included in an analysis (Meehl, 1995). That is one reason why simply analyzing all six TC items assessing proactive and reactive aggression would be unwarranted. There are three items assessing each of two conceptually distinct scales, and these item subsets should be aggregated to form two scale scores prior to analysis. Another concern is that, even if all available items were conceptually and empirically
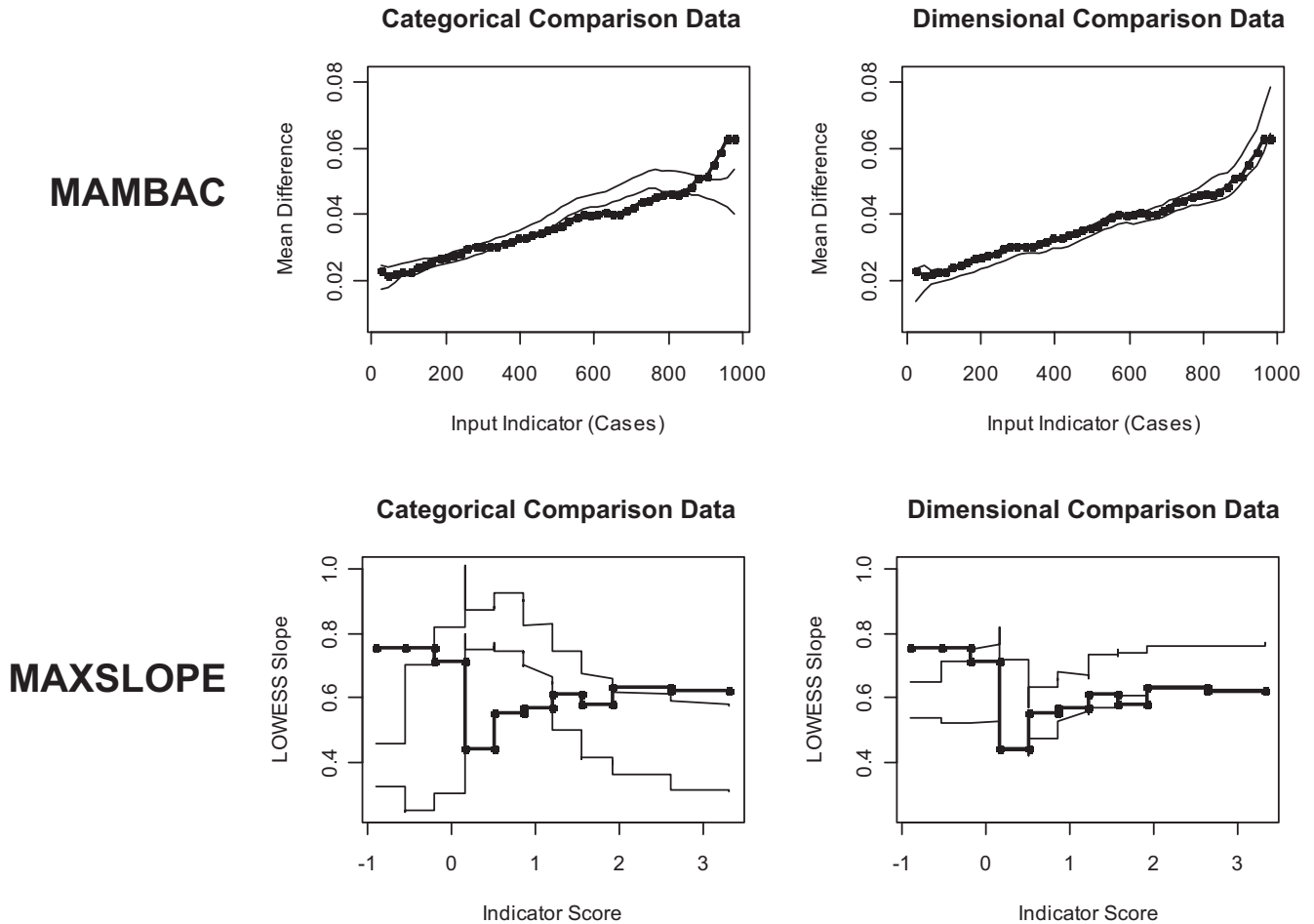
MAMBAC

Categorical Comparison Data

Dimensional Comparison Data

MAXSLOPE

Categorical Comparison Data

Dimensional Comparison Data

Figure 5. Taxometric curves for analyses of proactive and reactive aggression data. MAMBAC = mean above minus below a cut; MAXSLOPE = maximum slope.

nonredundant, a larger number of items varied along a more constrained response scale could yield less informative results than could a smaller number of items varied along a broader response scale. As the results of Study 1 show, the number of ordered categories exerts a strong influence on the accuracy rate of taxometric analyses. For example, the six TC items vary along 3-point response scales. By summing item subsets, two variables that vary along 7-point scales are created. Among the samples analyzed by Walters and Ruscio (2009) with six variables that varied along 3-point scales, the accuracy of the mean CCFI for MAMBAC and MAXCOV was 83.0%. In the present Study 1 with two variables that varied along 7-point scales, the accuracy of the mean CCFI for MAMBAC and MAXSLOPE was 84.2%. This is a relatively small difference, but it demonstrates that the number of ordered categories can be as (or more) important a consideration as the number of variables in a taxometric analysis.

In Study 2, we performed taxometric analyses on a large sample of data collected using Dodge and Coie's (1987) TC rating scale for proactive and reactive aggression. These analyses illustrate the possibility of obtaining informative taxometric results when only two variables are available. The graphs for MAMBAC and MAXSLOPE, when accompanied by those for parallel analyses of

comparison data, clearly favor dimensional latent structure over categorical latent structure. This was confirmed by the very low CCFI values, well outside an ambiguous range of intermediate values. We have argued that, given the nature of this target construct and the measures available to study it, only two variables will be available for analysis. Is this a fairly unique situation, in which case the present research addresses a question that will seldom arise in practice?

We believe that there are more than a few situations in which only two variables will be available due to the nature of the target construct and/or the measures that can be used to assess it. For example, to determine whether introversion–extraversion (I–E) forms a categorical system, as suggested by Jung (1921/1971) and consistent with the common practice of interpreting responses to the Myers–Briggs Type Indicator in a categorical fashion, or forms a dimensional model as espoused by Eysenck (1971), one could perform a taxometric analysis of two variables (I and E scores) from the same measure. There are a number of different measures of I–E, but still only two conceptually distinct variables, so multiple measures of I and E should be aggregated into two variables for analysis. Another example involves the study of the structure of psychopathy, specifically research using the Psychopathy Check-

list—Revised (e.g., Edens, Marcus, Lilienfeld, & Poythress, 2006) or the Psychopathy Checklist: Screening Version (e.g., Walters et al., 2007). Though standard practice has been to use four facet scores as variables in taxometric analyses, one might choose instead to analyze the two-factor scores (Selfish, Callous, and Remorseless Use of Others vs. Chronically Unstable and Antisocial Lifestyle) because of their greater reliability (more items per factor than per facet) and because of the unresolved controversy surrounding the relative merits of three- versus four-factor models of psychopathy as assessed by these measures (Hare & Neumann, 2010; Skeem & Cooke, 2010).

A final class of examples is considerably more extensive: The *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–IV–TR*; 4th ed., text rev.; American Psychiatric Association, 2000) provides multiple criteria for many disorders, but only two criteria for others. The criteria for body dysmorphic disorder are (a) preoccupation with an imagined defect in personal appearance and (b) significant clinical distress and/or impaired functioning as a result of this preoccupation. The criteria for delusional disorder are (a) nonbizarre delusion and (b) absence of severe impairment as in schizophrenia. The criteria for several sexual dysfunction disorders pair (a) a specific behavior (decreased sexual desire or aversion to genital sexual contact) with (b) distress or interpersonal difficulties. The criteria for sleep disorders are similar to those for the sexual dysfunction disorders mentioned above. Specifically, several of these disorders (e.g., nightmare sleep disorder, circadian rhythm sleep disorder, primary insomnia) have two primary criteria: (a) a specific sleep-related behavior paired with (b) distress/impairment. This is by no means a comprehensive list of psychological constructs for which only two variables might be available for analysis, but it demonstrates that this is a situation that will be encountered in many important areas of research. The present findings provide the first empirical guidance for implementing the taxometric method in this situation.

References

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

Beck, A. T., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., . . . Vitaro, F. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology, 39,* 222–245.

Brown, K., Atkins, M. S., Osborne, M. L., & Milnamow, M. (1996). A revised teacher rating scale for reactive and proactive aggression. *Journal of Abnormal Child Psychology, 24,* 473–480.

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology, 63,* 452–459.

Cleveland, W. S. (1979). Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74,* 829–836.

Dodge, K. A., & Coie, J. D. (1987). Social information processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53,* 1146–1158.

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.

Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). New York, NY: Oxford University Press.

Eysenck, H. J. (1971). *Readings in extraversion–introversion.* New York, NY: Wiley.

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

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

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

Hare, R. D., & Neumann, C. S. (2010). The role of antisociality in the psychopathy construct: Comment on Skeem and Cooke. *Psychological Assessment, 22,* 446–454.

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, NY: Wiley.

Hubbard, J. A., McAuliffe, M. D., Morrow, M. T., & Romano, L. J. (2010). Reactive and proactive aggression in childhood and adolescence: Precursors, outcomes, processes, experiences, and measurement. *Journal of Personality, 78,* 95–118.

Jung, C. G. (1971). *Collected works of C. G. Jung: Vol. 6. Psychological types* (G. Adler & R. F. C. Hull, Eds. & Trans.). Princeton, NJ: Princeton University Press. (Original work published 1921)

Kempes, M., Matthys, W., Maassen, G., van Goozen, S., & van Engeland, H. (2006). A parent questionnaire for distinguishing between reactive and proactive aggression in children. *European Child and Adolescent Psychiatry, 15,* 38–45.

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

McCutcheon, A. L. (1987). *Latent class analysis.* Newbury Park, CA: Sage.

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

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

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

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

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

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods, 13,* 19–30.

Ruscio, J. (2009). Assigning cases to groups using taxometric results: An empirical comparison of classification techniques. *Assessment, 16,* 55–70.

Ruscio, J. (2010). *Taxometric programs for the R computing environment: User's manual.* Retrieved from http://www.tcnj.edu/~ruscio/TaxProgManual%202010-07-26.pdf

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., & Kaczetow, W. (2009). Differentiating categories and dimensions: Evaluating the robustness of taxometric analysis. *Multivariate Behavioral Research, 44,* 259–280.

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., & Keane, T. M. (2004). Using taxometric analysis to distinguish a small latent taxon from a latent dimension with positively skewed indicators: The case of involuntary defeat syndrome. *Journal of Abnormal Psychology, 113,* 145–154.

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

Ruscio, J., & Walters, G. D. (2009). Using comparison data to differentiate categorical and dimensional data by examining factor score distributions: Resolving the mode problem. *Psychological Assessment, 21,* 578–594.

Ruscio, J., Walters, G. D., Marcus, D. K., & Kaczetow, W. (2010). Comparing the relative fit of categorical and dimensional latent variable models using consistency tests. *Psychological Assessment, 22,* 5–21.

Skeem, J. L., & Cooke, D. J. (2010). Is criminal behavior a central component of psychopathy? Conceptual directions for resolving the debate. *Psychological Assessment, 22,* 433–445.

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

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., McGrath, R. E., & Knight, R. A. (2010). Taxometrics, polytomous constructs, and the comparison curve fit index: A Monte Carlo study. *Psychological Assessment, 22,* 149–156.

Walters, G. D., Ronen, T., & Rosenbaum, M. (2010). The latent structure of childhood aggression: A taxometric analysis of self-reported and teacher-rated aggression in Israeli schoolchildren. *Psychological Assessment, 22,* 628–637.

Walters, G. D., & Ruscio, J. (2009). To sum or not to sum: Taxometric analysis with ordered categorical assessment items. *Psychological Assessment, 21,* 99–111.

Walters, G. D., & Ruscio, J. (2010). Where do we draw the line: Assigning cases to subsamples for MAMBAC, MAXCOV, and MAXEIG taxometric analyses. *Assessment, 17,* 321–333.