



Published in final edited form as:

Struct Equ Modeling. 2010 October 1; 17(4): 605–628. doi:10.1080/10705511.2010.510050.

Latent Class Detection and Class Assignment: A Comparison of the MAXEIG Taxometric Procedure and Factor Mixture Modeling Approaches

Gitta Lubke[#] and Stephen Tueller[#]

University of Notre Dame

[#] These authors contributed equally to this work.

Abstract

Taxometric procedures such as MAXEIG and factor mixture modeling (FMM) are used in latent class clustering, but they have very different sets of strengths and weaknesses. Taxometric procedures, popular in psychiatric and psychopathology applications, do not rely on distributional assumptions. Their sole purpose is to detect the presence of latent classes. The procedures capitalize on the assumption that, due to mean differences between two classes, item covariances within class are smaller than item covariances between the classes. FMM goes beyond class detection and permits the specification of hypothesis-based within-class covariance structures ranging from local independence to multidimensional within-class factor models. In principle, FMM permits the comparison of alternative models using likelihood-based indexes. These advantages come at the price of distributional assumptions. In addition, models are often highly parameterized and susceptible to misspecifications of the within-class covariance structure.

Following an illustration with an empirical data set of binary depression items, the MAXEIG procedure and FMM are compared in a simulation study focusing on class detection and the assignment of subjects to the latent classes. FMM generally outperformed MAXEIG in terms of class detection and class assignment. Substantially different class sizes negatively impacted the performance of both approaches, whereas low class separation was much more problematic for MAXEIG than for the FMM.

One of the classic and long-standing debates in psychology revolves around the question of whether individual differences should be conceived of in terms of typologies or in terms of continuous traits (e.g., Kendell, 1991; Meehl, 1992; Rutter & Shaffer, 1980; Wilson, 1993). Recently, this question has regained momentum in psychiatry (Pickles & Angold, 2003). For example, Attention Deficit/Hyperactivity Disorder (ADHD) has been defined in terms of distinct subtypes in the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2000), but it has also been argued that it is more adequate to assume gradual differences with respect to the severity of the disorder (Hudziak, Achenbach, Althoff, & Pine, 2007; Hudziak et al., 1998; Lubke, Hudziak, Derks, van Bijsterveldt, & Boomsma, 2009; Lubke et al., 2007; Rohde et al., 2001). The distinction between types and traits has practical relevance not only in diagnosis, prevention, or intervention, but also in genetic research where it is important to consider the usefulness of searching for subtype-specific genes (Lasky-Su et al., 2008).

It is often unknown whether or how many subtypes exist, and several methodologies have been developed to determine whether individual differences are more adequately described in terms of traits or types. Independent of the method, differences between types can typically be detected only when the mean differences between the types are sufficiently large in a given set of data. Types can be formalized as latent classes, and mean differences between the classes can be measured for instance with the multivariate Mahalanobis distance. Methods can differ with respect to their sensitivity of detecting the latent clustering. When classes are detected, an ensuing challenge is how to accurately assign subjects to their true class. Taxometric procedures and factor mixture models (FMMs) are two widely used approaches designed to detect latent classes and assign subjects to classes.

This study compares FMM and taxometric procedures with respect to the detection of latent classes and with respect to class assignment. Previous comparisons are limited. Cleland, Rothschild, and Haslam (2000) found that using taxometric methods and fitting unconstrained finite mixture models performed roughly the same in their respective abilities to detect whether data were produced from a one- or two-class structure. The FMM is a special case of the normal finite mixture model. In an applied study, Lenzenweger, McLachlan, and Rubin (2007) found that taxometrics and an unconstrained finite mixture model produced consistent results, and Marcus, Ruscio, Lilienfeld, and Hughes (2008) reported taxometrics and latent class analysis (LCA) yield consistent results. LCA is a submodel of the more general FMM that resembles most closely the main idea behind taxometrics. None of the three studies have evaluated class assignment.

Taxometric procedures were first developed by Meehl and colleagues (Meehl & Yonce, 1996; Waller & Meehl, 1998). Taxometrics aim at discriminating between two latent classes, the taxon and its complement. Taxometric procedures have been used frequently in psychiatric, psychopathology, and personality research (see Ruscio, 2008; Ruscio, Haslam, & Ruscio, 2006, pp. 266–267, for a comprehensive list of applications of taxometric procedures). FMMs fall into the broad category of latent variable models that are typically fitted using maximum likelihood. The FMM combines LCA and confirmatory factor analysis, and can in principle be used to discriminate between any number of user-specified latent classes. Variations of this model have been proposed by several different researchers (e.g., Arminger, Stein, & Wittenberg, 1999; Dolan & Van der Maas, 1998; Heinen, 1996; Jedidi, Jagpal, & DeSarbo, 1997; B. Muthén & Shedden, 1999; Vermunt & Magidson, 2003; Yung, 1997). An attractive feature of FMM is the possibility to specify and compare different within-class models. FMM has been applied in an increasing number of substantive areas such as developmental psychology (e.g., Nylund, Bellmore, Nishina, & Graham, 2007), psychopathology and addiction (e.g., Greenbaum & Dedrick, 2007; Neale, Aggen, Maes, Kubarych, & Schmitt, 2006), criminology (e.g., Nagin & Land, 1993), and psychiatric applications (e.g., Lubke et al., 2009; Lubke et al., 2007).

Taxometric procedures typically address only one of the questions that can be investigated with FMMs, namely whether or not subjects in a given data set are best described in terms of two clusters or in terms of a single homogeneous population. Because the FMM is a general model permitting the specification of a large number of alternative submodels, the range of applications of FMMs is much wider. This study focuses mainly on data conditions in the area of application of taxometric procedures, and investigates the sensitivity of the two methods in detecting two clusters when present in the data. The comparison is carried out with simulated data and includes a range of conditions known to be unproblematic but also some known to be problematic for either or both approaches. In addition, we evaluate classification accuracy, which is especially important given the increasing number of empirical studies reporting post-hoc comparisons between latent classes carried out after assigning subjects to classes. Class assignment is usually based on some probability measure

of belonging to a class, and, consequently, contains uncertainty (Tueller & Lubke, 2010).¹ Because the validity of post-hoc comparisons between classes will depend on the size of the assignment error, we compute error rates for both taxometrics and FMM.

The next section introduces the two methodologies. This is followed by an empirical example using data that are common in psychopathology research. The empirical data are analyzed by applying multiple taxometric procedures and by comparing the fit of a number of alternative FMMs. The empirical analyses also illustrate the flexibility of FMMs to conduct analyses that go beyond determining whether two classes are more appropriate than a single class. The remainder of the article describes the simulation study comparing the taxometric procedure MAXEIG and FMM with respect to class detection and class assignment.

TAXOMETRIC PROCEDURES AND THE FMM

This section provides a conceptual description of the essential features of taxometric procedures and the FMM. Much more detail can be found in Meehl (1973, 1995), Waller and Meehl (1998), and Ruscio et al. (2006) for taxometric procedures, and B. Muthén and Shedden (1999), Yung (1997), Jedidi et al. (1997), Dolan and Van der Maas (1998), and Lubke and Muthén (2005) for the FMM. In addition, the measures of class separation used in the taxometric and FMM literature are described and compared.

Taxometric Procedures

Taxometric procedures were developed by Meehl and colleagues (Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998), and have received recent attention by Ruscio and colleagues (Ruscio, 2007; Ruscio et al., 2006; Ruscio & Marcus, 2007), who have provided user-friendly functions to carry out the procedures in R (R Development Core Team, 2008). This study uses the MAXimum EIGenvalue procedure (MAXEIG; Waller & Meehl, 1998) and the Comparison Curve Fit Index (CCFI; Ruscio & Marcus, 2007; Ruscio, Ruscio, & Meron, 2007) for detecting types, and uses the base rate classification technique (Ruscio, 2007, 2009; Ruscio et al., 2006) for assigning subjects to classes. Computation times for the CCFI restricted this study to one taxometric procedure. MAXEIG was selected because of its multivariate treatment of data, and because of its bivariate special case maximum covariance (MAXCOV), which is the most widely studied and applied taxometric procedure (see Ruscio et al., 2006, for a detailed review of studies using each taxometric procedure).

The general idea of taxometric procedures can be illustrated by means of the MAXCOV procedure (Meehl, 1973). MAXCOV evaluates the covariance of two indicators called output indicators for different ranges (e.g., “windows”) of a third indicator called the input indicator. The input indicator is assumed to be a proxy of a latent dimension on which the two classes (if they exist) should differ. If subjects are ordered on the input indicator, then sliding a window over the range of the output indicators will initially contain only subjects of one class, then a mix of two classes, and, finally, only subjects of the second class. Assuming local independence within class, the covariance between the two output variables is zero if a window of the indicator variable only contains subjects of the first class. The covariance increases as the window contains subjects from two classes because it deviates from zero due to mean differences between the classes, and then decreases again when the sliding window contains only subjects of the second class. The user specifies the number of overlapping windows. The degree of overlap is determined by the subsample size and the number of subsamples. The covariances between the two output variables can then be

¹See Tueller and Lubke (2010) for methods to estimate the uncertainty in classification probabilities in applied settings.

plotted and visually inspected. A flat plot should indicate absence of latent clustering, whereas a peaked plot is interpreted as evidence of a taxon and its complement. Maraun, Slaney, and Goddyn (2003) and Maraun and Slaney (2005) have investigated data types that are exceptions to the general hypothesis that one-class data will produce a flat plot and that two-class data will produce a single peak.²

Although MAXCOV can in principle be repeated for all pairs of variables in a multivariate data set, the approach becomes cumbersome as the number of variables increases. MAXEIG has been developed to treat larger numbers of indicators more parsimoniously (Waller & Meehl, 1998). As in MAXCOV, subjects are first ordered on the input variable to select overlapping subsamples with increasing means on the input variable. Instead of taking the covariance of a pair of output indicators in each window, MAXEIG uses the largest eigenvalue of the covariance matrix of the variables. The eigenvalues are computed using a modified covariance matrix. The variances on the diagonal of the covariance matrix are replaced with zeros, leaving only the covariances. The removal of variances enhances the difference in the largest eigenvalue across overlapping subsamples if local independence within class holds, and therefore eases the decision making when evaluating the plots. Just as with MAXCOV, a single-peaked MAXEIG plot is taken as evidence that the data come from two classes and that a flat MAXEIG plot indicates that data come from a continuous underlying trait.

This process is repeated with each variable acting as the input indicator, resulting in as many MAXEIG plots as there are indicators. Examining each plot can reveal indicators that do or do not discriminate between classes, and an average plot can be produced and examined to inform the general decision of whether there is sufficient evidence to conclude that there are one or two classes. The point of the maximum eigenvalues (called the HITMAX) is used to compute estimates of the base rate (i.e., class proportions), and the average of the base rates across all input indicators is typically used as the final estimate of the base rate of the sample.

Recent taxometric studies have used the base rate classification technique to assign subjects to classes. Base rate classification requires the raw data and an estimate of the base rate, and is therefore independent of the clustering method beyond estimation of the base rate. The total sum score for all observed variables is computed and sorted from the smallest to the largest value. Then cases with the highest scores are assigned to the second class such that the proportion of cases in the second class equals the estimate of the base rate for the second class (Ruscio, 2007; Ruscio et al., 2006).

After assigning subjects to classes, within-class means and correlations can be computed to assess model fit with the CCFI (Ruscio & Marcus, 2007; Ruscio et al., 2007). More specifically, determining model fit is accomplished by computing the CCFI. Bootstrapping is used to generate one- and two-class data sets that reproduce distributions of the indicators. For both the two-class and the one-class simulated data sets, the root mean square residual (RMSR) is calculated. The CCFI is the ratio of the one-class RMSR and the sum of the one-class and two-class RMSRs. It ranges from 0 to 1, where 0 indicates best fit of the data with the one-class model, 1 indicates the best fit of the data with the two-class model, and .5 indicates the same evidence of (mis)fit for both the one- and two-class models. Although the CCFI is not the only means of assessing whether the one-class or two-class model fits better,

²Older references to latent classes typically construe a class as a set of degenerate distributions (e.g., a two-class model has two degenerate or zero variance distributions). We advocate generalizing the definition of class to be the number of component distributions in any mixture model, whether degenerate or not. Hence, a one-class model refers to any traditional analysis that assumes population homogeneity such as CFA or multiple regression.

the CCFI has recently been shown to perform better than previously used indexes or judgment-based graphical inspection procedures (Ruscio & Marcus, 2007; Ruscio et al., 2007).

An advantage of taxometric procedures is that their application does not necessitate complex model specifications or distributional assumptions. The procedures have several limitations. First, the procedures are primarily designed and used to detect only two types within a population at one time. Grath (2008) showed that when there are three classes in the data, standard implementation of taxometric procedures will lead to incorrect or inconclusive results. A second limitation is that measurement error in the observed variables is not taken into account. Third, the procedures depend on low correlations of observed variables within class. Taxometric procedures are reported to perform ideally for within-class correlations up to .3 (Meehl, 1995); however, they have been shown to perform well for within-class correlations up to .6 for a certain implementation of MAXCOV (Beauchaine & Beauchaine, 2002). It is unknown whether this generalizes to other taxometric procedures and correlations up to .6 are included in the following simulation study to investigate this issue. Fourth, detection of latent classes is most commonly not a research goal per se. Using taxometric procedures, further investigation of the latent classes such as the model fit comparison of one- and two-class models with the CCFI, or the comparison of within-class factor structures, latent class means, or the relation of latent classes to covariates, can only be accomplished post-hoc after assigning subjects to classes (Bernstein et al., 2007). One of the two goals of this study is to quantify the quality of class assignment. The effects of incorrect assignment on post-hoc testing are investigated in ongoing research (Lubke, Carey, Lessem, & Hewitt, 2008).

Factor Mixture Modeling

FMM combines latent class and latent factor models, and permits researchers to compare the fit of different within-class structures such as factor models versus local independence (Lubke & Neale, 2006, 2008). Different forms of the general model have been described by Heinen (1996), Yung (1997), Jedidi et al. (1997), Dolan and Van der Maas (1998), B. Muthén and Shedden (1999), and Arminger et al. (1999). Within each class, a standard common factor model for a single homogeneous population is specified. Assuming normality for the factors and residual variances of observed variables, linearity of the regression of manifest variables on the factors, uncorrelatedness of factors, and errors, the manifest variables within-class are multivariate normal. The joint distribution of observed variables is consequently a mixture of these multivariate normals.

$$f(\mathbf{y}) = \sum_{k=1}^K \pi_k \phi_k(\mathbf{y}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (1)$$

where \mathbf{y} is a vector of p observed continuous variables, K is the number of classes, π_k are the class proportions with $\sum_{k=1}^K \pi_k = 1$, and ϕ_k are multivariate normal probability density functions (PDFs) with class-specific mean vectors $\boldsymbol{\mu}_k$ and class-specific covariance matrices $\boldsymbol{\Sigma}_k$.

A factor model is imposed on each of the component distributions:

$$\boldsymbol{\mu}_k = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \boldsymbol{\alpha}_k \quad (2)$$

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}_k \boldsymbol{\Psi}_k \boldsymbol{\Lambda}_k' + \boldsymbol{\Theta}_k, \quad (3)$$

where \mathbf{v}_k is a $p \times 1$ vector of equation intercepts in class k , p is the number of observed variables, $\mathbf{\Lambda}_k$ is a $p \times m_k$ matrix of factor loadings in class k , m_k is the number of factors in class k , $\boldsymbol{\alpha}_k$ is an $m_k \times 1$ vector of factor means in class k , $\boldsymbol{\Psi}_k$ is the $m_k \times m_k$ covariance matrix for the factors in class k , and Θ_k is a $p \times p$ covariance matrix of the measurement errors with error variances on the diagonal.

Note that local independence can be specified as a factor model with zero loadings or zero factor variance(s). The general model therefore offers great flexibility with respect to the within-class structure and the number of classes.

In practice, using the FMM usually involves fitting models with different numbers of classes and sometimes different within-class factor structures. The best fitting model or models are selected using information criteria such as the Bayesian Information Criterion (BIC; Schwarz, 1978) or bootstrapped or adjusted versions of the likelihood ratio test (Lo, Mendell, & Rubin, 2001; Vuong, 1989). Nylund, Asparouhov, and Muthén (2007) showed that the BIC performs well under a variety of conditions, and is only sometimes slightly outperformed by the bootstrapped LRT. The adjusted LRT (aLRT) only performs well for simple models, and under those conditions can outperform the BIC. We will base our comparison mainly on the BIC. Due to computation times the bootstrapped LRT is not feasible in this simulation. We also report the aLRT. In the context of FMM, subjects can be assigned to their most likely class using the highest posterior probability of belonging to a class. This is commonly called *modal assignment*.

The FMM has a set of strengths and limitations that is very different from those of taxometric procedures. FMM has the advantages of easily accommodating more than two latent classes, of explicitly accounting for measurement error in the observed variables, and of explicitly modeling within-class covariance structures. Generalizations, for instance, to include structural equations between factors have been described (Henson, Reise, & Kim, 2007; Jedidi et al., 1997; Tueller & Lubke, 2010), and the model enjoys great popularity.

Limitations include the requirement of the specification of a factor model for the within-class covariance matrices and mean vectors. A model is generally a simplification of the true data-generating process, and can also contain more or less severe misspecifications. In addition to misspecifications of the within-class factor structure, factors or errors in a given cluster of subjects might not be normally distributed, or observed variables might not be linearly related to the factors. Furthermore, each cluster of subjects in the population is thought to correspond to one of the K component distributions, and, consequently, estimates of the parameters π_k are interpreted as the relative class size. When fitting FMMs to empirical data, this one-to-one correspondence of clusters of subjects and mixture components is not necessarily given (see, e.g., Bauer & Curran, 2003). On the positive side, Lubke and Neale (2006, 2008) showed that comparisons of alternative models lead to correct model choice in a wide variety of scenarios.

Class Separation and Within-Class Covariance

Class separation is a crucial determinant of the success of any clustering approach. The taxometric and FMM literature use different measures to indicate the minimal separation needed for an adequate analysis. To connect the taxometric and FMM literatures, their respective definitions for class separation are compared here. In the taxometric literature, class separation is defined as the standardized mean difference (Cohen's distance) between classes for each indicator. The taxometric literature generally recommends that each indicator has a minimum Cohen's distance of 1.25 (Meehl, 1995; Ruscio et al., 2006). Cohen's distance is defined by

$$d_p = \frac{(\mu_{p1} - \mu_{p2})}{\sigma_p} \quad (4)$$

where the subscript p corresponds to the p th observed variable, μ_{p1} and μ_{p2} are the first- and second-class means for the p th observed variable, and σ_p is the pooled variance for the p th observed variable.

In the FMM literature the multivariate Mahalanobis distance (MD) is often used to measure class separation. Different authors show that $MD = 1.5$ seems to suffice when classes have equal size (Lubke & Neale, 2006, 2008; Yung, 1997), but larger distances are necessary for unbalanced class sizes (Tueller & Lubke, 2010). The MD is a multivariate extension of Cohen's distance and is given as

$$MD = \sqrt{(\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2)}, \quad (5)$$

where μ_1 is the p -dimensional vector of item means in the first class, μ_2 is the p -dimensional vector of item means in the second class, and Σ is the $p \times p$ pooled covariance matrix. The MD takes into account the covariances between variables. For example, all else being equal, two classes with large within-class correlations have a smaller MD compared to two classes with low within-class correlations.³ In addition, the MD increases with the number of variables. The relation between MD and Cohen's distance is shown in Table 1. As can be seen, the MD increases substantially with increasing numbers of indicators when item correlations are zero. This effect tapers off with increasing correlations. Table 1 also shows that the prior recommendations for taxometric procedures (e.g., zero correlations within class, $d = 1:25$) correspond to $MD > 2:0$, which is a rather unproblematic setting for FMM models if class sizes are equal. In our simulation we choose a sufficiently large range of settings such that a deterioration of performance can be detected for both methods. First, we illustrate the methods using empirical data.

DEPRESSION DATA ILLUSTRATION

The depression data illustration demonstrates how taxometric procedures and the FMM can be used for a given data set. The taxometric literature provides much advice on exploratory steps that should be taken prior to taxometric analyses. If data do not meet certain standards, the advice is to obtain other data. In practice it is often difficult to obtain additional data, or to construct measures that meet taxometric requirements. Also, taxometric procedures are limited to testing a two-class hypothesis against a one-class hypothesis. The illustration demonstrates how alternative ideas can be translated into different FMMs that can be compared using, for example, the BIC. We use standard questionnaire items that are common in psychiatric and psychopathological research focusing on the investigation of subtypes, which is the main area of application of taxometric procedures.

The data for the empirical illustration are 10 binary items that are matched to symptoms of Major Depressive Disorder (MDD) as defined in the Diagnostic and Statistical Manual of Mental Disorders (3rd ed., rev.; *DSM-III-R*; American Psychiatric Association, 1987), and come from a subset of the Virginia Twin Registry (Kendler & Prescott, 1999). The Virginia

³Anderson and Bahadur (1962) provided a generalization of the MD that is more accurate for imbalanced class covariance matrices. The simulations in this work examine balanced class covariance matrices, and the authors are currently examining imbalanced class covariance matrices. Note that Maraun and Slaney (2005) showed analytically that MAXCOV is not guaranteed to produce a single-peaked plot in the presence of two classes under imbalanced class covariance matrices, and by extension, MAXEIG will have the same problem.

Twin Registry is a population-based register formed from a systematic review of all birth certificates in the Commonwealth of Virginia from 1918 onward. Twins were eligible for participation in each of the studies if one or both twins were successfully matched to birth records and were born between 1940 and 1974. To address dependence of twin pair data, and to avoid the potential confounding of a latent class structure with known gender differences of MDD prevalence, in this example we only use data from one male twin from each male–male or male–female twin pair. A factor mixture analysis of female twins has been reported by Lubke and Neale (2008) to illustrate the effect of power when comparing models with and without class-specific model parameters. Here, the emphasis is on demonstrating how factor mixture models can be used to model psychiatric data from a general population sample, and how results can inform a researcher about the disorder. The taxometric analysis of the depression items is presented first.

Taxometric Analyses

A strong recommendation found in the taxometric literature is to rely on converging sources of evidence from multiple taxometric procedures. In addition to MAXEIG, the taxometric procedures MAXCOV, L-Mode (Waller & Meehl, 1998), and Mean Above and Mean Below a Cut (MAMBAC; Meehl & Yonce, 1994) were used for the depression data. We use the CCFI to assess latent structure detection for the depression data.

The L-Mode procedure fits a unidimensional factor model to a set of indicators and examines the distribution of Bartlett factor scores (Bartlett, 1937). A bimodal distribution is taken as evidence that data come from two latent classes (Ruscio et al., 2006; Waller & Meehl, 1998). Note, however, that only large mean differences lead to bimodality (for examples, see McLachlan & Peel, 2000). L-Mode has not been systematically studied, but initial evidence indicates that it performs well under several conditions (Meehl & Yonce, 1994).

The MAMBAC procedure searches for an optimal cutting score. One output indicator is sorted on an input indicator, and the input indicator is used to make a series of cutting scores. For each cutting score, the mean of the output indicator below and the mean of the output indicator above the cut is computed. The mean differences are then plotted. If data come from two classes, the plot should be single peaked at the point where the data can be optimally cut into two groups with the largest class separation. If data come from a single class, MAMBAC typically produces a U-shaped plot (Meehl & Yonce, 1994; Ruscio et al., 2006). In general, MAMBAC does not perform as well as MAXCOV or MAXEIG (Ruscio et al., 2006).

Because the depression items are binary, composite input indicators are formed using the sum of all variables not being used as output indicators for MAXEIG, MAXCOV, and MAMBAC analyses. Composite input indicators ensure there is sufficient variation for reliable sorting of the data when categorical indicators are used (Ruscio, Haslam, & Ruscio, 2006). MAXEIG, MAXCOV, and MAMBAC analyses all used 50 windows, and MAXEIG and MAXCOV analyses set window overlap to be 90%.

Taxometric results—Results of the taxometric analyses of the depression data are summarized in Table 2. In the current application, the four taxometric methods yielded diverging conclusions about the data. L-Mode and MAMBAC did not support either one-class or two-class conclusions (CCFI = .52 and CCFI = .50, respectively), whereas the MAXCOV and MAXEIG analyses supported a one-class conclusion (CCFI = .34 and CCFI = .33, respectively). The estimates of the correlations between indicators within class are somewhat higher in the complement for MAXCOV and MAXEIG, equal for MAMBAC, and higher in the taxon for L-Mode.

FMM Analyses

We fit the following initial set of models to the data: latent class models (i.e., local independence within class) with two to five classes, a single-factor, single-class model (analogous to the dimensional model in taxometric procedures), single-factor models that constrain loadings and item thresholds to be class-invariant (e.g., measurement invariant models) with two to four classes, and single-factor models that relax the measurement invariance (MI) constraints and permit differences in loadings and thresholds with two to four classes. The results are presented in Table 3.

Consistent with previous findings, the Akaike's Information Criterion (AIC) favors more complex models, whereas the BIC penalizes the increase in parameters more heavily. Adding classes with many class-specific parameters such as in the case of the measurement noninvariant models can therefore result in a larger BIC, which, in turn, can lead to a potentially erroneous conclusion that MI holds across classes (Lubke & Neale, 2008). Especially in this example, where symptom endorsements of MDD are observed in a population sample, it is questionable that items (e.g., "thoughts of suicide") discriminate equally well within the unaffected part of the population and within the affected part. In fact, the data contain a large group of subjects with zero scores on (almost) all items. This means that when fitting single-factor, two-class models, there is little or no information to estimate factor loadings and thresholds for subjects with low scores on the depression factor. Fitting MI models simply equates the loading and threshold estimates obtained from subjects with varying levels on the depression factor to those of the unaffected class. More realistic is a model that imposes a structure reflecting mainly zero scores in the unaffected class, and a single-factor model in the class that contains subjects with severity differences in depression.

A preponderance of zeros can be modeled with two-part models, and also with a model that fixes thresholds to large values in the unaffected class while estimating class-specific factor loadings and factor variances. We fitted such a model, and obtained a fit with $LL = -5427.77$ with 32 estimated parameters. The BIC was comparable to the more parsimonious MI models (see Table 3), and equaled 11094.52. Both AIC and saBIC were clearly better than any of the MI models, namely 10919.54 and 10992.86, respectively. Evidence supporting MI seems weak given the observed data pattern of a large zero-scoring group of subjects and previous simulation results showing that the BIC penalty on model complexity might be too high to reject MI. Importantly, MI can only be investigated adequately if the set of items covers the range of the trait in both groups or classes. The data in this analysis are best described by a model that acknowledges the fact that the items have high item difficulties and do not adequately measure the trait in the unaffected group. For comparison, this pattern of results was less pronounced in the analysis of the females described in Lubke and Neale (2008), which is likely due to a higher prevalence of depression in females.

FMM provides the possibility of investigating differences between classes with respect to variables of interest in a single analysis. Taxometric procedures require that the investigator first assign subjects classes such that post-hoc comparisons can be carried out in a second step. In FMMs, effects can be investigated by estimating different types of covariate effects. The effect of a covariate on depression might, for instance, be completely mediated by class membership, or might essentially consist of a direct effect of the covariate on the depression factor within class. We fitted a model where (a) class membership and (b) depression within class in the more severe class is regressed on age. The results show that that both effects are small but significant given the sample size. Although the age in the unaffected class is somewhat higher in our sample (the log odds of being in the affected class predicted by age are .975), depression within the affected class is positively related to age. This is a well-known effect.

Using taxometric procedures, a more fine-grained analysis of within-class structures as described in this section can only be accomplished post-hoc after assigning subjects to classes, and hinges on low error rates in the assignment. The evaluation of correct class assignment is one of the two goals of the simulation study.

SIMULATION STUDY

The following sections outline the design of the main simulation study, the data generation procedures, and the analyses carried out on the generated data.

Simulation Study Design and Data Generation

The conditions for the simulation study were selected to compare MAXEIG and FMM performance under conditions that have been reported to be ideal for either method, and conditions that were expected to be problematic but that are quite common in empirical data: balanced versus imbalanced class proportions, locally independent data versus data having within-class variability, and large versus moderate class separations.

Data were generated with either balanced class proportions or with a .95/.05 split reflecting the presence of a majority and a minority class. With $N = 400$ subjects in each data set, the balanced class proportion conditions will have on average 200 subjects in each class, and the imbalanced class proportion conditions will have on average 380 subjects in the first class and 20 in the second class. Prior studies have shown poor FMM performance with severely imbalanced class proportion data (Tueller & Lubke, 2010), whereas the taxometric literature has been more optimistic on detecting small classes (Ruscio & Marcus, 2007). Thus, taxometric procedures are expected to outperform the FMM for locally independent data with large class separations and unbalanced class sizes.

Multivariate normal data were generated under five different two-class models. Models 1 and 2 are LCA models that differ by the number of indicators (i.e., 5 and 10 indicators), Model 3 is a single-factor FMM, and Models 4 and 5 are two two-factor FMMs that differ with respect to the presences or absence of cross-loadings. The five models are denoted as LCA5, LCA10, 1F, and 2Fcl, and 2Fss, respectively, where cl indicates cross-loadings and ss indicates simple structure. We focus on two-class models because taxometric procedures have not yet been extended to three or more classes and can lead to incorrect conclusions when there are three classes (Grath, 2008). Within-class variances were balanced in all five models.

The LCA data are locally independent given class membership and have zero within-class correlations. LCA conditions are favorable for taxometric procedures and FMM alike. The one- and two-factor data illustrate within-class individual differences, which are common in empirical data (e.g., Lubke et al., 2009; Lubke et al., 2007). The FMM has the advantage of being able to model these individual differences using within-class confirmatory factor analysis (CFA) models. The within-class item correlations in our simulated data range from .3 to .6. Correlations of .3 fall at the high end of favorable conditions of taxometrics (e.g., Ruscio et al., 2006) but lead to only moderate within-class item reliabilities, which could be problematic for within-class CFAs in the FMM, whereas correlations of .6 have been shown to be tolerable for taxometric procedures in prior research (Bernstein et al., 2007) and lead to reasonable within-class item reliabilities for the FMM.

As noted in the section on class separation, both methods require sufficient separation to correctly detect latent classes. In the simulation study, class separations of $MD = 1.5$ and $MD = 3.0$ were examined. MD was controlled by manipulating the factor mean difference between the two classes. For the LCA models, the manipulation was carried out at the level

of the observed variables. As noted earlier, FMM research suggests $MD = 1.5$ is a minimal condition needed for the FMM. $MD = 3.0$ is a very large effect. As can be seen in Table 4, the corresponding Cohen's d values range from moderately below to well above the recommended minimum of $d = 1:25$. Table 4 suggests that the FMM is likely to perform better for smaller class separations.

In sum, the simulation design is a 5 (models) \times 2 (MD distances) \times 2 (class proportion settings) design. For each cell of the design, we generated 100 data sets. Population parameter values are given in the Appendix.

Analyses

The following sections provide implementation details for taxometric and FMM analyses, followed by the methods used to assess the performance regarding class detection and assignment.

Taxometric analyses—As described earlier, MAXEIG, the base rate classification technique, and the CCFI were used for the taxometric analyses in this study. Computation times for the CCFI restricted this study to one taxometric procedure, and MAXEIG was selected because of its multivariate treatment of data, and because its bivariate special case, MAXCOV, is the most widely applied taxometric procedure. The MAXEIG analyses were performed using 50 windows with 90% overlap, and each variable was used as the input indicator once with the remaining nine indicators acting as output indicators. In computing the CCFI, 30 two-class and 30 one-class comparison data sets were generated. Following Ruscio and Marcus (2007), if the CCFI for a given data set was greater than .5, the analysis was recorded as favoring the two-class taxonic rather than the dimensional structure. The taxometric analyses were carried out using R (R Development Core Team, 2008) code written and maintained by John Ruscio, available at <http://www.taxometricmethod.com/>.

FMM analyses—Because in practice FMM analyses are subject to potential misspecifications of the number of classes, the number of factors, and the within-class factor structure, class enumeration is obtained by comparing the fit of a set of alternative models. For the LCA data, two models were fit to each data set: (a) a single-factor, one-class model, and (b) a two-class LCA model. These two models represent the competing hypotheses of a single underlying continuum versus latent clustering in the form of two subtypes. For the one- and two-factor two-class FMM data, four alternative models were fit to each data set: (a) a single-class, one-factor model, (b) a two-class, one-factor model, (c) a two-class, two-factor model with simple structure, and (d) a two-class, two-factor model with cross-loaded indicators. We determined that FMM correctly concluded that two classes exist in a data set if (a) the correct model has converged, and (b) the BIC selected the correct FMM from among the four models. Class assignment is implemented using modal assignment; that is, subjects are assigned to the class with the highest posterior class probability.

The FMM models are fit to the data using the software program Mplus version 5.0 (L. K. Muthén & Muthén, 2007) on 20 dual-processor PC workstations managed by the Condor High-Throughput Computing System (Thain, Tannenbaum, & Livny, 2005).

Assessment of class detection and assignment—The goal of the analyses in this study is to investigate to what extent taxometrics and FMM are able to (a) detect the correct number of classes, and (b) accurately assign subjects to classes. Because all data sets were generated under a two-class model, the first goal is assessed by computing the proportion of data sets in which two classes were detected. To assess the accuracy of the assignment of subjects to classes in the second goal, sensitivity, specificity, and the Hubert–Arabie

Adjusted Rand Index (ARI_{HA}) are computed. Because in the imbalanced class proportion conditions the second class is the (affected) minority class, we define sensitivity as the proportion of Class 2 subjects correctly assigned to Class 2. Specificity is defined as the proportion of Class 1 subjects correctly classified as being in Class 1.

The ARI_{HA} is an overall proportion correct assignment measure adjusted for chance. The index was proposed by Hubert and Arabie (1985). In a large simulation study, Steinley (2004, p. 392) developed a heuristic for interpreting ARI_{HA} values where “(a) values greater than 0.90 can be viewed as excellent recovery, (b) values greater than 0.80 can be considered good recovery, (c) values greater than 0.65 can be considered moderate recovery, and (d) values less than 0.65 reflect poor recovery.” This heuristic is used in interpreting the ARI_{HA} . The ARI_{HA} values in this study were computed using the `adjustedRandIndex()` function in the R package MCLUST version 3 (Banfield & Raftery, 1993; Fraley & Raftery, 1999, 2002, 2003, 2006; R Development Core Team, 2008).

RESULTS

The results are presented first for the first goal, correct detection of two classes, followed by results for the second goal, the accuracy of class assignment. Tables are structured according to the simulation design described previously.

Latent Class Detection

Class detection and assignment in FMM analyses is contingent on model convergence. Convergence rates are given in Table 5, and the proportion of data sets for which taxometric and FMM analyses correctly concluded that there were two classes are given in Table 6. Comparing the two tables shows that conditional on convergence, FMM almost always detected the two classes. Convergence rates were good except for the more complex two-factor models in the imbalanced class proportion conditions. Parameter estimates including the class proportions can diverge from the population values substantially if class separation is small. Because each subject’s contribution to the likelihood is weighted by the posterior class probability, this can have the effect that only very few subjects contribute substantial information to the estimation of the within-class model, which, in turn, will affect convergence. The two-factor models require the estimation of more parameters than the one-factor models, which increases the chance of empirical underidentification for the minority class. A similar result has been reported by Tueller and Lubke (2010). Note that increased class separation alleviates much of the problem due to imbalanced class proportions. Larger sample sizes are needed, especially when small classes are expected.

In general, the FMM outperformed taxometric procedures in detecting the two classes. For the LCA data, the taxometric procedures performed better for the 5-indicator data than for the 10-indicator data. Even though the MD is constant for the 5- and 10-indicator conditions, the univariate class separation is smaller for the 10-indicator data (see Table 1). Although MAXEIG is a multivariate technique, final MAXEIG results are averaged over runs where each indicator acts as the *univariate* input indicator once. Hence, MAXEIG results are expected to be affected by univariate class separation.

As expected, the larger class separation conditions ($MD = 3.0$) resulted in more accurate class detection for both methods than the smaller class separation conditions ($MD = 1.5$). Taxometric procedures rarely detected the two classes in small class separation conditions, as can be seen in the first column of Table 6. The FMM results concerning class detection in Table 6 are based on the BIC. The one-factor FMM models also had poor detection for the smaller class separation conditions when using the BIC as a criterion. This result is consistent with Nylund, Asparouhov, and Muthén (2007). The aLRT provided much better

detection rates for the single-factor model, namely 99%. However, the aLRT is not consistent across different model types and clearly underperforms the BIC when considering more complex models (see also Nylund, Asparouhov, & Muthén, 2007). Using the BIC, the FMM accurately detected the two classes of the more complex models for balanced class proportion conditions. This is shown in the upper half of Table 6. For the imbalanced class proportion conditions, the FMM performed well for the LCA data but deteriorated for the one- and two-factor data when $MD = 1.5$. For imbalanced class proportions and $MD = 3.0$, both FMM and taxometric class detection were moderate to good, as can be seen in the lower right of Table 6.

Class Assignment

Class assignment is evaluated conditional on class detection (see Table 5), and is therefore based on different numbers of data sets for the different cells of the simulation design. In the small class separation conditions, class detection for taxometrics was between .1 and .5 for all generated data types except the LCA with 5 indicators, hence results regarding sensitivity and specificity are not generalizable.

Sensitivity and specificity—The sensitivity and specificity for taxometric and FMM class assignment are summarized for each condition in Table 7 and Table 8. Results that are based on fewer than six data sets are marked in italics. The definition of sensitivity as the number of true positives given diseased status is translated here as the number of true minority class members being assigned to the minority class (i.e., Class 2). Similarly, specificity is computed as correct assignment to Class 1.

Taxometric procedures and the FMM have comparable sensitivity and specificity in the ideal conditions with large class separation and balanced class proportions, as shown in the upper right blocks of Tables 7 and 8. The FMM performed slightly better for data generated under the LCA models. Taxometrics had better sensitivity (i.e., correctly assigning minority class subjects) for the large-separation imbalanced-class proportion conditions, namely .74–.95 versus .69–.70 for the FMM (see lower right block of Table 7).

For the small class separation, $MD = 1.5$, the FMM only performed well in the balanced class proportions. In the unbalanced condition, FMM had a sensitivity around .3 for all data types (i.e., about 70% of truly minority class members are incorrectly assigned). Specificity was acceptable (i.e., around .95); however, in practice the main interest is usually in the smaller classes. Taxometrics only detected two classes in 1 to 5 out of 100 generated data sets. The results for the successfully detected two-class data, given in italics in Table 7 and Table 8, should not be generalized as they might reflect advantageous sampling fluctuation.

Hubert–Arabie Adjusted Rand Index—In general, overall class assignment was poor in the small separation conditions as seen in the left half of Table 9. Both methods had moderate class membership recovery in the balanced-class proportion and large separation conditions as seen in the upper right of Table 9. The FMM also had moderate class recovery for the large separation imbalanced-class proportion conditions as seen in the lower right of Table 9, whereas taxometric procedures had poor recovery for these conditions.

CONCLUSIONS

The empirical example shows that taxometric procedures were inconclusive whether a taxonic or dimensional structure was supported. The FMM approach indicated two classes, a majority class with a preponderance of zero scores, and a single-factor model in the minority class containing subjects with positive item endorsements. The analysis illustrated the flexibility of the FMM to test more specific hypotheses than taxometric procedures. We also

showed that inclusion of covariate effects is useful to investigate differences between classes without the need of assigning subjects to their most likely class.

The simulation study shows that FMMs and MAXEIG provide similar results for class detection and assignment under ideal conditions of balanced class proportions and large class separation. Under more realistic conditions of small class separation or imbalanced class proportions, the FMM outperforms taxometrics in the detection of the two latent classes. The study also underlines the limitations of FMMs. In conditions of small class separation and imbalanced class proportions, FMM does not detect the second class with the sample sizes used in this simulation. FMM class detection is likely to improve with larger sample sizes as class detection is a matter of power (Lubke & Neale, 2008). Nonconvergence in this simulation is likely due to empirical underidentification of the minority class, which is a threat that should not be overlooked in empirical studies involving small minority classes.

Similar to class detection, the results concerning correct class assignment are comparable for the two methods under ideal conditions of large class separation. When class sizes are unequal, using the highest posterior probability to assign subjects to classes resulted in somewhat lower sensitivity for FMM compared to taxometrics. Correct assignment to a majority class (specificity) using FMM is better than correct assignment to a minority class (sensitivity), which is expected given the prior probability of correct assignment. Keeping the distance between classes and within-class variances equal, then decreasing the class size of the minority class will result in an increasing proportion of minority subjects with $P_{majority} > P_{minority}$, where p is the probability under the normal weighted by the prior class probability, and where majority and minority refer to the component distributions of the larger and the smaller class. In our study, sensitivity is poor for the FMM when class separation is small and class size is unequal. Sensitivity and specificity for taxometrics under these conditions could not be evaluated because taxometrics did not detect the second class in a sufficient proportion of the data sets.

In general, our results show that MAXEIG requires greater class separation than the FMM, and that within-class correlations do not seem to have a systematic negative effect on MAXEIG performance when the MD is held constant (e.g., compare the LCA10 results to the FMM results in Tables 6–9). In addition, our results show that error rates in class assignment accuracy can be substantial. The impact of assignment error on post-hoc testing needs is currently under investigation.

Limitations of this study include the following. The simulation study was limited to a small set of conditions that are, however, quite common in practice. Other interesting conditions would include within-class nonnormality and imbalanced class variances. Imbalanced class variances are investigated in Tueller and Lubke (2010) for generalizations of the FMM and are currently being investigated in the context of comparing the FMM and MAXEIG. Second, mixture distributions can approximate nonnormal distributions. When fitting mixture models to data that have within-class nonnormality, the additional classes needed to approximate the distribution might not necessarily reflect true subgroups. Third, our study limited the data generation to two classes. Assignment error might have different patterns in cases with more than two classes, and will likely depend on factors such as ordering of the classes along a common dimension versus qualitatively different classes, and the resulting mutual class separation. Taxometrics, however, are most commonly used to assess whether data come from a single homogenous population versus a population consisting of two classes. The weaker class detection compared to FMM, combined with potentially high error rates when class separation is small, limits the utility of taxometric procedures, especially

given the fact that evidence of a taxonic structure by itself is usually only the first step rather than the final goal of a study.

Acknowledgments

The research of the first author was supported through grant DA018673 by NIDA.

APPENDIX PARAMETER VALUES

Latent Class Models

residual variances [.5 .5 .5 .5 .5 .5 .5 .5 .5 .5]'

Intercepts in the second class (first class intercepts set to zero): MD = 1.5

5 item [.4744 .4744 .4744 .4744 .4744 .4744 .4744 .4744 .4744 .4744]

10 item [.3355 .3355 .3355 .3355 .3355 .3355 .3355 .3355 .3355 .3355]

MD = 3.0

5 item [.675 .675 .675 .675 .675 .675 .675 .675 .675 .675]

10 item [.95 .95 .95 .95 .95 .95 .95 .95 .95 .95]

One-Factor Model

Class-invariant parameters:

factor loadings [1 .8 .8 .8 .8 .8 .8 .8 .8 .8]'

factor variance 1

residual variances [.5 .5 .5 .5 .5 .5 .5 .5 .5 .5]'

Class-specific parameters:

factor mean in the second class MD = 1.5 [1.57]

factor mean in the second class MD = 3.0 [2.1]

Two-Factor Models

Class-invariant parameters:

factor loadings, simple structure $\begin{bmatrix} 1 & .8 & .8 & .8 & .8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & .8 & .8 & .8 & .8 \end{bmatrix}'$

factor loadings, cross-loadings $\begin{bmatrix} 1 & .8 & .6 & .4 & .8 & 0 & .2 & 0 & .4 & 0 \\ 0 & 0 & .2 & .4 & 0 & 1 & .6 & .8 & .4 & .8 \end{bmatrix}'$

factor covariance matrix, positive factor correlation $\begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}'$

residual variances [.5 .5 .5 .5 .5 .5 .5 .5 .5]'

Class-specific parameters:

factor means in the second class

MD = 1.5 [1.399 1.399]

MD = 3.0 [2.799 2.799]

REFERENCES

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd ed. Author; Washington, DC: 1987.
- American Psychiatric Publishing. Diagnostic and statistical manual of mental disorders. 4th ed. Author; Washington, DC: 2000.
- Anderson TW, Bahadur RR. Classification into two multivariate normal distributions with different covariance matrices. *The Annals of Mathematical Statistics*. 1962; 33:420–431.
- Arminger G, Stein P, Wittenberg J. Mixtures of conditional mean- and covariance-structure models. *Psychometrika*. 1999; 64:475–494.
- Banfield JD, Raftery AE. Model-based Gaussian and non-Gaussian clustering. *Biometrics*. 1993; 49:803–821.
- Bartlett MS. The statistical conception of mental factors. *British Journal of Psychology*. 1937; 28:97–104.
- Bauer DJ, Curran PJ. Overextraction of latent trajectory classes: Much ado about nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003). *Psychological Methods*. 2003; 8(3):384–393.
- Beauchaine TP, Beauchaine RJ. A comparison of maximum covariance and k-means cluster analysis in classifying cases into known taxon groups. *Psychological Methods*. 2002; 7:245–261. [PubMed: 12090413]
- Bernstein A, Zvolensky MJ, Norton PJ, Schmidt NB, Taylor S, Forsyth JP, et al. Taxometric and factor analytic models of anxiety sensitivity: Integrating approaches to latent structural research. *Psychological Assessment*. 2007; 19:74–87. [PubMed: 17371124]
- Cleland CM, Rothschild L, Haslan N. Detecting latent taxa: Monte Carlo comparison of taxometric, mixture model, and clustering procedures. *Psychological Reports*. 2000; 87:37–47. [PubMed: 11026388]
- Dolan CV, Van der Maas HLJ. Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*. 1998; 63:227–253.
- Fraley C, Raftery AE. MCLUST: Software for model-based cluster analysis. *Journal of Classification*. 1999; 16:297–306.
- Fraley C, Raftery AE. Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association*. 2002; 97:611–631.
- Fraley C, Raftery AE. Enhanced software for model-based clustering, density estimation, and discriminant analysis: MCLUST. *Journal of Classification*. 2003; 20:263–286.
- Fraley, C.; Raftery, AE. MCLUST version 3 for R: Normal mixture modeling and model-based clustering. Department of Statistics, University of Washington; Seattle: Sep. 2006 (Tech. Rep. No. 504)
- Grath REM. Inferential errors in taxometric analyses of ordered three-class constructs. *Journal of Personality Assessment*. 2008; 90(1):11–25. [PubMed: 18444091]
- Greenbaum PE, Dedrick RF. Changes in use of alcohol, marijuana, and services by adolescents with serious emotional disturbance: A parallel-process growth mixture model. *Journal of Emotional and Behavioral Disorders*. 2007; 15(1):21–32.
- Heinen, T. Latent class and discrete latent trait models: Similarities and differences. Sage; Newbury Park, CA: 1996.

- Henson JM, Reise SP, Kim KH. Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling*. 2007; 14:202–226.
- Hubert L, Arabie P. Comparing partitions. *Journal of Classification*. 1985; 2:193–218.
- Hudziak JJ, Achenbach TM, Althoff RR, Pine DS. A dimensional approach to developmental psychopathology. *International Journal of Methods in Psychiatric Research*. 2007; 16:S16–S23. [PubMed: 17623391]
- Hudziak JJ, Heath AC, Madden PF, Reich W, Bucholz KK, Slutske W, et al. Latent class and factor analysis of DSM–IV ADHD: A twin study of female adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1998; 37:848–857. [PubMed: 9695447]
- Jedidi K, Jagpal HS, DeSarbo WS. STEMM: A general finite mixture structural equation model. *Journal of Classification*. 1997; 14:23–50.
- Kendell, RE. The major functional psychoses: Are they independent entities or part of a continuum? Philosophical and conceptual issues underlying the debate. In: Kerr, A.; McClelland, H., editors. *Concepts of mental disorder: A continuing debate*. Gaskell; London: 1991. p. 1-16.
- Kendler KS, Prescott CA. A population-based twin study of lifetime major depression in men and women. *Archives of General Psychiatry*. 1999; 56(1):39–44. [PubMed: 9892254]
- Lasky-Su J, Neale B, Franke B, Anney R, Zhou K, Chen W, et al. Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *American Journal of Medical Genetics Part B*. 2008; 147B: 1345–1354.
- Lenzenweger MF, McLachlan G, Rubin DB. Resolving the latent structure of schizophrenia endophenotypes using expectation-maximization-based finite mixture modeling. *Journal of Abnormal Psychology*. 2007; 116:16–29. [PubMed: 17324013]
- Lo YT, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. *Biometrika*. 2001; 88:767–778.
- Lubke, GH.; Carey, G.; Lessem, J.; Hewitt, J. Using observed genetic variables to predict latent class membership: A comparison of two methods. Paper presented at the Behavior Genetics Association Conference; Louisville, KY. Jun 25–28. 2008
- Lubke GH, Hudziak JJ, Derks EM, van Bijsterveldt TCEM, Boomsma DI. CBCL attention problems and the relation to DSM diagnoses: Lack of evidence for categorically distinct subtyping. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2009; 48:1085–1093. [PubMed: 19797980]
- Lubke GH, Muthén BO. Investigating population heterogeneity with factor mixture models. *Psychological Methods*. 2005; 10:21–39. [PubMed: 15810867]
- Lubke GH, Muthén BO, Moilanen I, McGough JJ, Loo SK, Swanson JM, et al. Subtypes vs. severity differences in Attention Deficit Hyperactivity Disorder in the Northern Finnish Birth Cohort (NFBC). *Journal of the American Association of Child and Adolescent Psychiatry*. 2007; 46:1584–1593.
- Lubke GH, Neale MC. Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood. *Multivariate Behavioral Research*. 2006; 41:499–532.
- Lubke GH, Neale MC. Distinguishing between latent classes and continuous factors with categorical outcomes: Class-invariance of parameters of factor mixture models. *Multivariate Behavioral Research*. 2008; 43:592–620. [PubMed: 20165736]
- Maraun MD, Slaney K. An analysis of Meehl’s MAXCOV-HITMAX procedure for the case of continuous indicators. *Multivariate Behavioral Research*. 2005; 40:489–518.
- Maraun MD, Slaney K, Goddyn L. An analysis of Meehl’s MAXCOV-HITMAX procedure for the case of dichotomous indicators. *Multivariate Behavioral Research*. 2003; 38:81–112.
- Marcus DK, Ruscio J, Lilienfeld SO, Hughes KT. Converging evidence for the latent structure of antisocial personality disorder: Consistency of taxometric and latent class analyses. *Criminal Justice and Behavior*. 2008; 35:284–293.
- McLachlan, G.; Peel, D. *Finite mixture models*. Wiley-Interscience; New York: 2000.

- Meehl, PE. MAXCOV-HITMAX: A taxonomic search method for loose genetic syndromes. In: Meehl, PE., editor. *Psychodiagnosis: Selected Papers*. University of Minnesota Press; Minneapolis: 1973. p. 200-224.
- Meehl PE. Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality*. 1992; 60:117–174.
- Meehl PE. Bootstraps taxometrics—Solving the classification problem in psychopathology. *American Psychologist*. 1995; 50:266–275. [PubMed: 7733538]
- Meehl PE, Yonce LJ. Taxometric analysis: I. Detecting taxonicity with 2 quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*. 1994; 74:1059–1274.
- Meehl PE, Yonce LJ. Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports*. 1996; 78:1091–1227.
- Muthén BO, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*. 1999; 55:463–469. [PubMed: 11318201]
- Muthén, LK.; Muthén, BO. *Mplus Version 5.0* [Computer program]. Muthén & Muthén; Los Angeles: 2007.
- Nagin DS, Land KC. Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology*. 1993; 31(3):327–362.
- Neale MC, Aggen SH, Maes HH, Kubarych TS, Schmitt JE. Methodological issues in the assessment of substance use phenotypes. *Addictive Behaviors*. 2006; 31(6):1010–1034. [PubMed: 16723188]
- Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*. 2007; 14:535–569.
- Nylund KL, Bellmore A, Nishina A, Graham S. Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development*. 2007; 78(6):1706–1722. [PubMed: 17988316]
- Pickles A, Angold A. Natural categories or fundamental dimensions: On carving nature at the joints and the rearticulation of psychopathology. *Development and Psychopathology*. 2003; 15:529–551. [PubMed: 14582931]
- R Development Core Team. *R: A language and environment for statistical computing*. Author; Vienna, Austria: 2008.
- Rohde LA, Barbosa G, Polanczyk G, Eizirik M, Rasmussen ER, Neuman RJ, et al. Factor and latent class analysis of DSM-IV ADHD symptoms in a school sample of Brazilian adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2001; 40:711–718. [PubMed: 11392350]
- Ruscio, J. Taxometric programs for the R computing environment: User’s manual. 2007. Available at: <http://www.tcnj.edu/~ruscio/taxometrics.html>
- Ruscio, J. [Retrieved October 13, 2007] Taxometrics references. 2008. from <http://www.taxometricmethod.com/TaxometricsReferences.pdf>
- Ruscio J. Assigning cases to groups using taxometric results: An empirical comparison of classification techniques. *Assessment*. 2009; 16:55–70. [PubMed: 18607008]
- Ruscio, J.; Haslam, N.; Ruscio, AM. *Introduction to the taxometric method: A practical guide*. Lawrence Erlbaum Associates, Inc; Mahwah, NJ: 2006.
- Ruscio J, Marcus DK. Detecting small taxa using simulated comparison data: A reanalysis of Beach, Amir, and Bau’s (2005) data. *Psychological Assessment*. 2007; 19:241–246. [PubMed: 17563206]
- Ruscio J, Ruscio AM, Meron M. Applying the bootstrap to taxometric analysis: Generating empirical sampling distributions to help interpret results. *Multivariate Behavioral Research*. 2007; 42:349–386.
- Rutter M, Shaffer D. DSM-III: A step forward or back in terms of the classification of child psychiatric disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1980; 19:371–394.
- Schwarz G. Estimating the dimension of a model. *Annals of Statistics*. 1978; 6(2):461–464.

- Steinley D. Properties of the Hubert–Arabie Adjusted Rand Index. *Psychological Methods*. 2004; 9:386–396. [PubMed: 15355155]
- Thain D, Tannenbaum T, Livny M. Distributed computing in practice: The Condor experience. *Concurrency and Computation: Practice & Experience*. 2005; 17:323–356.
- Tueller SJ, Lubke GH. Evaluation of structural equation mixture models in a cross-sectional setting: Parameter estimates and correct class assignment. *Structural Equation Modeling*. 2010; 17(2):165–192. [PubMed: 20582328]
- Tueller SJ, Lubke GH. Estimating classification accuracy for latent variable mixture models (under review). 2010
- Vermunt JK, Magidson J. Latent class models for classification. *Computational Statistics and Data Analysis*. 2003; 41:531–537.
- Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*. 1989; 52(2):307–333.
- Waller, NG.; Meehl, PE. *Multivariate taxometric procedures: Distinguishing types from continua*. Sage; Thousand Oaks, CA: 1998.
- Wilson M. DSM–III and the transformation of American psychiatry: A history. *American Journal of Psychiatry*. 1993; 150:399–410. [PubMed: 8434655]
- Yung YF. Finite mixtures in confirmatory factor-analysis models. *Psychometrika*. 1997; 62:297–330.

TABLE 1

Mahalanobis Distances

No. of Indicators	r = 0	r = .2	r = .4	r = .6	r = .8
<i>d</i> = 0.6					
4	1.20	0.95	0.81	0.72	0.65
6	1.47	1.04	0.85	0.73	0.66
8	1.70	1.10	0.87	0.74	0.66
10	1.90	1.13	0.88	0.75	0.66
12	2.08	1.16	0.89	0.75	0.66
<i>d</i> = 1.2					
4	2.40	1.90	1.62	1.43	1.30
6	2.94	2.08	1.70	1.47	1.31
8	3.39	2.19	1.74	1.49	1.32
10	3.79	2.27	1.77	1.50	1.33
12	4.16	2.32	1.79	1.51	1.33
<i>d</i> = 1.8					
4	3.60	2.85	2.43	2.15	1.95
6	4.41	3.12	2.55	2.20	1.97
8	5.09	3.29	2.61	2.23	1.98
10	5.69	3.40	2.65	2.25	1.99
12	6.24	3.49	2.68	2.26	1.99

Note. Mahalanobis distances (MD) at various levels of Cohen's distance *d* between groups for all individual indicators. Note that the univariate *d* is constant across indicators but the MD increases as the number of indicators increases. When the number of indicators is 1, MD = *d*. As the correlation among all indicators approaches 1, MD approaches *d*.

TABLE 2

Taxometric Analysis of the Depression Data

<i>Procedure</i>	CCFI	Base Rate	\bar{r}_{tax}	\bar{r}_{comp}
MAXCOV	0.34	0.16	0.022	0.124
MAXEIG	0.33	0.15	0.022	0.136
MAMBAC	0.52	0.25	0.080	0.064
L-Mode	0.50	0.41	0.163	0.007

Note. Results of the taxometric analyses of the depression data. The base rate estimate is the mean base rate across all curves for MAXCOV, MAXEIG, and MAMBAC. The base rate estimate for L-Mode was computed using assigned taxon/complement class membership. \bar{r}_{tax} and \bar{r}_{comp} are the mean within-taxon and within-complement class correlations, respectively. CCFI = Comparison Curve Fit Index.

TABLE 3

Fit Indexes for Depression Data Factor Mixture Models

Model	LL	Par	AIC	BIC	saBIC
F1C1	-5,565.02	20	11,170.04	11,279.40	11,215.86
F0C2	-5,748.18	21	11,538.35	11,653.18	11,586.46
F0C3	-5,461.69	32	10,987.38	11,162.36	11,060.69
F0C4	-5,427.51	43	10,941.02	11,176.14	11,039.53
F0C5	-5,410.08	54	10,928.16	11,223.43	11,051.88
F0C6	-5,390.50	65	10,911.00	11,266.42	11,059.92
F1C2MI	-5,453.19	23	10,952.39	11,078.15	11,005.08
F1C3MI	-5,447.76	26	10,947.52	11,089.69	11,007.09
F1C4MI	-5,447.29	29	10,952.57	11,111.14	11,019.01
F1C2MnI τ	-5,417.81	32	10,899.61	11,074.59	10,972.93
F1C3MnI τ	-5,391.02	44	10,870.03	11,110.62	10,970.84
F1C4MnI τ	-5,377.63	56	10,867.26	11,173.47	10,995.56
F1C2MnI $\tau\lambda$	-5,415.97	41	10,913.94	11,138.13	11,007.87
F1C3MnI $\tau\lambda$	-5,374.70	62	10,873.40	11,212.41	11,015.44
F1C4MnI $\tau\lambda$	-5,352.00	83	10,870.00	11,323.83	11,060.15
F1C2fix τ	-5,427.77	32	10,919.54	11,094.52	10,992.86

Note. LL = log-likelihood; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion, saBIC = sample size adjusted Bayesian Information Criterion. Models are denoted as F_iC_j where i indicates the number of factors and j the number of classes (e.g., F0C2 indicates a two-class latent class model). MI = measurement invariant; MnI = measurement noninvariant with noninvariance in the thresholds τ or thresholds and factor loadings λ . The F1C2fix τ is a model with thresholds fixed to large values in one class to model high probabilities of zero responses in that class. Values in bold indicate the best fitting model according the corresponding index. The F1C2fix τ model is bold for comparison.

TABLE 4Population Cohen's d Values for Each Simulation Condition

	MD = 1.5	MD = 3.0
LCA5	0.67	1.34
LCA10	0.47	0.95
1F	1.18	2.35
2Fcl	1.09	2.18
2Fss	1.07	2.13

Note. d does not change between balanced and imbalanced class proportion conditions. Table entries are averages across all items within condition. MD = Mahalanobis distance; LCA5 = latent class analysis model with 5 indicators; LCA10 = latent class analysis model with 10 indicators; 1F = one-factor factor mixture model; 2Fcl = two-factor factor mixture model with cross-loadings; 2Fss = two-factor factor mixture model with simple structure.

TABLE 5

Convergence Rates for Latent Class Analysis and Factor Mixture Modeling Data

	MD = 1.5	MD = 3.0
$\pi_1 = .5$		
LCA5	1.00	1.00
LCA10	0.99	1.00
1F	1.00	1.00
2Fcl	0.99	1.00
2Fss	0.99	1.00
$\pi_1 = .95$		
LCA5	1.00	1.00
LCA10	0.94	1.00
1F	0.99	1.00
2Fcl	0.40	0.84
2Fss	0.39	0.85

Note. Convergence rates for the latent class analysis and factor mixture modeling data generating models. MD = Mahalanobis distance; LCA5 = latent class data with 5 indicators; LCA10 = latent class data with 10 indicators; 1F = one-factor data with 10 indicators; 2Fcl = two-factor data with cross-loadings with 10 indicators; 2Fss = two-factor data with simple structure with 10 indicators; π_1 = proportion of subjects in the first class where $\pi_2 = 1 - \pi_1$. In each condition, 100 data sets are generated.

TABLE 6

Proportions of Data Sets Correctly Detecting Two Latent Classes

	MD = 1.5		MD = 3.0	
	TAX	FMM	TAX	FMM
$\pi_1 = .5$				
LCA5	0.34	1.00	0.96	1.00
LCA10	0.03	1.00	1.00	1.00
1F	0.04	0.09	0.91	1.00
2Fcl	0.03	0.99	0.98	1.00
2Fss	0.05	0.99	1.00	1.00
$\pi_1 = .95$				
LCA5	0.32	1.00	0.89	1.00
LCA10	0.01	1.00	0.67	1.00
1F	0.01	0.16	0.74	0.99
2Fcl	0.04	0.20	0.75	1.00
2Fss	0.03	0.39	0.97	0.85

Note. Proportions of data sets for which taxometric procedures (TAX) and the factor mixture model (FMM) correctly detected two classes. For the FMM, table entries are conditional on (a) convergence of the data generating model and (b) selection of the data generating model using the Bayesian Information Criterion. MD = Mahalanobis distance; LCA5 = latent class data with 5 indicators; LCA10 = latent class data with 10 indicators; 1F = one-factor data with 10 indicators; 2Fcl = two-factor data with cross-loadings with 10 indicators; 2Fss = two-factor data with simple structure with 10 indicators; π_1 = proportion of subjects in the first class where $\pi_2 = 1 - \pi_1$.

TABLE 7

Sensitivity

	<u>MD = 1.5</u>		<u>MD = 3.0</u>	
	TAX	FMM	TAX	FMM
$\pi_1 = .5$				
LCA5	0.69	0.74	0.75	0.93
LCA10	0.75	0.75	0.89	0.94
1F	0.71	0.74	0.93	0.94
2Fcl	0.74	0.79	0.93	0.93
2Fss	0.75	0.79	0.93	0.93
$\pi_1 = .95$				
LCA5	0.56	0.28	0.74	0.69
LCA10	0.86	0.21	0.84	0.68
1F	0.63	0.32	0.92	0.69
2Fcl	0.65	0.20	0.93	0.69
2Fss	0.69	0.30	0.95	0.70

Note. Average sensitivity of taxometric procedures (TAX) and the factor mixture model (FMM). Sensitivity is defined as the proportion of subjects in the second (minority) class correctly assigned to the second class. Table entries are conditional on the detection rates of the two-class structure as summarized in Table 6, which are close to zero for taxometric procedures for some conditions. Results based on fewer than six data sets are presented in italics. MD = Mahalanobis distance; LCA5 = latent class data with 5 indicators; LCA10 = latent class data with 10 indicators; 1F = one-factor data with 10 indicators; 2Fcl = two-factor data with cross-loadings with 10 indicators; 2Fss = two-factor data with simple structure with 10 indicators; π_1 = proportion of subjects in the first class where $\pi_2 = 1 - \pi_1$.

TABLE 8

Specificity

	<u>MD = 1.5</u>		<u>MD = 3.0</u>	
	TAX	FMM	TAX	FMM
$\pi_1 = .5$				
LCA5	0.72	0.71	0.74	0.93
LCA10	<i>0.69</i>	<i>0.69</i>	0.86	0.92
1F	<i>0.73</i>	0.77	0.93	0.93
2Fcl	<i>0.74</i>	0.68	0.93	0.93
2Fss	<i>0.75</i>	0.70	0.93	0.93
$\pi_1 = .95$				
LCA5	0.60	0.85	0.75	0.99
LCA10	<i>0.48</i>	0.89	0.86	0.99
1F	<i>0.78</i>	0.97	0.94	0.99
2Fcl	<i>0.76</i>	0.95	0.93	0.99
2Fss	<i>0.75</i>	0.92	0.91	0.99

Note. Average specificity of taxometric procedures (TAX) and the factor mixture model (FMM). Specificity is defined as the proportion of subjects in the first (majority) class correctly assigned to the first class. Table entries are conditional on the detection rates of the two-class structure as summarized in Table 6, which are close to zero for taxometric procedures for some conditions. Results based on fewer than six data sets are presented in italics. MD = Mahalanobis distance; LCA5 = latent class data with 5 indicators; LCA10 = latent class data with 10 indicators; 1F = one-factor data with 10 indicators; 2Fcl = two-factor data with cross-loadings with 10 indicators; 2Fss = two-factor data with simple structure with 10 indicators; π_1 = proportion of subjects in the first class where $\pi_2 = 1 - \pi_1$.

TABLE 9

Hubert-Arabie Adjusted Rand Index

	<u>MD = 1.5</u>		<u>MD = 3.0</u>	
	TAX	FMM	TAX	FMM
$\pi_1 = .5$				
LCA5	0.17	0.22	0.24	0.74
LCA10	<i>0.19</i>	0.20	0.56	0.74
1F	<i>0.19</i>	0.24	0.74	0.75
2Fcl	<i>0.24</i>	0.25	0.74	0.74
2Fss	<i>0.25</i>	0.25	0.74	0.74
$\pi_1 = .05$				
LCA5	0.02	0.09	0.10	0.70
LCA10	<i>0.00</i>	0.05	0.09	0.57
1F	<i>0.11</i>	0.17	0.55	0.71
2Fcl	<i>0.10</i>	0.16	0.51	0.71
2Fss	<i>0.09</i>	0.19	0.45	0.71

Note. Table entries are the Hubert–Arabie Adjusted Rand Index, a chance-corrected measure of correct class assignment. Table entries are conditional on the detection rates of the two-class structure as summarized in Table 6, which are close to zero for taxometric procedures for some conditions. Results based on fewer than six data sets are presented in italics. MD = Mahalanobis distance; TAX = taxometric procedures; FMM = factor mixture model; LCA5 = latent class data with 5 indicators; LCA10 = latent class data with 10 indicators; 1F = one-factor data with 10 indicators; 2Fcl = two-factor data with cross-loadings with 10 indicators; 2Fss = two-factor data with simple structure with 10 indicators; π_1 = proportion of subjects in the first class where $\pi_2 = 1 - \pi_1$.