MULTIVARIATE BEHAVIORAL RESEARCH, *41*(4), 499–532

Distinguishing Between Latent Classes and Continuous Factors: Resolution by Maximum Likelihood?

Gitta Lubke
University of Notre Dame

Michael C. Neale
Virginia Commonwealth University

Latent variable models exist with continuous, categorical, or both types of latent variables. The role of latent variables is to account for systematic patterns in the observed responses. This article has two goals: (a) to establish whether, based on observed responses, it can be decided that an underlying latent variable is continuous or categorical, and (b) to quantify the effect of sample size and class proportions on making this distinction. Latent variable models with categorical, continuous, or both types of latent variables are fitted to simulated data generated under different types of latent variable models. If an analysis is restricted to fitting continuous latent variable models assuming a homogeneous population and data stem from a heterogeneous population, overextraction of factors may occur. Similarly, if an analysis is restricted to fitting latent class models, overextraction of classes may occur if covariation between observed variables is due to continuous factors. For the data-generating models used in this study, comparing the fit of different exploratory factor mixture models usually allows one to distinguish correctly between categorical and/or continuous latent variables. Correct model choice depends on class separation and within-class sample size.

Starting with the introduction of factor analysis by Spearman (1904), different types of latent variable models have been developed in various areas of the social sciences. Apart from proposed estimation methods, the most obvious differences between these early latent variable models concern the assumed distribution of the

Correspondence concerning this article should be addressed to Gitta H. Lubke, Department of Psychology, 118 Haggar Hall, University of Notre Dame, Notre Dame IN 46556. E-mail: glubke@nd.edu

observed and latent variables. Bartholomew and Knott (1999) and Heinen (1993) classified models in a 2×2 table according to whether the observed and latent variables are categorical or continuous. Models with categorical latent variables encompass the classic latent class model, which has categorical observed variables, and the latent profile model (LPM), which has continuous observed variables (Heinen, 1993; Lazarsfeld & Henry, 1968). Models with continuous latent variables include latent trait models, factor models, and structural equation models. Attempts have been made to show similarities between these models, and to propose general frameworks that encompass the different models as submodels (Bartholomew & Knott, 1999; Heinen, 1996; Langenheine & Rost, 1988; Lazarsfeld & Henry, 1968).

A common characteristic of categorical and continuous latent variable models is that both are designed to explain the covariances between observed variables. More precisely, latent variable models are constructed such that the observed variables are independent conditional on the latent variables. In the context of latent trait and latent class analysis, this characteristic has been called "local independence" (Lazarsfeld & Henry, 1968; Mellenbergh, 1994). In exploratory and confirmatory factor models, local independence corresponds to the assumption of uncorrelated residuals. Consequently, in both continuous and categorical latent variable models, nonzero covariances between observed variables are due to differences between participants with respect to the latent variables. The general question addressed in this article is whether, based on observed data, it is possible to distinguish between categorical and continuous latent variables.

The distinction between categorical and continuous *observed* variables refers to the response format of the data, whereas the distinction between categorical and continuous *latent* variables can be of considerable importance on a theoretical level. For example, whether individuals with clinical diagnoses of psychiatric disorders differ *qualitatively* or *quantitatively* from those without is a topic of current debate (Hicks, Krueger, Iacono, McGue, & Patrick, 2004; Krueger et al., 2004; Pickles & Angold, 2003). Therefore, a model comparison approach that can reliably discriminate between latent class or LPMs, on one hand, and factor models, on the other hand, would have great practical value.

Although the difference between categorical and continuous latent variables can be significant on a conceptual level, the distinction is less clear on a statistical level. As shown in Bartholomew (1987), the model implied covariances of a K-class latent class model for continuous observed variables are structurally equivalent to the model implied covariances of a $K - 1$ factor model for continuous observed variables (see also Molenaar & van Eye, 1994, and more recently, Bauer & Curran, 2004; Meredith & Horn, 2001). Consequently, in an analysis of covariances (or correlations), a K-class model and a $K - 1$ factor model fit equally well, regardless of whether the true data-generating model is a K-class model or a $K - 1$ factor model. This fact has serious consequences when making assumptions regarding the distribution of the data. First, assuming population heterogeneity and restricting an analysis to latent class models may result in an overextraction of classes if the sample is homogeneous and covariances of observed variables are due to underlying continuous factors (see Bauer & Curran, 2004, for some interesting examples). Second, assuming homogeneity of the sample, and conducting an exploratory factor analysis might result in overextraction of factors if the sample is heterogenous. In fact, if a population consists of, say, six classes, an exploratory factor analysis of the pooled covariance (or correlation) matrix derived from a reasonably large sample will likely result in a nicely fitting five-factor solution (see Molenaar & van Eye, 1994, for an illustration). Given that exploratory factor analysis is more popular than latent class analysis, overextraction of factors may be a more common mistake than overextraction of classes.

Exploratory and confirmatory factor analyses can be carried out using covariance matrices of the observed variables because these are sufficient statistics given the usual assumptions, namely, homogeneity of the sample, multivariate normality of the underlying factors, normality of residuals, independence of factors and residuals, zero autocorrelations of residuals, and linear relations between observed variables and factors. Under these assumptions, the resulting distribution of the observed variables is multivariate normal, and has therefore zero skewness and kurtosis. Latent class models, on the other hand, are fitted to raw data, because class membership is not observed, and therefore within-class covariance matrices and mean vectors are not available. The distribution of observed variables in a heterogenous sample consisting of several latent classes is a mixture distribution. Skewness, kurtosis, and other higher order moments deviate from zero as differences between classes (e.g., differences in means, variances, and covariances) increase. Raw data contain information concerning all higher order moments of the joint distribution of the observed data. Therefore, a comparison of latent class models and exploratory factor models based on maximum likelihood analysis of the raw data should in principle reveal the correct model. The likelihood of the data should be larger under the correctly specified distribution than under an incorrectly specified distribution. However, in an actual analysis of empirical data, it is not clear whether the distinction can be made with acceptable accuracy.

Recently, general models have been introduced that include both continuous and categorical latent variables and that can handle both categorical and continuous observed variables (Arminger, Stein, & Wittenberg, 1999; Dolan & van der Maas, 1998; Heinen, 1996; Jedidi, Jagpal, & DeSarbo, 1997; Muthén & Shedden, 1999; Vermunt & Magidson, 2003; Yung, 1997). These general models have been introduced under various names and are referred to as factor mixture models in the remainder of this article. Factor mixture models allow the specification of two or more classes, and within-class structures that can range from local independence to complex structural relations between latent continuous variables. These models have the potential to estimate structural relations in the presence of population heterogeneity, and to compare different subpopulations with respect to the parameters of the within-class distribution. Using factor mixture models in an analysis of empirical data represents an additional challenge regarding the distinction between latent categorical and latent continuous variables. Specifically, the task is to determine simultaneously the number of latent classes *and* the dimensionality of the latent factor space within each class.

Factor mixture models are based on several assumptions including multivariate normality within class. The joint distribution of observed variables is a mixture distributions where the mixture components represent the latent classes. It should be noted that mixture distributions have two main areas of applications: One is to use mixture components to model clusters in a population, whereas the other is to use mixture components to approximate non-normal distributions (McLachlan & Peel, 2000). When using factor mixture models, one should be aware of the fact that deviations from the assumed multivariate normality within class can lead to overextraction of classes. This fact has been demonstrated by Bauer and Curran (2003a) in the context of factor mixture models for longitudinal data (see also McLachlan & Peel, 2000, for some classic examples).

In this article, we focus on the recovery of the correct within-class factor structure in case of possible population heterogeneity. Mixture components are thought to represent clusters of subjects. Our approach is to compare models in an analysis that is exploratory with respect to *both* the within-class structure *and* the number of latent classes. This article addresses the question whether model comparisons lead to a correct model choice regarding the nature of the latent variables (continuous, categorical, or both), and their dimensionality (number of classes and number of factors) when model assumptions are not violated. Artificial data are generated under different models including (a) latent class models with two and three classes, (b) single- and two-factor models where the population is homogenous (as in conventional factor models), and (c) single- and two-factor models where the population consists of two latent classes. Correct and incorrect models are fitted to the different types of data and compared with respect to model fit. The accuracy of model selection is likely to depend on the separation between classes when the population is heterogeneous (Lubke & Muthén, 2007; Yung, 1997). Therefore, data are drawn from heterogeneous populations with different degrees of separation. Model parameters such as intercepts and factor loadings are class invariant or class specific. Class proportions are equal or unequal. In addition, sample sizes are varied for a subset of the generated data to investigate minimum sample size requirements for accurate model selection.

MODELS

General models that encompass continuous and categorical latent variables can be conceptualized as generalizations of classic latent class models. In classic latent class models, observed variables within class are assumed to be independent. The generalization consists of the possibility to impose more complicated structures on observed variables within class, such as factor models. Various types of these generalized latent class models have been described elsewhere (Arminger et al., 1999; Dolan & van der Maas, 1998; Heinen, 1996; Jedidi et al., 1997; Muthén & Shedden, 1999; Vermunt & Magidson, 2005; Yung, 1997).[1] Here, we follow an approach comparable to Bartholomew and Knott (1999) and start with the joint distribution of observed variables **Y** in the total population. The joint distribution is a mixture distribution where each mixture component corresponds to the distribution of observed variables within a particular latent class. The joint probability distribution of the data is a weighted sum of the probability distributions of the components. The weights are the proportions of the latent classes in the total population, and sum to unity. The mixture with $k = 1,\ldots,K$ components can be denoted as

$$f(\mathbf{y}) = \sum_{k=1}^{K} \pi_k f(\mathbf{y}_k), \tag{1}$$

where π_k is the proportion of class k and $f(\cdot)$ is a probability distribution. The within-class distribution $f(\mathbf{y}_k)$ can be parameterized in different ways to derive specific submodels.

Latent Profile Model

In the LPM, observed variables are assumed to be continuous, and independent conditional on class membership. Specifically, $\mathbf{Y}_k \sim N(\boldsymbol{\mu}_k, \Sigma_k)$ where the within-class mean vectors $\boldsymbol{\mu}_k$ equal the observed within-class means, and where the within-class covariance matrices $\boldsymbol{\Sigma}_k$ are diagonal. That is, within each class, the observed variables are uncorrelated.

Exploratory Factor Mixture Models

The exploratory factor model for a single homogeneous population is derived by letting $K = 1$. Hence, there is a single within-class distribution $f(\mathbf{y}_k)$ and the subscript k can be omitted. The observed variables **Y** are normally distributed with $\mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma)$. In the exploratory factor model the mean vector $\boldsymbol{\mu}$ is unstructured such that each observed variable has a single unique free parameter for its mean, and the model implied covariance matrix is structured as

[1]In a less technical article, Lubke and Muthén (2005) illlustrated how the latent class model can be extended stepwise to derive a factor mixture model.

$$\mathbf{\Sigma} = \mathbf{\Lambda \Psi \Lambda}^{\mathrm{t}} + \mathbf{\Theta}. \tag{2}$$

The factor loading matrix $\mathbf{\Lambda}$ has dimension $J \times L$ where J is the number of observed variables and L is the number of factors. We assume uncorrelated normally distributed factors with unit variance such that the factor covariance matrix $\mathbf{\Psi}$ is an identity matrix. Some restrictions on $\mathbf{\Lambda}$ are necessary for reasons of identification (Lawley & Maxwell, 1971). The residuals are assumed to be uncorrelated with the factors, and normally distributed with zero means and diagonal covariance matrix $\mathbf{\Theta}$.

By letting $k = 1,\ldots, K$, exploratory factor mixture models can be specified for K classes. As before, the within-class mean vectors $\boldsymbol{\mu}_k$ equal the observed within-class means, and the within-class covariance matrices $\mathbf{\Sigma}_k$ are structured as in Equation 2. Factor loadings and residual variances may differ across classes.

Confirmatory factor models for a homogeneous population with $K = 1$ classes, or for a heterogeneous population with $k = 1,\ldots, K$ classes can be derived by imposing restrictions on the within-class factor loading matrices. Confirmatory factor models are not further considered in this article.

SIMULATION STUDY DESIGN

The simulation study presented in this article is designed to investigate the merits of exploratory factor mixture analysis. Exploratory factor mixture analysis involves fitting a series of different mixture models, and has the potential to recover the within-class factor structure in the presence of possible population heterogeneity. In this study, mixture models are fitted with an increasing number of classes. Within-class, *exploratory* factor models are specified with an increasing number of factors. Note that the within-class model of the LPM can be conceptualized as a zero-factor model, or as an L-factor model where the factors have zero variance. LPMs impose the most restrictive structure within class in this study (i.e., local independence within class). Increasing the number of factors within class relaxes this restriction. In this study, the dimension of the within-class factor space, and the number of classes of the true models used for data generation is kept low.

Exploratory factor models are lenient because all observed variables have loadings on all factors, but they are not entirely unstructured. Because restrictions on the within-class distribution can result in overextraction of latent classes (Bauer & Curran, 2004), it might be preferable to start an analysis by fitting mixtures with unstructured within-class mean vectors and covariance matrices. However, such an approach would be impractical for questionnaire data with a large number of observed variables. Estimation may be problematic due to the fact that the number of parameters to be estimated increases rapidly with increasing numbers of classes. Furthermore, the likelihood surface of a (normal) mixture with class-specific covariance matrices often has many singularities (McLachlan & Peel, 2000). Indeed, in an initial pilot simulation we found that fitting unrestricted within-class models led to high nonconvergence rates (e.g., more than 70%) for data with more than eight observed variables.

The simulation study has two parts, which are organized in a similar way. The general approach consists of fitting a set of exploratory factor mixture models to artificial data to investigate whether comparing the fit of different models results in the correct model choice. In Part 1, detection of the correct model is investigated for different types of data-generating models, for different class separations in the true data, and for class-specific versus class-invariant intercepts and loadings. In Part 2, detection of the correct model is investigated for varying sample sizes and for unequal class proportions.

Throughout, data are generated without violating assumptions regarding the within-class factor models (viz., within-class homogeneity, multivariate normality of the underlying factors, normality of residuals, independence of factors and residuals, zero autocorrelations of residuals, and linear relations between observed variables and factors). It should be noted that, for example, deviations from normality within-class can lead to overextraction of classes. This fact has been demonstrated in a simulation study concerning confirmatory factor mixture models (e.g., growth mixture models) by Bauer and Curran (2003a) and has been discussed by Bauer and Curran (2003b), Cudeck and Henly (2003), Muthén (2003), and Rindskopf (2003). Our article represents a first step to evaluate the potential of exploratory factor mixture models and focuses on the possibility to discriminate between relatively simple models that differ with respect to the distribution of latent variables and some other aspects such as class-specific or invariant model parameters, class proportions, and sample size. Although beyond the scope of this article, extending our study presented here to include different possible violations of underlying model assumptions is of clear interest for future research.

Data Generation

For both parts of the simulation, 100 data sets are generated, which differ only with respect to factor scores and/or residual scores drawn from (multivariate) normal distributions and with respect to a score drawn from the uniform distribution, which is used together with the prior class probabilities to assign participants to a given class. The number of observed variables is 10 throughout. The observed variables are continuous, and multivariate normally distributed conditional on class membership. A detailed overview of all within-class parameter values used for the data generation can be found in the Appendix.

For the first part of the study, data are generated under nine different models. The first two models are LPMs with two and three classes, abbreviated LPMc2 and LPMc3, respectively. Models 3, 4, and 5 are factor models for a single population, namely, a single-factor/single-class model, denoted as F1c1; a two-factor/single-class model with simple structure (F2c1SS); and a two-factor/single-class model with cross-loadings (F2c1CL). Models 6, 7, and 8 are single-factor/two-class models. Specifically, F1c2MI is measurement invariant across classes (i.e, intercepts, factor loadings, and residual variances of the within-class factor model are invariant across classes), F1c2nMI1 has class-specific intercepts (i.e., νk in Equation 3), and F1c2nMI2 has noninvariant intercepts and factor loadings. The latter two models are not measurement invariant (Lubke, Dolan, Kelderman, & Mellenbergh, 2003; Meredith, 1993; Widaman & Reise, 1997). Finally, Model 9 is a two-factor/two-class model. The factor structure is equal to the two-factor/single-class model.

In this part of the study, the number of participants within class is approximately 200 for all models. Consequently, models with more than one class have a higher total number of participants. This choice is made because the emphasis of the simulation is on detecting the correct within-class parameterization. The prior class probabilities in the first part of the study are .5 for two-class models and .333 for the three-class model. The distance between the classes is varied such that the impact of class separation on model selection can be evaluated. The separation between classes for the two-class models is a separation of either 1.5 or 3 as measured by the multivariate Mahalanobis distance.[2] The only three-class model in the set of data-generating models is the three-class LPM. The Mahalanobis distance between consecutive classes equals 1.5 for the small class separation and 3 for the large separation. The difference between Classes 1 and 3 equals 3.0 and 6.0 for small and large separation, respectively.

The second part of the study focuses on different sample sizes and unequal class proportions. Data are generated under the single-factor/two-class model with intercept differences (i.e., F1c2nMI). While keeping the prior class probabilities at .5, the within-class sample sizes are 25, 50, 75, and 1000. While keeping the within-class sample size at 200, prior class probabilities are set to .9 and .1, respectively.

Model Fitting

A set of models were fitted to each type of data. For each fitted model, 50 sets of random starting values are provided, and 10 initial iterations are computed for these 50 sets. The 10 sets of starting values that result in the highest log likelihood are then iterated until convergence, and the best solution is chosen as the final result. Each set of fitted models comprises models with an increasing number of within-class factors and increasing number of classes, including LPMs with an increasing number of classes. The number of factors within class and the number of classes is increased until the higher dimensional model is rejected in favor of the lower dimensional model. Factor variances of fitted exploratory factor models are

[2]The Mahalanobis distance between two classes equals $M = (\mu_1 - \mu_2)^t\Sigma^{-1}(\mu_1 - \mu_2)$.

fixed to unity, and the matrix of factor loadings is constrained such that factor indeterminacy within class is avoided. Note that all fitted models are exploratory in the sense that model parameters such as loadings, intercepts, and residual variances (if applicable) are class specific. The criteria used for model comparisons are the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the sample size adjusted BIC (saBIC), the Consistent AIC (CAIC), and the adjusted likelihood ratio test (aLRT) statistic (Akaike, 1974, 1987; Bozdogan, 1987; Lo, Mendell, & Rubin, 2001; Schwarz, 1978). The formulae for the information criteria can be found in the Appendix. Information criteria differ in the way they correct for sample size and the number of free parameters of the fitted model. The number of free parameters of the models fitted in this study are shown in Table 1.

For each of the data-generating models, we provide the percentage of Monte Carlo (MC) replications for which the correct model is the first or second choice based on the information criteria AIC, BIC, CAIC, and saBIC, and the aLRT. The percentages of correct model selection, however, do not reveal which models are competing models, or how "close" competing models are with respect to their fit indices. For this reason, for each of the fitted models, we also compute fit indices averaged over the 100 MC replications. Regarding the average fit indices, the BIC and CAIC never lead to diverging results. The CAIC is therefore omitted from the tables reporting averages, but can be easily calculated by adding the number of free parameters (see Table 1) to the average BIC.

In addition, parameter estimates are evaluated that are informative with respect to model choice. For instance, factor correlations approaching unity may be regarded as an indication of a single underlying dimension. Similarly, very small factor loadings may indicate absence of a common underlying factor. Models are fitted using an extended version of the RUNALL utility designed for MC simulations with Mplus. The original Runall utility is available at www.statmodel.com/ runutil.html.

RESULTS

Part 1: Identifying the Correct Model

Convergence rates were above 95% if not otherwise mentioned. Results of the first part of the study are presented separately for each of the nine data-generating mod-

TABLE 1
Number of Free Parameters for All Fitted Models

F1C1	*F2C1*	*F3C1*	*F1C2*	*F2C2*	*LPAc2*	*LPAc3*	*LPAc4*
30	39	47	51	69	41	62	83

Note. For the two-class factor models, intercept differences between classes are estimated. Residual variances, factor loadings, and factor correlations (if part of the model) are estimated for each class.

els. The results of the nine data-generating models are summarized each in two tables. The first of the two tables shows the proportion of times the correct model is the first or second choice with respect to the likelihood value and different information criteria. In this table we also show the proportion of MC replications for which the aLRT would lead to a correct test result. Note that more restrictive models may emerge as first choice with respect to the likelihood value even though less restrictive, nested models were part of the set of fitted model. This is due to nonconvergence of the less restrictive models (see, e.g., Tables 2 and 6). The second, larger table shows the average fit measures corresponding to each of the fitted models. Informative average parameter estimates are discussed in the accompanying text. Results of the second part of the study are presented in a similar fashion.

Latent Profile Model Two-Class Data (Tables 2 and 3)

Seven different models were fitted to the LPM two-class data. These are LPMs with one, two, or three classes; exploratory factor models for a single class with one, two or three factors; and an exploratory factor model for two classes with a single factor. The results are shown in Tables 2 and 3.

Small class separation. Table 3 shows that if the analysis had been restricted to a latent profile analysis (LPA), the correct model would have been chosen based on most indices. Only the AIC favors the three-class LPM, although the necessity of the third class is rejected by the aLRT. If the analysis had been restricted to an exploratory factor analysis for a homogeneous population, the results are less clear. When comparing the corresponding models (i.e., exploratory factor models for a single class), it can be seen that the AIC favors the three-factor/single-class model (F3C1). BIC and adjusted BIC (aBIC) favor the single-factor/single-class model (F1C1). Given the fact that the true model has zero factors, at least one artificial factor would be extracted.

The fact that the aBIC and the BIC favor the single-factor/single-class model (F1C1) when comparing the full set of fitted models highlights the difficulty in discriminating between LPMs and factor models. Table 2 shows that in an overall comparison of all fitted models, the proportion of correct model choice is very low. Even as the second choice, the proportions for the different fit indices do not rise to acceptable levels. Recall that the first- and second-order moments of the F1C1 are structurally equivalent to the true two-class LPM (Bartholomew, 1987). Raw data contain information concerning higher order moments. However, in the case of small class separation, the higher order moments of the data may not deviate sufficiently from the zero values that are expected if data were generated under the single-factor/single-class model.

Compelling evidence against the factor models, however, is provided by the parameter estimates. The factor loadings of all factor models were small, and the fac-

TABLE 2
Data Generated Under the Two-Class Latent Profile Model: Proportions of Monte Carlo Replications the Correct Model Is First and Second Choice Using Information Criteria, and Proportion the aLRT Indicates the Necessity of the Second Class When Fitting the Correct Model

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
Small class separation						
First choice	0.06	0.11	0.00	0.14	0.00	0.97
Second choice	0.52	0.34	0.60	0.50	0.33	0.00
Larger class separation						
First choice	0.06	0.12	0.97	0.77	0.97	1.00
Second choice	0.56	0.73	0.00	0.2	0.00	0.00

Note. A proportion of .11 indicates that when fitting the correct model the corresponding information criterion had a minimum value 11% of the time when comparing all fitted models. The last column shows the proportion of Monte Carlo replications for which the aLRT provided the correct test result. aLRT = adjusted likelihood ratio test; logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC.

TABLE 3
Averaged Fit Statistics for 200 Data Sets Generated by a Two-Class Latent Profile Model (LPM), Fitting Latent Profile Models LPMc*j* With *J* Classes and Latent Factor Models F*i*c*j* With *I* Factors and *J* Classes

	logL-val	*AIC*	*BIC*	*saBIC*	*aLRT*
Small class separation					
LPMc1	–4559.22	9158.45	9238.27	9174.81	NA
LPMc2	**–4484.82**	**9051.63**	**9215.28**	**9085.19**	**0.01**
LPMc3	–4458.66	**9041.32**	9288.79	9092.06	0.47
F1c1	–4496.85	9053.70	**9173.45**	**9078.25**	NA
F2c1	–4493.23	9064.46	9220.13	9096.38	NA
F3c1	–4476.24	9046.48	9234.07	9084.94	NA
F1c2	–4452.22	**9026.45**	9269.93	9076.37	0.60
Larger class separation					
LPMc1	–5100.32	10240.63	10320.46	10257.00	NA
LPMc2	**–4599.16**	**9280.33**	**9443.98**	**9313.88**	**0.00**
LPMc3	–4573.00	**9270.00**	9517.47	9320.74	0.43
F1c1	–4692.42	9444.84	9564.59	9469.40	NA
F2c1	–4682.15	9442.30	9597.96	9474.21	NA
F3c1	–4685.49	9464.99	9652.59	9503.45	NA
F1c2	–4577.46	**9276.91**	9520.39	9326.83	0.06

Note. Lower information criteria correspond to better fitting models. A significant aLRT indicates that the model with one class less fits significantly worse; presented are the average *p* values. Results for the true model are in bold, as are the values of fit indices that favor misspecified models. logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; aLRT = adjusted likelihood ratio test.

tor(s) explained only a small percentage of the variance of the observed variables. For instance, for the fitted single-factor/single-class model, the observed variable R-square averaged over the 100 MC replications ranged between .02 and .23 for the 10 items, with an average standard deviation of .04. Therefore, when comparing the entire set of fitted models, and integrating information provided by fit indices and parameter estimates, the correct model without underlying continuous factors within class would be the most likely choice. Furthermore, on average, the factor mixture model with two classes is not supported by the aLRT.

Larger class separation. For larger separation, the correct model is favored by most indices, and the proportion of correct model choice are greater than .95 for BIC and CAIC. The AIC prefers the LPM with three classes, but the necessity of an additional class is again rejected by the aLRT. For large separation, comparing only the exploratory factor models for a single population based on information criteria is inconclusive. The single-factor/two-class model can be regarded as a competing model based on the AIC and the fact that the aLRT does not reject the second class. However, the amount of variation in the observed variables explained by the factors for the single-factor/two-class model is low. The observed variable R-square averaged over the 100 replications for the 10 items for the two classes range between 0.01 and 0.16, with an average standard deviation of 0.03. As for the small separation, the correct model without underlying continuous factors within class would be the logical choice.

When comparing the full set of models, there is no convincing evidence indicating an overextraction of factors. Apparently, with larger separation, higher order moments deviate sufficiently from zero for a clearer distinction between latent class models and factor models.

Latent Profile Model Three-Class Data (Tables 4 and 5)

The fitted models were LPMs with two, three, and four classes; exploratory factor models for a single class with one, two, and three factors; exploratory factor models for two classes with a single and two factors; and an exploratory factor model for three classes with a single factor (see Tables 4 and 5). The four-class LPMs had high nonconvergence (e.g., 43% and 39% for the two distances).

Small class separation. Restricting the analysis to fitting LPMs would result in a correct model choice. As in the case of the LPM two-class data, the AIC points to a model with one additional class, which is rejected by the aLRT. Table 4 shows that adding the proportions of first and second choice of the AIC leads to a proportion similar to the BICs and CAIC. Note that restricting the analysis to exploratory factor models for a single population would result in choosing the single-factor/single-class model (F1C1) based on the information criteria. The observed

TABLE 4
Data Generated Under the Three-Class Latent Profile Model: Proportions of Monte Carlo Replications the Correct Model Is First and Second Choice Using Information Criteria, and Proportion the aLRT Indicates the Necessity of the Second Class When Fitting the Correct Model

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
Small class separation						
First choice	0.43	0.46	0.87	0.86	0.87	1.00
Second choice	0.44	0.41	0.00	0.01	0.00	0.00
Larger class separation						
First choice	0.39	0.40	0.71	0.68	0.71	1.00
Second choice	0.32	0.31	0.00	0.03	0.00	0.00

Note. aLRT = adjusted likelihood ratio test; logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC.

TABLE 5
Averaged Fit Statistics for 200 Data Sets Generated by a Three-Class Latent Profile Model (LPM), Fitting Latent Profile Models LPMc*j* With *j* Classes and Latent Factor Models F*i*c*j* With *i* Factors and *j* Classes

	logL-val	*AIC*	*BIC*	*saBIC*	*aLRT*
Small class separation					
LPMc2	–7026.45	14134.91	14315.18	14185.02	0.00
LPMc3	**–6715.59**	**13555.18**	**13827.79**	**13630.95**	**0.00**
LPMc4	–6693.15	**13552.30**	13917.25	13653.75	0.42
F1c1	–7127.11	14314.21	14446.12	14350.88	NA
F2c1	–7119.31	14316.61	14488.09	14364.28	NA
F3c1	–7116.47	14326.94	14533.59	14384.38	NA
F1c2	–6832.88	13787.76	14055.97	13862.32	0.03
Larger class separation					
LPMc2	–7413.06	14908.12	15088.39	14958.23	0.00
LPMc3	**–6776.34**	**13676.67**	**13949.28**	**13752.45**	0.00
LPMc4	–6756.27	13678.54	14043.49	13779.98	0.42
F1C1	–7308.00	14676.00	14807.91	14712.67	NA
F2C1	–7302.81	14683.61	14855.09	14731.28	NA
F3C1	–7300.56	14695.11	14901.77	14752.56	NA
F1C2	–6981.01	14084.02	14352.23	14158.58	0.02

Note. Results for the true model are in bold, as are the values of fit indices that favor misspecified models. logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC; aLRT = adjusted likelihood ratio test.

R-square for the 10 items averaged over the 100 replications ranges between .26 and .49, with an average standard deviation of 0.03. The explained factor variance is higher than for the two-class LPM data. This difference is most likely due to the following. Although consecutive classes in the three-class data have a separation equivalent to the two-class data, Classes 1 and 3 have twice the distance. Hence, multivariate kurtosis is clearly more pronounced in the three-class data. As classes in the data-generating model are more separated, overextraction of factors becomes a more serious threat if an analysis is restricted to exploratory factor models assuming a homogeneous population. Tables 4 and 5 show that when comparing the whole set of fitted models, the correct three-class LPM would be chosen. Note that the convergence rate of the two-factor/two-class model and the single-factor/three-class model were below 40%, and average results are not presented.

Large class separation. All indices identify the correct model as the best-fitting model. The proportions of correct model choice are slightly lower than for the smaller separation. This is due to the fact that the four-class model is a competing model. However, the fourth class is rejected by the aLRT. Restricting the analysis to exploratory factor model for a homogeneous population would again result in choosing the single-factor/single-class model. The average observed variable R-square corresponding to this fitted model is higher for large separation than for small separation and ranges between .51 and .67, with an average standard deviation of .03. This shows that the probability of overextracting factors may increase with larger class separation.

Single-Factor/Single-Class Data (Tables 6 and 7)

Six different models were fitted to the single-factor/single-class data: LPMs with two and three classes; exploratory single-class models with one, two, or three factors; and an exploratory single-factor/two-class model. The results are shown in Tables 6 and 7.

Comparing only the results for the LPMs, the LPA model with three classes would have been chosen. A closer look at individual output files from the three-class LPA model does not reveal any obvious signs of a possible model misspecification such as inadmissible parameter estimates or extremely small classes. Therefore, the results provide evidence for an overextraction of classes to account for covariances that are due to an underlying continuous dimension *if* the analyses are restricted to an LPA.

When comparing results of all fitted models, the three information criteria consistently reject the LPMs in favor of a factor model. The BIC and CAIC select the correct model 93% of the time. Based on AIC and saBIC, the correct model would be the first choice 33% and 42% of the time, respectively, and the second choice 39% of the time. The reason for the poorer performance of the AIC and the saBIC

TABLE 6
Data Generated Under the Single-Factor/Single-Class Model: Proportions of Monte Carlo Replications the Correct Model Is First and Second Choice Using Information Criteria, and Proportion the aLRT Indicates the Necessity of the Second Class When Fitting the Correct Model

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
First choice	0.03	0.33	0.93	0.42	0.93	1.00
Second choice	0.17	0.39	0.00	0.39	0.00	0.00

Note. aLRT = adjusted likelihood ratio test; logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC.

TABLE 7
Averaged Fit Statistics for 200 Data Sets Generated by a Single-Class/Single-Factor Model, Fitting Latent Profile Models LPMc*j* With *j* Classes and Latent Factor Models F*i*c*j* with *i* Factors and *j* Classes

	logL-val	*AIC*	*BIC*	*saBIC*	*aLRT*
LPAc2	–2691.47	5464.94	5600.17	5470.27	0.02
LPAc3	–2527.07	5178.15	5382.64	5186.22	0.08
F1C1	**–2445.65**	**4951.31**	**5050.26**	**4955.21**	NA
F2C1	–2433.13	**4944.25**	5072.89	**4949.33**	NA
F3C1	–2432.44	4958.87	5113.89	4964.99	NA
F1C2	–2421.10	**4944.19**	5112.41	**4950.83**	0.76

Note. Results for the true model are in bold, as are the values of fit indices that favor misspecified models. logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC; aLRT = adjusted likelihood ratio test.

is that they favor the overparameterized F1C2 and F2C1 models as can be seen in Table 7 showing the averaged fit indices. However, the two-class model is not supported by a nonsignificant aLTR, which indicates that the second class is not necessary. Individual output files show that the factor correlation in the F2C1 model is close to unity. More specifically, the factor correlation averaged over the 100 replications equals .99 with a standard deviation of .17. Taken together, the correct model would most likely have been chosen.

Two-Factor/Single-Class Data With Simple Structure (Tables 8 and 9)

The same models were fitted as for the single-factor/single-class data (see Table 9). As shown previously, if an analysis had been restricted to an LPA, too many classes would have been extracted. A comparison of LPMs based on information criteria would have resulted in choosing the three-class LPM when the population in fact consists of only a single class. The aLRT favored the two-class LPM. However, if exploratory factor mixture models are part of the set of fitted models, the correct model would have been chosen. On average, all indices are in agreement and point to the two-factor/single-class model. Note that, again, the AIC and saBIC only select the correct model 51% and 66% of the time. Adding the second-choice proportion leads to acceptable values (i.e., > .95). As can be seen in Table 9 containing the average values, this is due to the competing three-factor single-class model. That model would, however, be rejected due to high factor correlations pertaining to the third factor.

TABLE 8
Data Generated Under the Two-Factor/Single-Class Model With Simple Structure: Proportions of Monte Carlo Replications the Correct Model Is First and Second Choice Using Information Criteria, and Proportion the aLRT Indicates the Necessity of the Second Class When Fitting the Correct Model

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
First choice	0.00	0.51	1.00	0.66	1.00	1.00
Second choice	0.17	0.44	0.00	0.30	0.00	0.00

Note. aLRT = adjusted likelihood ratio test; logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC.

TABLE 9
Averaged Fit Statistics for 200 Data Sets Generated by a Two-Factor/Single-Class Model With Simple Structure, Fitting Latent Profile Models LPMc*j* With *j* Classes and Latent Factor Models F*i*c*j* With *i* Factors and *j* Classes

	logL-val	*AIC*	*BIC*	*saBIC*	*aLRT*
LPAc2	–2888.88	5859.76	5995.00	5865.10	0.02
LPAc3	–2780.10	5684.19	5888.69	5692.26	0.14
F1C1	–2788.44	5636.89	5735.84	5640.79	NA
F2C1	**–2589.50**	**5257.01**	**5385.64**	**5262.08**	NA
F3C1	–2584.58	5263.17	5418.19	5269.29	NA
F1C2	–2623.65	5369.31	5570.51	5377.25	0.13

Note. Results for the true model are in bold, as are the values of fit indices that favor misspecified models. logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC; aLRT = adjusted likelihood ratio test.

Two-Factor/Single-Class Data With Cross Loadings (Tables 10 and 11)

The same set of models were fitted as for the preceding simple structure case. The results show no real difference with the models fitted to the two-factor single-class data with simple structure (cf. Tables 8 and 9 to Tables 10 and 11). The proportions of correct model choice are slightly improved for the AIC and saBIC.

TABLE 10
Data Generated Under the Two-Factor/Single-Class Model With Cross Loadings: Proportions of MC Replications the Correct Model Is First and Second Choice Using Information Criteria, and Proportion the aLRT Indicates the Necessity of the Second Class When Fitting the Correct Model

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
First choice	0.032	0.78	1.00	0.84	1.00	1.00
Second choice	0.168	0.22	0.00	0.16	0.00	0.00

Note. aLRT = adjusted likelihood ratio test; logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC.

TABLE 11
Averaged Fit Statistics for 200 Data Sets Generated by a Two-Factor/Single-Class Model Wth Cross Loadings, Fitting Latent Profile Models LPMc*j* With *j* Classes and Latent Factor Models F*icj* With *i* Factors and *j* Classes

	logL-val	*AIC*	*BIC*	*saBIC*	*aLRT*
LPMc2	–2790.98	5663.96	5799.19	5669.30	0.01
LPMc3C3	–2695.18	5514.36	5718.86	5522.43	0.13
F1C1	–2676.90	5413.90	5512.90	5417.80	NA
F2C1	**–2558.30**	**5194.60**	**5323.3**0	**5199.70**	NA
F3C1	–2552.10	5198.20	5353.30	5204.40	NA
F1C2	–2603.37	5308.73	5476.95	5315.38	0.24

Note. Results for the true model are in bold, as are the values of fit indices that favor misspecified models. logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC; aLRT = adjusted likelihood ratio test.

Measurement Invariant Single-Factor/Two-Class Data (Tables 12 and 13)

Here, the classes in the true data only differ with respect to the factor mean. The set of fitted models includes LPMs with two, three, and four classes; exploratory factor models with a single, two, and three factors; and a two-class exploratory factor model with a single factor. Tables 12 and 13 depict the results for this data type.

Small separation. Restricting the analysis to fitting only LPMs would lead to an overextraction of classes. Among LPMs, information criteria favor the four-class model. Similar to the previous results, a model with one class less would be chosen if the decision is based on the aLRT, that is, a three-class model would be chosen where the true data consist of two classes. Exploratory factor analysis for a single population lead to the following average results. The AIC favors the three-factor/single-class model (F3C1), but this would in practice be rejected due to very high factor correlations. Averaged over the 100 replications, all factor correlations are .99 with standard deviations of .13, .09, and .14 for the three different factor correlations. The BIC and saBIC are smallest for the single-factor/single-class model (F1C1). The aLRT for the correct single-factor/two-class model (F1C2) versus the incorrect F1C1 is only marginally significant, hence the necessity of a second class would be debatable in practice. Apparently, with small separation between classes and a within-class sample size of 200, for the measurement invariant F1C2 model there is a possibility of underestimating the number of classes. Table 13 shows that there are a number of competing models with small differences in average information criteria. This is consistent with the very low proportions of correct model choice presented in Table 12.

TABLE 12
Data Generated Under the Measurement Invariant Single-Factor/Two-Class Model: Proportions of Monte Carlo Replications the Correct Model Is First and Second Choice Using Information Criteria, and Proportion the aLRT Indicates the Necessity of the Second Class When Fitting the Correct Model

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
Small class separation						
First choice	0.85	0.18	0.00	0.02	0.00	0.63
Second choice	0.09	0.26	0.18	0.26	0.17	0.00
Larger class separation						
First choice	1.00	0.99	0.02	0.90	0.00	0.99
Second choice	0.00	0.00	0.47	0.09	0.37	0.00

Note. aLRT = adjusted likelihood ratio test; logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC.

TABLE 13
Averaged Fit Statistics for 200 Data Sets Generated by a Measurement Invariant Single-Factor/Two-Class Model, Fitting Latent Profile Models LPMc*j* With *j* Classes and Latent Factor Models F*i*c*j* With *i* Factors and *j* Classes

	logL-val	*AIC*	*BIC*	*saBIC*	*aLRT*
Small class separation					
LPMc2	–5791.85	11665.70	11829.35	11699.25	0.01
LPMc3	–5352.87	10829.74	11077.21	10880.48	0.04
LPMc4	–5152.52	10471.03	10802.33	10538.96	0.11
F1C1	–5002.60	10065.20	**10185.00**	**10089.80**	NA
F2C1	–4994.50	10066.90	10222.60	10098.90	NA
F3C1	–4984.60	**10063.30**	10250.90	10101.80	NA
F1C2	**–4985.56**	**10073.12**	**10276.68**	**10114.86**	**0.08**
Larger class separation					
LPMc2	–6265.20	12612.39	12776.04	12645.95	0.00
LPMc3	–5759.67	11643.35	11890.82	11694.09	0.09
LPMc4	–5445.28	11056.56	11387.85	11124.48	0.05
F1C1	–5153.00	10366.00	10485.70	10390.50	NA
F2C1	–5144.50	10367.00	10522.70	10398.90	NA
F3C1	–5133.80	10361.50	10549.20	10400.00	NA
F1C2	**–5113.49**	**10328.97**	**10532.54**	**10370.71**	**0.00**

Note. Results for the true model are in bold, as are the values of fit indices that favor misspecified models. logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC; aLRT = adjusted likelihood ratio test.

Larger separation. With larger class separation, the risk of overextracting classes when fitting only latent class models is higher than for small separation. All indices point to the LPM with four classes. Exploratory factor analysis for a single population would be inconclusive as the information criteria do not consistently favor the same model. However, comparing the entire set of fitted models would, on average, lead to a correct model choice. The proportions of correct model choice are good for AIC and saBIC (i.e., > .95) but low for BIC and CAIC. As we see, the situation is much improved when the data-generating model deviates from measurement invariance (e.g., class-specific intercepts and/or loadings).

Noninvariant Single-Factor/Two-Class Data: Class-Specific Intercepts (Tables 14 and 15)

Small separation. The same models were fitted as to the measurement invariant single-factor model/two-class data. In the noninvariant case, with small class separation, the AIC favors the correct model in a much higher proportion of MC replications than in the measurement invariance (MI) case (cf. Tables 14

TABLE 14
Data Generated Under the Single-Factor/Two-Class Model With Class-Specific Intercepts: Proportions of Monte Carlo Replications the Correct Model Is First and Second Choice Using Information Criteria, and Proportion the aLRT Indicates the Necessity of the Second Class When Fitting the Correct Model

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
Small class separation						
First choice	0.94	0.52	0.00	0.19	0.00	1.00
Second choice	0.05	0.29	0.41	0.55	0.19	0.00
Larger class separation						
First choice	1.00	1.00	1.00	1.00	1.00	1.00
Second choice	0.00	0.00	0.00	0.00	0.00	0.00

Note. aLRT = adjusted likelihood ratio test; logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC.

TABLE 15
Averaged Fit Statistics for 200 Data Sets Generated by a Single-Factor/Two-Class Model With Intercept Differences, Fitting Latent Profile Models LPMc*j* With *j* Classes and Latent Factor Models F*i*c*j* With *i* Factors and *j* Classes

	logL-val	*AIC*	*BIC*	*saBIC*	*aLRT*
Small class separation					
LPMc2	–5560.68	11203.36	11367.01	11236.92	0.01
LPMc3	–5280.87	10685.73	10933.20	10736.47	0.07
LPMc4	–5161.74	10489.48	10820.77	10557.40	0.13
F1C1	–5114.00	10288.10	10408.00	10313.01	NA
F2C1	–5056.90	10191.80	**10347.50**	**10223.70**	NA
F3C1	–5050.40	10194.80	10382.40	10233.30	NA
F1C2	**–5043.93**	**10189.86**	**10393.42**	**10231.60**	0.00
Larger class separation					
LPMc2	–5946.46	11974.92	12138.57	12008.47	0.01
LPMc3	–5715.29	11554.58	11802.05	11605.32	0.06
LPMc4	–5506.48	11178.96	11510.25	11246.88	0.06
F1C1	–5639.00	11338.01	11458.01	11363.03	NA
F2C1	–5252.50	10583.00	10738.60	10614.90	NA
F3C1	–5249.10	10592.21	10779.80	10630.60	NA
F1C2	**–5158.86**	**10419.73**	**10623.29**	**10461.46**	0.00

Note. Results for the true model are in bold, as are the values of fit indices that favor misspecified models. logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC; aLRT = adjusted likelihood ratio test.

and 15). The BIC and the aBIC do not perform well, and point to the two-factor/single-class model (see Tables 14 and 15) . However, the average factor correlation is again high, namely, .95 with a standard deviation of .04. Restricting the analysis to exploratory factor analysis for a single population would therefore be unlikely to result in an overextraction of factors. Comparison of the entire set of fitted models would likely result in a correct model choice because the aLRT clearly indicated the necessity of a second class.

Larger separation. For the large class separation, all indices are in accordance and would lead to a correct model choice. The proportion correct model selection equals unity for all information criteria. Overall, the results are clearly better than for the measurement invariant model.

Noninvariant Single-Factor/Two-Class Data: Class-Specific Intercepts and Factor Loadings (Tables 16 and 17)

Small separation. Class-specific factor loadings in addition to class-specific intercepts further increase the proportion correct model selection when compared to the measurement invariant model and the model with class-specific intercepts. Note that the class separation had been kept unchanged when introducing class specific parameters during data generation. The AIC and the sample size aBIC outperform the BIC and CAIC.

Larger separation. As for the model with class-specific intercepts, model choice is unproblematic for larger class separation.

TABLE 16
Data Generated Under the Single-Factor/Two-Class Model With Class-Specific Intercepts and Factor Loadings: Proportions of Monte Carlo Replications the Correct Model Is First and Second Choice Using Information Criteria, and Proportion the aLRT Indicates the Necessity of the Second Class When Fitting the Correct Model

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
Small class separation						
First choice	1.00	0.99	0.35	0.98	0.16	1.00
Second choice	0.00	0.01	0.62	0.01	0.66	0.00
Larger class separation						
First choice	1.00	1.00	1.00	1.00	1.00	1.00
Second choice	0.00	0.00	0.00	0.00	0.00	0.00

Note. aLRT = adjusted likelihood ratio test; logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC.

TABLE 17
Averaged Fit Statistics for 200 Data Sets Generated by a Single-Factor/Two-Class Model With Intercept and Loading Differences, Fitting Latent Profile Models LPMc*j* With *j* Classes and Latent Factor Models F*i*c*j* With *i* Factors and *j* Classes

	logL-val	*AIC*	*BIC*	*saBIC*	*aLRT*
Small class separation					
LPAc2	–5568.18	11218.36	11382.01	11251.91	0.01
LPAc3	–5295.87	10715.74	10963.21	10766.48	0.07
LPAc4	–5180.56	10527.13	10858.42	10595.06	0.11
F1C1	–5142.26	10344.53	10464.27	10369.08	NA
F2C1	–5087.63	10253.27	**10408.94**	10285.19	NA
F3C1	–5079.22	10252.45	10440.05	10290.91	NA
F1C2	**–5054.18**	**10210.36**	**10413.92**	**10252.09**	**0.00**
Larger class separation					
LPAc2	–5949.27	11980.54	12144.19	12014.09	0.00
LPAc3	–5718.41	11560.82	11808.29	11611.56	0.06
LPAc4	–5511.03	11188.05	11519.35	11255.98	0.10
F1C1	–5654.92	11369.83	11489.58	11394.39	NA
F2C1	–5281.25	10640.49	10796.16	10672.41	NA
F3C1	–5271.35	10636.69	10824.29	10675.16	NA
F1C2	**–5159.15**	**10420.3**	**10623.86**	**10462.04**	**0.00**

Note. Results for the true model are in bold, as are the values of fit indices that favor misspecified models. logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC; aLRT = adjusted likelihood ratio test.

Two-Factor/Two-Class Data (Tables 18 and 19)

The data are generated under a model with moderate factor correlations in both classes (i.e., .5), and class-specific intercepts. The fitted models include LPMs with two to four classes, exploratory single-class factor models with one to three factors, and exploratory two-class factor mixture models with one and two factors. The results are shown in Tables 18 and 19.

Small separation. An LPA would lead to overextraction of classes. All information criteria favor the four-class LPM when considering only LPMs. The aLRT points to the three-class LPM. An exploratory factor analysis would lead to the correct number of factors. Although the fit measures indicate three factors, one of the factors has correlations with the other two factors approaching unity. When considering all fitted models, the correct model would be chosen. Although the percentage of correct model choice is low for the AIC, and close to zero for the BIC and CAIC, Table 19 shows that the competing models are two- and three-factor single-class models. The aLRT is significant in 83% of the MC replications (see

TABLE 18
Data Generated Under the Two-Factor/Two-Class Model: Proportions of Monte Carlo Replications the Correct Model Is First and Second Choice Using Information Criteria, and Proportion the aLRT Indicates the Necessity of the Second Class When Fitting the Correct Model

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
Small class separation						
First choice	0.99	0.57	0.00	0.17	0.00	0.83
Second choice	0.01	0.43	0.53	0.83	0.22	0.00
Larger class separation						
First choice	1.00	1.00	1.00	1.00	1.00	1.00
Second choice	0.00	0.00	0.00	0.00	0.00	0.00

Note. aLRT = adjusted likelihood ratio test; logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC.

TABLE 19
Averaged Fit Statistics for 200 Data Sets Generated by a Two Factor/two Class Model, Fitting Latent Profile Models LPMc*j* With *j* Classes and Latent Factor Models F*i*c*j* With *i* Factors and *j* Classes

	logL-val	*AIC*	*BIC*	*saBIC*	*aLRT*
Small class separation					
LPAc2	–5925.73	11933.47	12097.12	11967.02	0.01
LPAc3	–5753.79	11631.58	11879.05	11682.32	0.09
LPAc4	–5623.29	11412.59	11743.88	11480.52	0.13
F1C1	–5745.32	11550.63	11670.38	11575.18	NA
F2C1	–5400.01	10878.02	11033.68	10909.93	NA
F3C1	–5350.53	10795.06	**10982.66**	**10833.52**	NA
F1C2	–5556.19	11214.37	11417.94	11256.11	0.05
F2C2	**–5338.01**	**10794.02**	**11029.52**	**10842.31**	**0.04**
Larger class separation					
LPAc2	–6250.65	12583.31	12746.96	12616.86	0.02
LPAc3	–6083.58	12291.15	12538.62	12341.89	0.11
LPAc4	–5953.88	12073.76	12405.05	12141.68	0.16
F1C1	–6116.62	12293.23	12412.98	12317.79	NA
F2C1	–5759.40	11596.79	11752.46	11628.71	NA
F3C1	–5542.90	11179.80	11367.40	11218.27	NA
F1C2	–5812.71	11727.43	11930.99	11769.17	0.01
F2C2	**–5453.08**	**11024.17**	**11259.66**	**11072.45**	**0.00**

Note. Results for the true model are in bold, as are the values of fit indices that favor misspecified models. logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC; aLRT = adjusted likelihood ratio test.

Table 18), indicating the need for a second class. The three-factor model is rejected due to high factor correlations.

Larger separation. Correct model choice is unproblematic for larger class separation.

Part 2: Effects of Sample Size and Class Proportions

Varying Sample Size (Table 20)

The data-generating model used to investigate the effect of sample size on the detection of the correct model is the single-factor/two-class model with class-specific intercepts. The within-class sample size is denoted as N_{wc}. Results for N_{wc} = 200 for this model were also presented in Part 1. Here, results are presented for within-class sample sizes of N_{wc} =25, 50, 75, 150, 200, 300, and 1000. Table 20 shows the proportion correct model choice (i.e., "first choice" in the previous tables) as a function of increasing sample size. The averaged results are discussed in the text.

TABLE 20
Data Generated Under the Single-Factor/Two-Class Model With Intercept Differences: Proportions of Correct Model Choice as a Function of Increasing Sample Size, and Proportion the aLRT Indicates the Necessity of the Second Class When Fitting the Correct Model

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
Small class separation						
N_{wc} = 25	0.25	0.30	0.00	0.48	0.00	0.06
N_{wc} = 50	0.62	0.25	0.00	0.52	0.00	0.37
N_{wc} = 75	0.78	0.34	0.00	0.42	0.00	0.77
N_{wc} = 150	0.87	0.40	0.00	0.20	0.00	0.98
N_{wc} = 200	0.94	0.52	0.00	0.19	0.00	1.00
N_{wc} = 300	0.94	0.68	0.00	0.19	0.00	1.00
N_{wc} = 1,000	1.00	1.00	0.10	0.85	0.01	1.00
Larger class separation						
N_{wc} = 25	0.86	0.78	0.18	0.88	0.04	0.90
N_{wc} = 50	1.00	1.00	0.57	1.00	0.34	0.99
N_{wc} = 75	1.00	1.00	0.88	1.00	0.67	1.00
N_{wc} = 150	1.00	1.00	1.00	1.00	1.00	1.00
N_{wc} = 200	1.00	1.00	1.00	1.00	1.00	1.00
N_{wc} = 300	1.00	1.00	1.00	1.00	1.00	1.00
N_{wc} = 1,000	1.00	1.00	1.00	1.00	1.00	1.00

Note. aLRT = adjusted likelihood ratio test; logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC.

Small separation. For very small within-class sample sizes (i.e., $N_{wc} = 25$), the only distinction that can be made is between LPMs on one hand and factor models on the other hand. LPMs with two and three classes have clearly higher information criteria and would therefore be rejected in favor of the factor models. However, a comparison of the information criteria corresponding to the different factor models does not allow for a clear decision. Average information criteria show that single-class models with one or two factors are competing models. The BIC and CAIC favor the single-factor/one-class model, whereas the AIC and saBIC favor the two-factor model. Factor correlations are high on average (i.e., .88). The aLRT does not support the necessity of the second class. In only 6 of the 100 MC replications was the aLRT less than or equal to .05. In sum, most likely the population heterogeneity would not be detected, and a single-factor or two-factor model for a single class would be chosen.

Increasing the within-class sample size results in an increasing proportion correct model selection when considering AIC and aLRT. Interestingly, the saBIC proportion correct drops to .2 for sample sizes between 150 and 300 before picking up again for large sample sizes. The two-factor/one-class model remains the competing model for the BIC and CAIC. As sample size increases, the factor correlation approaches unity (i.e., at $N_{wc} = 200$, it equals .95). When considering information criteria, aLRT, and parameter estimates jointly, a within-class sample size of $N_{wc} = 200$ is a conservative choice to achieve correct model selection, and reasonable results can be expected with 75 participants within class for this type of data.

Larger separation. For larger class separation, the results in Table 20 show that AIC, saBIC, and aLRT perform well at a sample size as small as 25 participants within class. For the BIC and CAIC, the competing model is again the two-factor/single-class model. When considering the BIC or CAIC jointly with the factor correlations, a within-class sample size of 50 would be sufficient.

It is important to note that if model comparisons are restricted to LPMs, larger sample sizes lead to accepting models with a greater number of classes. For the largest sample size in this study, the four-class LPM would be accepted although the true data only have two classes. This indicates that the risk of overestimating the number of classes when using only latent class or LPMs is pronounced in studies with large sample sizes. This result emphasizes the importance of fitting both latent class models and models with continuous latent variables to a given data set.

Unequal class proportions (Table 21)

The data-generating model is again the single-factor/two-class model with intercept differences. Here, class proportions are .9 and .1, respectively. The results in Table 21 can therefore be compared to Table 14 showing the results of the same data-generating model with equal class proportions.

TABLE 21
Data Generated Under the Single-Factor/Two-Class Model With Intercept Differences: Effect of Unequal Class Proportions on the Proportions of Monte Carlo Replications the Correct Model Is First and Second Choice

	logL-val	*AIC*	*BIC*	*saBIC*	*CAIC*	*aLRT*
Small class separation						
First choice	0.89	0.77	0.00	0.39	0.00	0.05
Second choice	0.06	0.11	0.09	0.24	0.06	0.00
Larger class separation						
First choice	1.00	1.00	1.00	1.00	1.00	0.98
Second choice	0.00	0.00	0.00	0.00	0.00	0.00

Note. logL-val = log likelihood value; AIC = Akaike information criterion; BIC = Bayesian information criterion; saBIC = sample size adjusted BIC; CAIC = Consistent AIC; aLRT = adjusted likelihood ratio test.

Except for the aLRT, unequal class proportions do not seem to have a major impact on the results. Interestingly, the AIC and saBIC perform slightly better than for equal class proportions, which is consistent with the findings concerning small sample sizes. The aLRT seems to be very sensitive to unequal class proportions and fails to support the second class 95% of the time when class separation is small. In general, however, the conclusions are similar to the ones concerning equal class proportions. The two-factor/single-class model is a competing model. This model would be rejected when considering factor correlations. For larger class separation, the correct model is detected without difficulty.

CONCLUSIONS

The results of this study demonstrate the value of exploratory factor mixture analysis for investigating the within-class factor structure while accounting for possible population heterogeneity. For the data-generating models and sample sizes used in this study, comparing the fit of a set of exploratory factor mixture models including LPMs correctly distinguishes between categorical and/or continuous latent variables in most cases.

A consistent finding in this study is that restricting analyses to fitting only latent class models results in overextraction of classes if continuous factors are part of the latent structure of the true data-generating model. The finding confirms examples provided in Bauer and Curran (2004), where fitting overly restricted within-class models resulted in spurious classes. In our simulation, overextraction of classes seems to depend neither on the type of within-class factor structure (i.e., simple structure vs. presence of cross loadings) nor on the restrictiveness of the within-class model for the means (i.e., measurement invariant vs. intercept differences). Overextraction occurred when the number of within-class factors was misspecified (i.e., specified as zero when in fact it was nonzero). It is noteworthy that the risk of overextracting classes is higher for large sample sizes.

Evidence for an overextraction of factors if an analysis is restricted to exploratory factor analysis and the population is incorrectly assumed to be homogeneous is less clear. The correct number of factors in a latent class model is zero. The researcher conducting factor analysis will likely conclude that there is at least one factor. That is, latent classes can generate data that appear to indicate a latent factor. This is a minimal level of overextraction. Sometimes two or more factors provide a better fit than a single-factor model, which is a more serious level of overextraction. However, although the fit indices can indicate the need for additional, spurious factors to account for covariation due to population heterogeneity, individual output files in this study show that factor structures resulting from fitting models with too many factors are rather unsatisfactory. High factor correlations suggest that the dimension of the factor space in the true model may be lower. Similarly, very low percentages of variance explained by the factor can be used as an indication that models without a continuous latent factor may be more appropriate.

It is a well-known fact in mixture analysis (McLachlan, Peel, & Bean, 2001) that correct model choice depends heavily on the separation between classes in the true model. In our study, classes were separated between 1.5 and 3 standard deviations. Note that when classes are separated by a 1.5 standard deviation factor mean difference and factors within class are normally distributed, the mixture distribution of factor scores would still look approximately normal, and not bimodal (McLachlan & Peel, 2000). Smaller class separations are investigated in Lubke and Muthén (2007). Our study clearly demonstrates that sample size plays an important role in addition to class separation. Regarding correct model choice, there is a clear trade-off between sample size and class separation. Interestingly, for the data-generating model used in the second part of our simulation, sample sizes as small as 75 participants within class often result in a correct model choice even if the distance between classes is so small that the heterogeneity would be hard to detect when plotting the data. For a larger separation, sample sizes as small as 25 participants within class result on average in correct model detection. Interestingly, the AIC and saBIC outperform the BIC and CAIC. The aLRT performs well at very small sample sizes. Further research is needed to investigate in how far results with respect to sample size can be generalized to other types of mixture models and other degrees of separation.

A limitation of this study is the fact that the data-generating models did not violate any of the model assumptions. There is no doubt that model assumptions such as multivariate normality are frequently unrealistic in practice. Our study represents a first step in evaluating exploratory factor mixture models when neither the number of factors nor the number of classes is known. Future research will show in how far our results are robust against violations of model assumptions. In addition, the emphasis in this study was on mean differences between classes. It remains to be seen if covariance structure differences can be detected in the absence of substantial mean differences.

In summary, comparing the fit of a set of exploratory factor mixture models including classic latent class models as proposed in this article helps to prevent overextraction of classes or factors. This study is could be extended in several ways. First, as previously mentioned, the data were generated in accordance with the assumptions of exploratory factor mixture models. Violations of some of these assumptions (e.g., normally distributed variables within class) have been shown to induce an overestimation of the number of classes (Bauer & Curran, 2003a) and therefore warrant further investigation. Second, the number of latent classes and the number of factors was kept low in our study. The choice between a four-factor/three-class model and a three-factor/four-class model based on information criteria, the aLRT, and inspection of individual output files may be more difficult than for one or two factors and/or classes. Higher dimensions of the latent space are currently under investigation. Third, the set of exploratory factor mixture models will in practice be unlikely to include the true model. Real data may well stem from more complex data-generating processes, including, for instance, differences between classes with respect to the structure of the within-class model. Hence, in practice, model selection will be between more or less misspecified models, and decisions are likely to be less clear cut. A broader set of candidate models could therefore be tested. Finally, it is not clear how reliable distinctions will be when the observed data consist of ordinal or mixed data instead of the continuous measures used here. This issue is also currently being investigated.

ACKNOWLEDGMENTS

The research of both authors was supported through grant MH65322 by the National Institutes of Mental Health and through grant DA018673 by the National Institute on Drug Abuse.

REFERENCES

Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on automatic Control, AU-19,* 719–722.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317–332.

Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional meanand covariance structure models. *Psychometrika, 64,* 475–494.

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

Bartholomew, D. J., & Knott, M. (1999). *Latent variables models and factor analysis* (2nd ed.). London: Arnold.

Bauer, D. B., & Curran, P. J. (2003a). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8,* 338–363.

Bauer, D. B., & Curran, P. J. (2003b). Overextraction of latent trajectory classes: Much ado about nothing? Reply to Rindskopf (2003), Muthen (2003), and Cudeck and Henly (2003). *Psychological Methods, 8,* 384–393.

Bauer, D. B., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9,* 3–29.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52,* 345–370.

Cudeck, R., & Henly, S. J. (2003). A realistic perspective on pattern prepresentation in growth data: Comment on Bauer and Curran. *Psychological Methods, 8,* 378–383.

Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika, 63,* 227–253.

Heinen, T. (1993). *Discrete latent variable models.* Tilburg, The Netherlands: Tilburg University Press.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences.* Thousand Oaks, CA: Sage.

Hicks, B. M., Krueger, R. F., Iacono, W. G., McGue, M., & Patrick, C. J. (2004). Family transmission and heritability of externalizing disorders: A twin-family study. *Archives of General Psychiatry, 61,* 922–928.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite mixture structural equation models for response based segmentation and unobserved heterogeneity. *Marketing Science, 16,* 39–59.

Krueger, R. F., Nichol, P. E., Hicks, B. M., Markon, K. E., Patrick, C. J., Lacono, W. G., et al. (2004). Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychological Assessment, 16,* 107–119.

Langenheine, R., & Rost, J. (1988). *Latent trait and latent class models.* New York: Plenum.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method.* London: Butterworth.

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

Lo, Y., Mendell, N., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88,* 767–778.

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). Absence of measurement bias with respect to unmeasured variables: An implication of strict factorial invariance. *British Journal of Mathematical and Statistical Psychology, 56,* 231–248.

Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10,* 21–39.

Lubke, G. H., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling, 14,* 26–47.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models.* New York: Wiley.

McLachlan, G. J., Peel, D., & Bean, R. W. (2001). *Modelling high-dimensional data by mixtures of factor analyzers* (Tech. Rep.). Brisbane, Australia: Department of Mathematics, University of Queensland.

Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychological Bulletin: Quantitative Methods in Psychology, 115,* 300–307.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58,* 525–543.

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 203–240). Washington, DC: American Psychological Association.

Molenaar, P. C. M., & van Eye, A. (1994). On the arbitrary nature of latent variables. In A. van Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 226–242). Thousand Oaks, CA: Sage.

Muthén, B. O. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8,* 369–377.

Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55,* 463–469.

Pickles, A., & Angold, A. (2003). Natural categories or fundamental dimensions: On carving nature at the joints and the rearticulation of psychopathology. *Development and Psychopathology, 15,* 529–551.

Rindskopf, D. (2003). Mixture or homogeneous? Comment on Bauer and Curran (2003). *Psychological Methods, 8,* 364–368.

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6,* 461–464.

Sclove, S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333–343.

Spearman. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Vermunt, J. K., & Magidson, J. (2003). Latent class models for classification. *Computational Statistics & Data Analysis, 41,* 531–537.

Vermunt, J. K., & Magidson, J. (2005). Factor analysis with categorical indicators: A comparison between traditional and latent class approaches. In A. K. d. Ark, M. A. Croon, & K. Sijtsma (Eds.), *New developments in categorical data analysis for the social and behavioral sciences* (pp. 41–62). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Yung, Y. F. (1997). Finite mixtures in confirmatory factor analysis models. *Psychometrika, 62,* 297–330.

APPENDIX

The parameter values used for the data generation are shown for each model. Only nonzero parameters are listed.

First Part

Two-class latent profile model

Class-invariant parameters:
residual variances $[0.7\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5]^t$
Class-specific parameters:
small separation, means class 2 $[0.35\ -.2\ .6\ -.75\ 0.35\ -.2\ .6\ -.75\ 0.35\ -.2]^t$
larger separation, means class 2 $[.8\ -.8\ 1.2\ -1.2\ .8\ -.8\ 1.2\ -1.2\ .8\ -.8]^t$

Three-Class Latent Profile Model

Class-invariant parameters:
residual variances $[0.7\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5]^t$
Class-specific parameters:
means class 1 $\boldsymbol{\upsilon} = [0\ 0\ 0\ 0\ 0]^t$
small separation, means class 2 $[.7\ -.8\ 1.2\ -1.1\ .7\ -.8\ 1.2\ -1.1\ .7\ -.8]^t$
small separation, means class 3 $[1.3\ -1.2\ 1.2\ -1.3\ 1.3\ -1.2\ 1.3\ -1.2\ 1.3\ -1.2]^t$
larger separation, means class 2 $[1.5\ -1.5\ 1.8\ -1.8\ 1.5\ -1.5\ 1.8\ -1.8\ 1.5\ -1.5]^t$
larger separation, means class 3 $[2.5\ -2.5\ 2.3\ -2.3\ 2.5\ -2.5\ 2.3\ -2.3\ 2.5\ -2.5]^t$

Single-Factor/Single-Class Model

factor loadings $[1\ .8\ .8\ .8\ .8\ .8\ .8\ .8\ .8\ .8]^t$
factor variance 1.2
residual variances $[0.7\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5]^t$

Two-Factor/Single-Class Model With Simple Structure

factor loadings $\begin{bmatrix} 1 & .8 & .8 & .8 & .8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & .8 & .8 & .8 & .8 \end{bmatrix}^t$
factor covariance matrix $\begin{bmatrix} 1.2 & .5 \\ .5 & 1.2 \end{bmatrix}$
residual variances $[0.7\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5]^t$

Two-Factor/Single-Class Model With Crossloadings

factor loadings $\begin{bmatrix} 1 & .8 & .6 & .4 & .8 & 0 & .2 & 0 & .4 & 0 \\ 0 & 0 & .2 & .4 & 0 & 1 & .6 & .8 & .4 & .8 \end{bmatrix}^t$

factor covariance matrix $\begin{bmatrix} 1.2 & .5 \\ .5 & 1.2 \end{bmatrix}$

residual variances $[0.7\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5]^t$

Measurement Invariant Single-Factor/Two-Class Model

Class-invariant parameters:
factor loadings $[1\ .8\ .8\ .8\ .8\ .8\ .8\ .8\ .8\ .8]^t$
factor variance 1.2
residual variances $[0.7\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5]^t$
Class-specific parameters:
small class separation factor mean class 2 is 1.8
larger class separation factor mean class 2 is 3.4

Single-Factor/Two-Class Model With Intercept Differences

Class-invariant parameters:
factor loadings $[1\ .8\ .8\ .8\ .8\ .8\ .8\ .8\ .8\ .8]^t$
factor variance 1.2
residual variances $[0.7\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5]^t$
Class-specific parameters:
small class separation intercepts class 2 $[0.35\ -.2\ .6\ -.75\ 0.35\ -.2\ .6\ -.75\ 0.35\ -.2]$
larger class separation intercepts class 2 $[.8\ -.8\ 1.2\ -1.2\ .8\ -.8\ 1.2\ -1.2\ .8\ -.8]$

Single-Factor/Two-Class Model With Intercept and Loading Differences

Class-invariant parameters:
factor variance 1.2
residual variances $[0.7\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5]^t$
Class-specific parameters:
small class separation intercepts class 2 $[0.35\ -.2\ .6\ -.75\ 0.35\ -.2\ .6\ -.75\ 0.35\ -.2]$
larger class separation intercepts class 2 $[.8\ -.8\ 1.2\ -1.2\ .8\ -.8\ 1.2\ -1.2\ .8\ -.8]$
factor loadings class 1 $[1\ .8\ .8\ .8\ .8\ .8\ .8\ .8\ .8\ .8]^t$
factor loadings class 2 $[1\ .49\ .68\ .95\ .95\ .58\ .92\ .61\ .88\ .96]$

Two-Factor/Two-Class Model With Intercept Differences

Class-invariant parameters:

factor loadings $\begin{bmatrix} 1 & .8 & .8 & .8 & .8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & .8 & .8 & .8 & .8 \end{bmatrix}^t$

factor covariance matrix $\begin{bmatrix} 1.2 & .5 \\ .5 & 1.2 \end{bmatrix}$

residual variances $[0.7\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5\ .5]^t$

Class-specific parameters:

small class separation intercepts class 2 [0.35 – .2 .6 – .75 0.35 – .2 .6 – .75 0.35 – .2]

larger class separation intercepts class 2 [.8 – .8 1.2 – 1.2 .8 – .8 1.2 – 1.2 .8 – .8]

Second Part

The parameter values used to generate data with varying sample sizes are the same as for the single-factor/two-class model with intercept differences.

Prior class probabilities used to generate data with unequal class sizes are .9 and .1.

Information Criteria

All information criteria used in our study are penalized log-likelihood functions with the general form $-2L + f(N)p$ where L is the loglikelihood of the estimated model with p free parameters and $f(N)$ is a function that may depend on the total sample size N (Sclove, 1987). The AIC does not depend on sample size, the penalty is $f(N)p = 2p$ (Akaike, 1974, 1987). The BIC, the CAIC, and the saBIC integrate N in different ways, the respective penalty terms are $\log(N)p$ and $(\log(N) + 1)p$ for the BIC and the CAIC (Bozdogan, 1987; Schwarz, 1978). The saBIC uses $(N^* = (N + 2)/24)$ instead of N.

Example Mplus Input File: Single-Factor/Two-Class Model

```
TITLE: F1 C2
DATA: FILE IS yinc.dat;
VARIABLE:
NAMES ARE ycont1-ycont10 eta1 eta2 c1 c2 tc;
USEVARIABLES ARE ycont1-ycont10;
CLASSES = c(2);
ANALYSIS: TYPE = MIXTURE;
ITERATIONS = 1000;
starts=50 10;
stiterations=10;
```

```
stscale=5;
stseed=198;
MODEL: %OVERALL%
f1 BY ycont1* ycont2-ycont10; f1@1;
%c#2%
f1 BY ycont1* ycont2-ycont10; f1@1; [F1@0];
!estimate class specific factor loadings
[ycont1-ycont10@0]; ycont1-ycont10;
! fix intercepts to zero in this class, estimate intercept differences in the other class
! estimate residual variances
%c#1%
f1 BY ycont1*  ycont2-ycont10; f1@1; [F1@0];
[ycont1-ycont10]; ycont1-ycont10;
OUTPUT: tech1 tech8 tech11 standardized;
```