

Investigating Population Heterogeneity With Factor Mixture Models

Gitta H. Lubke

University of Notre Dame

Bengt Muthén

University of California, Los Angeles

Sources of population heterogeneity may or may not be observed. If the sources of heterogeneity are observed (e.g., gender), the sample can be split into groups and the data analyzed with methods for multiple groups. If the sources of population heterogeneity are unobserved, the data can be analyzed with latent class models. Factor mixture models are a combination of latent class and common factor models and can be used to explore unobserved population heterogeneity. Observed sources of heterogeneity can be included as covariates. The different ways to incorporate covariates correspond to different conceptual interpretations. These are discussed in detail. Characteristics of factor mixture modeling are described in comparison to other methods designed for data stemming from heterogeneous populations. A step-by-step analysis of a subset of data from the Longitudinal Survey of American Youth illustrates how factor mixture models can be applied in an exploratory fashion to data collected at a single time point.

The populations investigated in the behavioral sciences and related fields of research are often heterogeneous. A sample may consist of explicitly defined groups such as experimental and control groups, and the aim is to compare these groups. On the other hand, the sources of population heterogeneity may not be known beforehand. Test scores on a cognitive test may reflect two types of children in the sample: those who master the knowledge required to solve the items (masters) and those who lack this critical knowledge (nonmasters). The interest may be to decide to which of the subpopulations a given child most likely belongs. In addition, it may be of interest to characterize masters and nonmasters using background variables to develop specific

preparatory courses. Psychopathology is another area in which unobserved population heterogeneity occurs. A disorder may consist of qualitatively different subtypes, and the interest may be to identify the subpopulations in order to develop subtype specific treatments or to find subtype specific genes. Meehl (1992) presents an extensive discussion of the type concept in psychology.

Subpopulation is a generic term indicating a cluster within a heterogeneous population. Population heterogeneity can be observed or unobserved. Heterogeneity is observed if it is possible to define the subpopulations based on an observed variable. For instance, it is known that gender often introduces heterogeneity in math achievement tests, and one can define two subpopulations (i.e., males and females) based on the observed variable gender. In the context of observed heterogeneity, subpopulations are called groups, and group membership is known for each participant. The data can be analyzed using models for multiple groups. Multiple-group analyses are appropriate if the interest is to compare explicitly defined groups such as gender groups, age groups, or experimental–control groups. Unobserved heterogeneity differs from the multiple-group situation. Here, the variables that cause the heterogeneity in the data are not known beforehand. Consequently, it is also not known to which of the subpopulations a participant belongs, and it is not possible to divide the sample into groups. The subpopulation membership of the participants has to be inferred from the data. In the context of unobserved heterogeneity, the subpopulations are called latent classes because subpopulation membership is not observed but latent. In this article, we use the term *subpopulation* in the context in which heterogeneity may or may not be observed. We reserve the term *group* in the context of

Gitta H. Lubke, Department of Psychology, University of Notre Dame; Bengt Muthén, Graduate School of Education & Information Studies, University of California, Los Angeles.

Additional materials are available on the Web at <http://dx.doi.org/10.1037/1082-989X.10.1.21.supp>

Bengt Muthén is co-developer of the M plus computer program used in this article. The research by Gitta H. Lubke was supported through National Institute of Mental Health (NIMH) Grant MH65322 and through a subcontract to National Institute of Child Health and Human Development Grant 5 R01 HD30995-07. The research of Gitta H. Lubke and Bengt Muthén was supported by NIMH and National Institute on Drug Abuse Grant MH40859 and by NIMH Grants MH01259, MH38725, and MH42968. Bengt Muthén was also supported by National Institute on Alcohol Abuse and Alcoholism Grant K02 AA 00230.

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

observed heterogeneity (i.e., gender defines two groups). Finally, we use the term *latent classes* to specifically address unobserved heterogeneity.

Different methods have been developed for the analysis of unobserved heterogeneity. Taxometric methods address the question of whether participants differ in degree with respect to an observed behavior or whether they belong to either of two qualitatively different types (Meehl, 1992; Waller & Meehl, 1998). The two types correspond to two latent classes. The taxometric approach does not offer the possibility of specifying models for observed data within each of the two latent classes. Classic latent class analysis allows for more than two latent classes but is also limited with respect to modeling observed variables within class. Factor mixture modeling is flexible with respect both to the number of latent classes and to modeling the observed variables within class.

The aim of the current article is to describe factor mixture models as a tool to explore unobserved population heterogeneity. Factor mixture models are a combination of the common factor model (Thurstone, 1947) and the classic latent class model (Lazarsfeld & Henry, 1968). The common factor model and the latent class model are latent variable models. Latent variables serve to model theoretical concepts or phenomena that are not directly observable. Observed variables that contain information about the theoretical concepts are used as indicators for the latent variables. There are important differences between the latent variables in the common factor model and the latent class model (for a historic overview, see Bartholomew, 1987; Heinen, 1996; Langenheine & Rost, 1988). In the context of the current study, the most important difference concerns the specific purpose of the latent variables in the common factor model compared with the latent class model. The common factor model is appropriate for data from a single homogeneous population. The common factor model is designed to investigate the common content of observed scores such as questionnaire items. Continuous latent variables called factors are used to model the common content of the observed variables. Participants are assumed to differ in degree with respect to the factors, and these differences produce the covariances of the observed items. The covariances are modeled by specifying regression relations between the observed items and the underlying continuous factors. Factor models serve to cluster items.

Latent class models, on the other hand, serve to cluster participants. This type of model is adequate if the sample consists of different subtypes and it is not known beforehand which participant belongs to which of the subtypes. The subtypes may differ qualitatively (e.g., qualitatively different subtypes of a psychiatric disorder) or quantitatively (e.g., a high- and a low-scoring class in an educational study). The latent variable in the latent class model is categorical, and the number of categories (i.e., number of

latent classes) represents the number of different clusters of participants in a sample. In other words, the latent categorical variable is used to model heterogeneity. In the classic form of the latent class model, observed variables within each latent class are assumed to be independent. Although in variations of the classic model this assumption is relaxed (see, for instance, Uebersax, 1999; Vermunt & Magidson, 2000; Vermunt & Magidson, 2002) no specific structure for the covariances of observed variables is specified.

The factor mixture model combines the latent class model and the common factor model and has a single categorical and one or more continuous latent variables. As in the latent class model, the categorical latent variable serves to model the unknown population heterogeneity. Different than in the latent class model, observed variables within class are not assumed to be independent, but a common factor model is specified to impose a structure on the covariance matrix and mean vector of the observed variables within class. In other words, observed variables within class are allowed to covary, and the covariation is modeled using underlying continuous factors.

In this article, special attention is devoted to different possibilities of integrating covariates in an investigation of unobserved population heterogeneity. Covariates can be used if some sources of heterogeneity are observed. If boys and girls differ with respect to their performance on a math test, then gender explains at least some of the heterogeneity in the data. The heterogeneity induced by gender can take several different forms. Gender may predict class membership (i.e., being a master vs. a nonmaster), the math factor score (i.e., mean differences in math within class), scores on one or more of the individual math items (i.e., absence of measurement invariance with respect to gender), or a combination of these. We show how these different options can be modeled and discuss their distinct interpretations.

Within this article, we first highlight characteristics of factor mixture models in comparison to other commonly used methods for data from heterogeneous populations. Next, the general factor mixture model is presented, and integration of covariates is described in detail. A step-by-step factor mixture analysis of empirical data illustrates how factor mixture models may be used to explore population heterogeneity.

Characteristics of Factor Mixture Models in Comparison to Other Methods for Heterogeneous Data

Different methods for the analysis of data from heterogeneous populations have been designed to address specific types of research questions. The following overview shows in which situation a given method is suitable and highlights the flexibility of factor mixture models in comparison to more traditional methods. A number of criteria can be used

to categorize methods for the analysis of data from heterogeneous populations. Here we focus on three important criteria, namely (a) whether population heterogeneity is observed or unobserved (i.e., multiple group vs. latent class methods), (b) whether observed variables within a subpopulation are assumed to be categorical, continuous, or both, and (c) whether the method is a latent variable approach. In a latent variable approach, observed scores are decomposed into a part that is due to the latent variable and a part that is the residual. Measurement error in the observed score is contained in the residual. The overview is summarized in Table 1.

Regarding the first criterion, well-known methods for observed group membership are discriminant analysis (DA), logistic (or ordered polytomous) regression (LR), multivariate analysis of variance (MANOVA), and multi-group common factor analysis (MG-CFA; Jöreskog, 1971; Stevens, 1992). MG-CFA is an extension of the common factor model to accommodate multiple groups. The groups in LR, MANOVA, and MG-CFA are defined using a single observed variable such as gender or a combination of several observed variables (e.g., different combinations of gender and highest educational degree). DA and LR are used if the interest is to identify variables that strongly predict group membership. MANOVA serves mean comparisons of groups with respect to a set of observed variables. MG-CFA is designed for group comparisons with respect to the means and covariances of a set of observed variables. MG-CFA encompasses MANOVA as a submodel.

Methods for observed group membership presuppose that each participant can be uniquely assigned to a single group based on one or more grouping variables. An example would be a group that consists entirely of males or a group that consists entirely of males with a college degree, depending on the definition of the grouping variable. Methods for unobserved heterogeneity are more flexible in that latent classes can consist predominantly of males, and the calculation of the proportion of males within a latent class is a result of the model estimation. Methods that do not require

heterogeneity to be observed are K-means clustering, latent class analysis (LCA), latent profile analysis (LPA), and factor mixture models (FMM; Arminger, Stein, & Wittenberg, 1999; Dolan & van der Maas, 1998; Jedidi, Jagpal, & DeSarbo, 1997a; Lazarsfeld & Henry, 1968; McLachlan & Peel, 2000; B. O. Muthén & Shedden, 1999; Vermunt & Magidson, 2002; Yung, 1997). These methods are designed to detect in a given data set clusters of participants with similar response patterns on a set of observed variables. The number of clusters or latent classes has to be prespecified. Clustering using the K-means method is achieved based on an arbitrarily chosen criterion, which aims at minimizing within-cluster variability while maximizing between-cluster variability. LCA, LPA, and FMM, on the other hand, are model based (see Magidson & Vermunt, 2002, for a comparison of LCA methods to K-means clustering). Model-based methods have the advantage that more rigorous methods can be applied for the comparison of alternative models (Vermunt & Magidson, 2002).

Second, the methods differ with respect to the assumptions concerning the distribution of the observed outcome variables within a subpopulation. Some methods are designed specifically for categorical or continuous variables, whereas others can handle different types of variables. DA, MANOVA, K-means, and LPA are methods for continuous outcome variables, whereas the classic LCA and LR are methods for categorical outcome variables. MG-CFA and factor mixture models can handle both categorical and continuous variables in the same analysis.

Third, an observed score can be decomposed into a score on an underlying latent variable and a residual containing measurement error. Only LPA, LCA, MG-CFA, and factor mixture models are latent variable models. Latent variables can be categorical or continuous. The classic LPA and LCA models have a single categorical latent variable, whereas MG-CFA has a single or several continuous latent variables. Factor mixture models are the only models with both a single categorical and a single or multiple continuous latent variables. Categorical and continuous latent variables serve

Table 1
Classification of Commonly Used Methods for Heterogeneous Populations

Method	Source of heterogeneity	Outcome variables	Latent variable model
Discriminant analysis	Observed	Continuous	No
Logistic regression	Observed	Categorical	No
MANOVA	Observed	Continuous	No
Multigroup CFA	Observed	Continuous and/or categorical	Yes
K-means clustering	Unobserved	Continuous	No
Latent class analysis	Unobserved	Categorical	Yes
Latent profile analysis	Unobserved	Continuous	Yes
Factor mixture modeling	Unobserved	Continuous and/or categorical	Yes

Note. MANOVA = multivariate analysis of variance; CFA = common factor analysis.

different goals and also differ with respect to their interpretation on a conceptual level. The single categorical latent variable in LPA, LCA, and factor mixture models serves to model class membership (Lazarsfeld & Henry, 1968). Models with a latent class variable are appropriate to analyze data from heterogeneous populations. The number of latent classes corresponds to the number of clusters of participants in the sample. In the classic LCA and LPA, all covariation between observed variables is modeled to be due to differences between classes. Within class, observed variables do not covary. This is called the assumption of local independence. The variances of the observed variables within class are residual variances. The common factor model, which is an important building block of MG-CFA and factor mixture models, has a single or multiple continuous latent variables called factors. The factors serve to capture the common content of the observed variables rather than to model clusters of participants. Observed variables in the common factor model are separated into factor scores and a residual that contains specific factors and measurement error. Factor models are appropriate, for instance, for data from questionnaire items, which are designed to measure an underlying continuous construct and are likely to contain measurement error. Because factor mixture models incorporate categorical and continuous latent variables, they may be used for the analysis of data with underlying continuous constructs while simultaneously modeling population heterogeneity.

In summary, factor mixture models are designed for heterogeneous data, in which the sources of heterogeneity are unobserved. The sources of heterogeneity are represented by the latent class variable. The covariance matrix and means of the observed variables within each latent class are structured in terms of the common factor model such that measurement error is taken into account. Factor mixture models include several well-known models as submodels. The common factor model can be derived by setting the number of classes to one. The latent class and latent profile models can be derived by setting the variances of within-class factors to zero. The growth mixture model can be derived by imposing a specific structure on the within-class covariance matrix (see B. O. Muthén, 2004; B. O. Muthén & Muthén, 2000). Factor mixture models are flexible because of the possibility of analyzing continuous and categorical observed variables in a single model and including covariates effects in different parts of the model. Comparisons of alternative models can be based on statistical tests and indicators of goodness of fit.

Models

This section consists of two parts: a description of the general factor mixture model and a discussion of the differ-

ent ways to include covariates. Throughout this section, the following hypothetical example is used as an illustration. Suppose a math test consists of several items, and high test scores require mastery of a skill. The population from which the sample is drawn is assumed to consist of masters and nonmasters. Although there is variation in skill level among the masters and the nonmasters, for the sake of this example we assume that masters and nonmasters differ with respect to the mean skill level substantially such that it would be inappropriate to model the underlying latent skill variable as a continuous factor. Hence, masters and nonmasters correspond to two different clusters. In addition to this mean difference with respect to the factor underlying the observed variables, masters and nonmasters may differ with respect to several background variables. One of the background variables is the test takers' attitude toward math. Suppose previous research has indicated that attitude is related to skill level. Attitude is a known source of heterogeneity, and higher levels of attitude are related to a higher skill level. With respect to the other background variables, it is not known whether they are related to the heterogeneity. Finally, the latent classes differ with respect to a distal outcome variable (e.g., passing or failing an admission test).

Both the mean differences in the skill factor and differences with respect to background variables can be investigated with factor mixture models. Factor mean differences may often be the primary focus of a study. However, differences in background variables provide information for an additional characterization of the classes (e.g., what are the characteristics of a master?).

Description of the General Factor Mixture Model

Several different factor mixture models have been proposed in the literature. Yung (1997) presents the confirmatory factor model with structured means as a within-class model. This model is similar to the multigroup confirmatory factor model (Sörbom, 1974), the main difference being that subpopulation membership is unobserved. In Arminger et al. (1999), the focus is on estimating the factor mixture model conditional on covariates such that multivariate normality has to be assumed only conditional on the covariates. The same treatment of covariates can be found in B. O. Muthén and Shedden (1999), who also include categorical outcome variables in their model that are predicted by class membership. Jedidi, Jagpal, and DeSarbo (1997b) propose a full structural equation model as a within-class model. Dolan and van der Maas (1998) extend this work by adding the possibility of imposing linear or nonlinear constraints on the model parameters. The following model description is based on the description of the

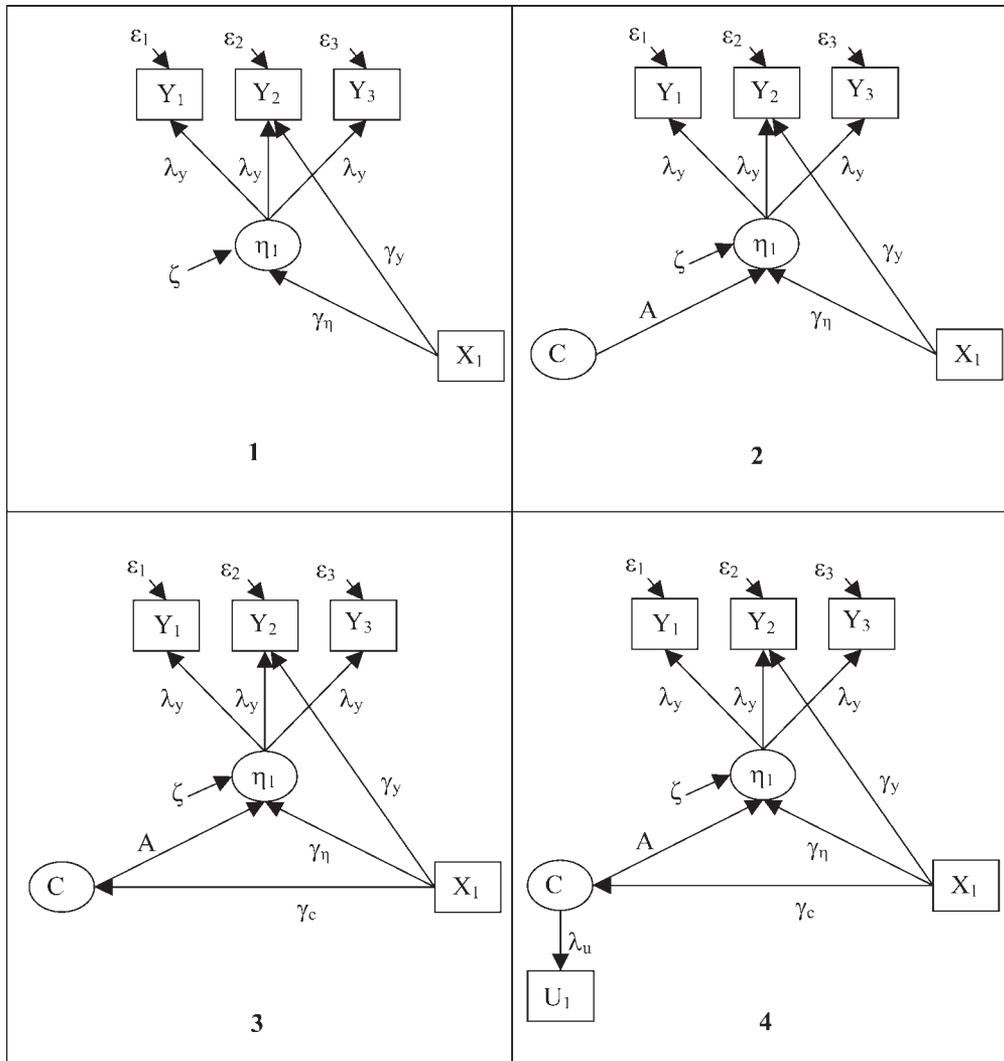


Figure 1. The factor mixture model shown as a stepwise extension of the common factor model.

factor mixture model provided in Muthén and Shedden (1999).¹

The factor mixture model can be decomposed into several parts. Part 1 is the common factor model for a single homogeneous population (Jöreskog, 1971). By adding Parts 2, 3, and 4 to the first part in consecutive steps, it is shown that the factor mixture model is an extension of the common factor model. The stepwise extension of the common factor model is depicted in Figure 1 (panels 1 through 4).

Part 1. The common factor model is a linear regression model in which observed variables are regressed on factors (see Figure 1, panel 1). In addition, observed variables and factors can be regressed on covariates. In what follows, the subscript i is used as an index for variables that vary across participants. An uppercase letter is used to indicate a random variable (e.g., Y is used to indicate an observed variable). The corresponding lowercase letter is used to repre-

sent a realization of the random variable (i.e., y_i is participant i 's score on variable Y). Observed outcome variables, which are always multivariate, are denoted in bold (i.e., matrix notation) as \mathbf{Y} . Covariates may be multivariate or univariate and are denoted in bold as \mathbf{X} .

The regression intercepts are denoted as $\boldsymbol{\nu}$, the regression slopes or factor loadings as $\boldsymbol{\Lambda}_y$, and the regression residuals as $\boldsymbol{\epsilon}$. The regression residuals in the common factor model are the sum of specific factors and measurement error (Meredith, 1993). The direct effect of \mathbf{X} on \mathbf{Y} is conveyed through the regression weight $\boldsymbol{\Gamma}_y$. Scores on the underlying factors are denoted as $\boldsymbol{\eta}$. The usual assumptions of the common factor model apply, including that the residuals

¹ The technically inclined reader may prefer the more thorough presentation in Muthén and Shedden (1999).

have zero autocorrelations and are uncorrelated with the factors. The common factor model is expressed as

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\Gamma}_y \mathbf{x}_i + \boldsymbol{\varepsilon}_i, \quad (1)$$

$$\boldsymbol{\eta}_i = \boldsymbol{\Gamma}_\eta \mathbf{x}_i + \boldsymbol{\zeta}_i. \quad (2)$$

The second equation shows that the factor scores $\boldsymbol{\eta}$ are regressed on covariates \mathbf{X} with regression weights $\boldsymbol{\Gamma}_\eta$. The residual factor score $\boldsymbol{\zeta}_i$ is the part of the factor score that is not explained by the covariates. The common factor model is shown in the first panel of Figure 1.

In the hypothetical example, the math test items are \mathbf{Y} , which measure a single underlying math factor η . Attitude is a single covariate \mathbf{X} . If one of the items is especially sensitive to attitude, this would be captured by a direct effect of \mathbf{X} on that item, and the corresponding γ_y would be nonzero. An effect of attitude on the underlying factor is expressed in γ_η .

Part 2. The common factor model is a model for a single homogeneous population (see Figure 1, panel 2). In the next part, it is extended with a latent class variable to model unobserved population heterogeneity. The latent classes are unordered and modeled using a multinomial variable. For $k = 1, \dots, K$ latent classes, a multinomial variable with $K - 1$ categories is needed, and

$$c_{ik} = \begin{cases} 1 & \text{if participant } i \text{ belongs to class } k \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

In the example, $K = 2$ (masters and nonmasters) such that C reduces to a binomial variable. A participant who is a master would have the score $c = 1$ and a nonmaster would have the score $c = 0$.

The second part of the factor mixture model extends the first part by regressing the factor scores on the latent class variable C (see Figure 1, panel 2). Integrating this extension in Equations 1 and 2 gives

$$\mathbf{y}_{ik} = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_{yk} \boldsymbol{\eta}_{ik} + \boldsymbol{\Gamma}_{yk} \mathbf{x}_i + \boldsymbol{\varepsilon}_{ik}, \quad (4)$$

$$\boldsymbol{\eta}_{ik} = \mathbf{A} \mathbf{c}_i + \boldsymbol{\Gamma}_{\eta k} \mathbf{x}_i + \boldsymbol{\zeta}_{ik}, \quad (5)$$

where the subscript k is attached to parameters that may vary across classes and to random variables that may have class specific distributions (i.e., different means, covariances). Note that the covariates have no subscript k because the model is estimated conditional on the covariates. \mathbf{A} contains the intercepts of the factors for each class and has dimensions Number of Factors \times Number of Classes. In terms of the example, the element in \mathbf{A} corresponding to the masters would be higher than that for the nonmasters. $\gamma_{\eta k}$ conveys the effect of attitude on the skill factor. The subscript k indicates that the effect size may or may not be the same for masters and nonmasters. The covariance matrix of

the residual factor scores, $\boldsymbol{\zeta}_{ik}$, and of the residuals of the observed scores, $\boldsymbol{\varepsilon}_{ik}$, may vary across classes. Note that it is not possible to estimate all class-specific intercepts and factor means simultaneously. The same restrictions apply as in multigroup analyses (Sörbom, 1974).

Part 3. Although a participant is either a master ($c = 1$, see *Part 2*) or a nonmaster ($c = 0$), the class membership of the individual participants is not known (see Figure 1, panel 3). This is integrated in the factor mixture model as follows. The probability of belonging to each of the classes is predicted for each participant using multinomial regression. The class probabilities are computed during the model estimation (for a slightly different approach, see, e.g., Dolan & van der Maas, 1998). For instance, after fitting a factor mixture model to the cognitive test data, a participant may have a 0.8 probability of belonging to the masters' class and a 0.2 probability of belonging to the nonmasters' class. Class membership may be predicted by covariates \mathbf{X} (e.g., attitude increases the probability of belonging to the masters), which is covered in the third part of the model (see Figure 1, panel 3). More precisely, covariates \mathbf{X} predict the log odds of the probability of belonging to a given class compared with the probability of belonging to the (arbitrarily chosen) K^{th} class. The third part of the model is denoted as

$$\ln \left[\frac{P(c_{ik} = 1 | \mathbf{x}_i)}{P(c_{iK} = 1 | \mathbf{x}_i)} \right] = \lambda_{c_k} + \boldsymbol{\Gamma}_{c_k} \mathbf{x}_i. \quad (6)$$

$\boldsymbol{\Gamma}_{c_k}$ contains the regression weights of the covariates. The intercept or location parameter λ_{c_k} can be class specific.

Part 4. The fourth and last part of the model incorporates observed categorical outcome variables that are predicted by class membership (see Figure 1, panel 4). In case the categorical variables are binary, the response categories do not have to be ordered, and the regression on class membership is a logistic regression. Say we have $j = 1, \dots, J$ binary outcomes U , then

$$\ln \left[\frac{P(u_{ij} = 1 | \mathbf{c}_i)}{1 - P(u_{ij} = 1 | \mathbf{c}_i)} \right] = \boldsymbol{\Lambda}_{uj} \mathbf{c}_i. \quad (7)$$

In case of more than two response categories, the response categories have to be ordered and the logistic regression is extended to an ordered polytomous regression (see B. O. Muthén & Shedden, 1999, for details). In the example, $P(u_i) = 1$ is participant i 's probability of passing the admission test. $\boldsymbol{\Lambda}_u$ captures the effect that being a master increases the odds of passing the admission test; hence, the element in $\boldsymbol{\Lambda}_u$ corresponding to the masters would be larger than that for the nonmasters.

It should be noted that the observed continuous variables \mathbf{Y} are assumed to be multivariate normally distributed conditional on class membership and the covariates. Condi-

tional normality is relevant in the context of using test statistics such as the adjusted likelihood ratio statistic (aLRT) for model comparisons (see Step-by-Step Analysis of LSAY Data at a Single Time Point below), which assume within-class normality. Note that violations of conditional normality can have implications for the number of classes that are extracted (Bauer & Curran, 2003a). Conditional normality is also relevant in the context of missing data. Because the model is estimated conditional on the covariates, participants with missing data on covariates are deleted from the analysis. Covariates may or may not be correlated. However, specific relations between covariates are not explicitly specified and estimated. To accommodate for an interaction between two covariates, it is necessary to include a variable representing the interaction (i.e., a product of the two interacting covariates).

In sum, the general factor mixture model features three types of observed random variables, namely continuous outcome variables \mathbf{Y} (Equation 4), categorical outcome variables \mathbf{U} (Equation 7), and covariates \mathbf{X} , which can be continuous or categorical. Furthermore, there are four types of unobserved random variables, namely the categorical latent class variable C (Equations 3, 5, 6, and 7), the continuous factor scores $\boldsymbol{\eta}$ (Equation 4), the residual factor scores $\boldsymbol{\zeta}$ (Equation 5), and the residual scores $\boldsymbol{\varepsilon}$ (Equation 4). Scores on the random variables vary across participants. Regression intercepts and regression weights, on the other hand, are fixed (i.e., they do not vary across participants within a given latent class). The general factor mixture model includes as submodels the common factor model, in which the number of classes equals 1, the classic latent class and latent profile model, which can be derived as a factor mixture model with zero factor variance, and the growth mixture model (Lazarsfeld & Henry, 1968; B. O. Muthén, 2004). Although the approach described in the remainder of the article is also applicable to these submodels, we focus on exploring heterogeneity with factor mixture models at a single time point.

Different Ways to Specify Effects of Covariates X and Class Variable C

The covariate effects on observed variables, within-class common factors, and the class variable may vary across classes. For instance, attitude may have a strong effect on the skill level of the masters but explain very little variance of the skill level of the nonmasters. Parameters of the within-class common factor model (i.e., intercepts, factor loadings, and residual variances) may also be class specific. Class-specific effects are closely related to the concept of measurement invariance. This section, therefore, starts with an overview of measurement invariance. The different covariate effects and other class-specific effects in the factor

mixture model are discussed with respect to measurement invariance.

Measurement invariance. If subpopulations are to be compared, one has to ensure that the observed variables on which the comparison is based are measurement invariant across the subpopulations. Measurement invariance (MI) as defined by Mellenbergh (1989) and Meredith (1993) implies that, given a certain score on the underlying factor, an observed score does not depend on subpopulation membership. Meredith has shown that MI holds only if the measurement model relating the observed variables to the underlying factors is identical across subpopulations (see Lubke, Dolan, Kelderman, & Mellenbergh, 2003, for a less technical account of MI). In the context of the common factor model, three restrictions have to be imposed on the model to achieve MI: equality of factor loadings, intercepts, and residual variances across subpopulations. If either of the three restrictions is not tenable, MI, as defined, is absent. Absence of MI implies that the observed measures or indicators may not be measuring the same constructs in the various subpopulations.

Meredith (1993) has defined three increasingly restrictive levels of factorial invariance by stepwise adding the three restrictions. Weak factorial invariance requires equality of factor loadings, strong factorial invariance adds the restriction of equal intercepts, and strict factorial invariance corresponds to the combination of all three restrictions. Because all three restrictions have to hold, only strict factorial invariance ensures MI (see Meredith, 1993, for examples in which even strict factorial invariance fails to imply MI). In what follows, the three levels of factorial invariance and their implication for MI are briefly reviewed.

Weak factorial invariance holds if factor loadings are identical across subpopulations. The common factor model is a linear regression model, and a factor loading is the slope of the regression of an observed variable Y on a factor η . If subpopulations differ with respect to a factor loading, then a unit increase in the factor score does not result in the same increase of the observed score across subpopulations. The differential increase in observed scores is an indication that one or more factors not included in the model augment (or attenuate) the observed scores in a subpopulation-specific way and that this effect increases (or decreases) with increasing factor scores. The observed variable measures different underlying factors in the subpopulations. Therefore, absence of weak factorial invariance violates MI.

Strong factorial invariance adds the requirement of equal intercepts to the equality of factor loadings. If subpopulations differ with respect to the intercepts of the regression of an observed variable Y on a factor η but not with respect to the factor loadings, then one subpopulation scores consistently higher than the other independent of the scores on the factor η . The observed mean differences between subpopulations are, therefore, not entirely due to mean differences in

the factor η , but they are at least partially due to the effects of factors not explicitly included in the model. Hence, absence of strong factorial invariance is also a violation of MI.

Strict factorial invariance adds the restriction of equal residual covariances to the equality of loadings and intercepts. In the common factor model, the residual variance is the sum of specific factor variance and random measurement error variance. These two sources of variance cannot be separated. Differences in residual variance can, therefore, be due to differences in specific factor variance, random measurement error variance, or a combination of the two. Differences in specific factor variance indicate that specific factors boost or attenuate individual differences in observed scores differentially across subpopulations. The observed scores may not measure the same constructs across subpopulations; hence, absence of strict factorial invariance violates MI.

Note that strong factorial invariance is often considered to be a sufficient basis for the comparison of subpopulations (Little, 1997; Widaman & Reise, 1997). This is due to the theoretical consideration that even in the presence of residual variance differences (a) loading equality ensures that factor variances and covariances can be compared across subpopulations and (b) intercept equality ensures that observed mean differences are due to factor mean differences. In the factor model, the residuals, which contain possible effects of specific factors, are assumed to have zero means in all subpopulations. Consequently, if observed mean differences are due to mean differences in specific factors in addition to factor mean differences, the specific factor mean differences would be manifest in terms of intercept differences. Hence, intercept equality implies absence of specific factor mean differences. In a practical situation, however, equality restrictions are evaluated based on likelihood ratio tests and measures of goodness of fit. Lubke and Dolan (2003) have shown that unrestricted residual variances across subpopulations may leave specific factor mean differences undetected. In the absence of substantive theory that would support differences in residual variances in a given study, it seems advisable not to neglect the third requirement of MI.

Assuming MI, differences in means of Y between subpopulations are exclusively due to differences in common factor means, and differences in covariances of Y are exclusively due to differences in common factor covariances. Because, in case of MI, the measurement model relating observed variables to the factors is the same for all subpopulations, the interpretation of the factors is the same for all subpopulations. This fact greatly simplifies comparisons between subpopulations.

Partial MI (PMI) refers to the absence of MI in one or more observed items of a test in which the remaining

items are invariant (Byrne, Shavelson, & Muthén, 1989). The interpretation of subpopulation differences is more cumbersome in case of PMI and depends on whether PMI concerns absence of loading equality, intercept equality, residual variance equality, or a combination of the three. As shown previously, in case of absence of intercept equality, higher scores on the noninvariant observed variable are not necessarily due to higher scores on the construct of interest. However, such a model may still be highly interesting in an exploration of subpopulation differences because it reveals a subpopulation difference in the response to the noninvariant item. To limit the number of model comparisons, one may compare measurement-invariant models with models in which the intercepts of an entire scale are unrestricted. This approach is illustrated in the analysis of the Longitudinal Study of American Youth (LSAY) data.

Class-specific effects. Seven different possible class-specific effects correspond to the parameters with a subscript k in Equations 4, 5, and 6. They are labeled in Figure 2 with numbers 1 through 7 and are discussed in terms of the example described in the beginning of this section.

The first three effects concern class-specific factor loadings, intercepts, and residual variances. The implication with respect to MI of these effects was discussed previously.

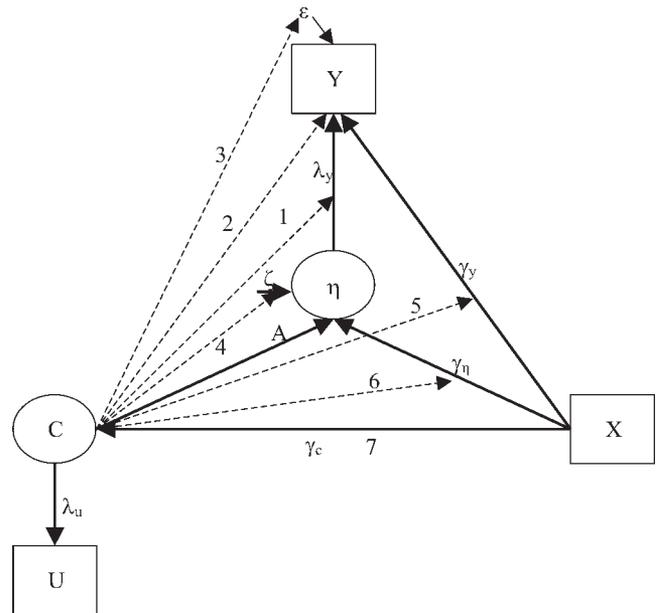


Figure 2. Different effects of covariate X and class variable C . Solid lines indicate class-invariant parameters. Dashed lines indicate the possibility to specify class-specific parameters.

Path 1. Path 1 refers to class-specific factor loadings. Using the example, a higher factor loading for the masters, for example, means that skill level interacts with being a master versus a nonmaster; with an identical increase in skill level, the observed scores of the masters would increase more than those of the nonmasters.

Path 2. Path 2 indicates class-specific intercepts. A class-specific intercept occurs if all masters have the same advantage on one of the items and if this advantage cannot be explained by a higher score on the skill factor or by the covariates \mathbf{X} .

Path 3. Path 3 represents class-specific residual variances, which arise if scores of masters vary more (or less) than the scores of nonmasters because of a difference in variation of specific factors or measurement error.

Path 4. Path 4 indicates the effects of the class variable on the common factors. These effects serve to model factor mean and covariance differences between masters and nonmasters that are not accounted for by one or more covariates. Effects of the class variable on the common factors do not violate MI. In fact, the comparisons of classes with respect to factor means and covariances are often the focus of a study.

There are three different possible ways to include covariate effects. All covariate effects can be specified as class-invariant or as class-specific effects. First, covariates can explain within-class variation by directly influencing observed variables \mathbf{Y} through regression coefficient γ_y . Second, covariates can influence the observed variables indirectly through latent continuous variables $\boldsymbol{\eta}$, which is modeled using coefficient γ_η . Third, covariates can explain between-class variation by influencing the latent class variable C , which is conveyed through coefficient γ_c .

Path 5. The path γ_y is the direct effect of the covariate on an observed variable. It indicates that within class some of the variation in the observed variable is due to variation in X . Equally, the mean of the observed variable depends at least partially on the covariate. The interpretation in terms of the hypothetical example is as follows. Attitude influences performance on one (or more) of the test items in addition to the influence of the latent cognitive skill factor. If this path is specified to be class invariant, the latent classes can still be compared in a straightforward way with respect to their factor intercepts (e.g., masters and nonmasters differ with respect to their skill level, there is no difference between masters and nonmasters in the influence of attitude on the particular test items). If, alternatively, the effect of the covariate is class specific, then there is noninvariance across classes. Path 5 means that there is an interaction between the class variable and the covariate with respect to the effect on the

observed variable. The measurement model, which relates observed variables to the factors, is not the same across classes; hence, class comparisons are more complicated to interpret. In terms of the example, masters and nonmasters differ with respect to skill, and the effect of attitude on one of the test items may be high for the nonmasters but almost absent for the masters.

Path 6. The path γ_η specifies an effect of the covariates on the factor means and covariances. In other words, attitude has an effect on the skill level. If this effect is class invariant, then the effect is the same for the masters and the nonmasters and there is no violation of MI. If the effect is allowed to vary across classes, we have an interaction between the class variable and the covariate with respect to the effect on the factor. This is not a violation of MI because the measurement model relating \mathbf{Y} to $\boldsymbol{\eta}$ does not vary across classes.

Path 7. The path γ_c concerns the prediction of the odds of belonging to one class versus belonging to a reference class. For instance, in a three-class model, there are two such regression coefficients corresponding to (a) the odds of belonging to Class 1 versus Class 3 and (b) the odds of belonging to Class 2 versus Class 3. It is possible to restrict these covariate effects to be equal for all possible odds, but usually the effect will differ.

Although the possibility of specifying different class-specific effects is clearly an advantage, it is also obvious that the interpretability of a model can rapidly decrease with an increasing number of effects. Ideally, decisions to include or not include class-specific effects in the model are theory driven. For instance, attitude may be expected to have an impact on the variation in skill level within the masters and the nonmasters classes and on the skill level differences between masters and nonmasters. Consequently, the factor and the class variable would be regressed on attitude. A covariate such as course participation may be important mainly in explaining differences between masters and nonmasters. In that case, only the class variable is regressed on the covariate. In an exploratory analysis, in which theory with respect to the sources of population heterogeneity is less specific, a series of models may be considered in which increasingly restrictive models are fitted to the data. Initial models may include effects of the covariates on both the within-class factors and the latent class variable. To explore mean differences between classes, which are generally a main interest of a mixture analysis, one may compare models with class-specific intercepts (i.e., partially invariant models) with models with class-invariant intercepts. This approach is illustrated in the analysis of the LSAY data.

Step-by-Step Analysis of LSAY Data at a Single Time Point

General Outline

The step-by-step analysis in this section provides an illustration of how population heterogeneity may be investigated with factor mixture models. The latent class variable serves to model unknown heterogeneity, whereas variables that are known to induce heterogeneity are included as covariates. In the current approach, the researcher has to specify the number of latent classes for each model beforehand. During the model estimation, not only are the parameters of the model estimated but the posterior probabilities with which each participant belongs to each of the classes are calculated. Hence, the results consist of the model parameters such as within-class factor loadings and intercepts, factor mean differences between classes and so on, and the posterior class probabilities for each participant. For instance, the resulting class probabilities for the first participant in a study may be .85, .10, and .05 of belonging to Classes 1, 2, and 3, respectively, and for the second participant .95, .05, .0, and so on. On the basis of their highest probability, participants can be assigned to one of the classes. This classification is the basis for post hoc analyses, which serve to compare classes, and to decide whether an additional class is informative.

Fitted models. In the main analysis below (see *Factor Mixture Modeling*), the within-class models are confirmatory factor models. Hence, the analysis is not exploratory with respect to the factor structure of the test. The pattern of factor loadings is assumed to be known from previous research. The analysis is exploratory with respect to the sources of population heterogeneity. Commonly, a key interest in an exploration of population heterogeneity is the investigation of mean differences between latent classes.

In the analysis of the LSAY data, we, therefore, fit four models with increasing restrictions on the model for the observed means within class. In all four models, loadings are class invariant such that factor variances and covariances can be compared across classes. Residuals are equally class invariant. Model 1 imposes no restrictions on the means of the observed variables (i.e., the intercepts of all indicators of the math and science factor are class specific). This is the noninvariant model. Next, two different partially invariant models are specified. In Model 2, all intercepts of the indicators of the math factor are specified to be class specific, but the science scale is specified to be measurement invariant. If this model holds, classes can be compared with respect to the means of the science factor. The class differences in the individual indicators of the math factor can be due to other factors in addition to the math factor and have to be interpreted accordingly. In Model 3, the pattern is reversed and the math scale is specified to be invariant

whereas the science scale is noninvariant. Model 4 is a fully measurement invariant model in which factor loadings, intercepts, and residual variances are specified to be equal across classes.

The main analysis is followed by a post hoc comparison with respect to background variables. Intercept differences may be due to mean differences in specific factors. To detect potential specific factors (or variables related to specific factors), it is useful to compare classes with respect to available background variables that had not been included in the main analysis and to investigate how classes differ with respect to these background variables. If classes differ substantially with respect to a background variable, it can be included as a covariate in a subsequent analysis.

In an empirical study, often large numbers of background variables are measured. Only some of these variables are known to induce heterogeneity, whereas others are measured for exploratory purposes. On the basis of a simulation study by Lubke and Muthén (2003), we propose to include only those variables as covariates in the factor mixture models that are expected to have a significant effect and to use all other background variables in post hoc analyses. The simulation study has shown that covariates with medium to large effects on class membership improve model performance in terms of correct class assignment and parameter recovery. Including large numbers of categorical covariates may lead to convergence problems, which are likely due to low frequencies of some combinations of response categories. In the current illustration, the choice of variables that are included as covariates in the main analysis and variables that are used in the post hoc analysis is entirely arbitrary.

Model comparisons. Models are fitted with an increasing number of classes, and the fit of the different models is compared. Regarding mixture model comparisons, the following issue deserves consideration. Model misspecifications can correspond to the imposed restrictions within class (e.g., restrictions of measurement invariance) or to the number of classes. To achieve an improvement in model fit, restrictions within class can be relaxed or the number of classes can be increased. Hence, model comparisons have to be carried out both between models with different numbers of classes and between models that differ with respect to their restrictions. Models that impose more restrictions on the means and covariances within class imply a different distribution of the data than a more lenient model with the same number of classes. More restrictive within-class models often require the addition of more classes than models that are more lenient. For instance, the fully invariant model may require more classes than the more lenient partially invariant models. In that case, models with different numbers of classes need to be compared. Models with different numbers of classes cannot be compared using the usual

likelihood ratio test because regularity conditions are not met (Cramér, 1946; McLachlan & Peel, 2000).²

There are different approaches to compare models and to decide on the number of classes (see, for instance, Nagin, 1999). Commonly used are fit indexes such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), sample size-adjusted BIC (aBIC), and aLRT. These indexes are described in Akaike (1974), Schwarz (1978), L. K. Muthén and Muthén (2001), and Lo, Mendell, and Rubin (2001), respectively. Because AIC, BIC, and aBIC have different penalties on the number of parameters and adjust differently for sample size, different decisions may be reached based on those criteria. In addition to the information criteria and the aLRT, we approach the comparison of models also from a more content-oriented point of view. An additional class may reveal an interesting subpopulation. However, an additional class may also result in the splitting of a well-interpretable latent class into two poorly interpretable classes. The utility of an additional class with respect to substantive theory can be assessed by comparing classes with respect to factor means or intercepts of observed variables and the means of the covariates. The stability of a given class across solutions with increasing numbers of classes can be traced with transition matrices. Transition matrices show how many participants from a given class of, for example, a two-class model are assigned to each of the classes in a three-class model.

In the following analysis, the numbers of factors of the test are assumed to be known, and confirmatory within-class models are specified. If the measurement model that relates observed variables to underlying factors is unknown, it is possible to specify exploratory factor models with increasing numbers of factors as within-class models (McLachlan & Peel, 2000). The exploratory factor mixture model is a submodel of the general factor mixture model described previously. In the exploratory factor mixture model, both the factor structure and the sources of population heterogeneity are explored simultaneously. Measurement models may also differ with respect to their factor structure across classes (e.g., see Dolan & van der Maas, 1998). Note that if the factor structure within class has to be investigated, the possibility of specifying equivalent models has to be taken into account. Equivalent models imply the same distribution but differ with respect to the structure they impose on the data and may differ considerably with respect to their conceptual interpretation. In the context of models for data from single homogeneous populations, the problem of equivalent models has led to a considerable body of research (see, for instance, Hershberger, 1994; Luijben, 1993; MacCallum, Wegener, Uchino, & Fabrigar, 1993; Raykov & Marcoulides, 2001). In the context of mixture models, additional problems may arise because models for different numbers of classes may imply the same mixture distribution (for examples, see McLachlan & Peel, 2000). Addressing the

problem of equivalent models in detail, however, is beyond the scope of this study.

Data Description

Starting in 1987, LSAY data have been collected for two cohorts through 1994 (Miller, Kimmel, Hoffer, & Nelson, 2000). Both cohorts are national samples. As part of the LSAY, a math and a science achievement test was administered each year. For current purposes, data from the second cohort obtained during the last year at high school are analyzed. More specifically, we use the scores derived by Miller et al. (2000) based on item response theory (IRT) for the four math subscales (i.e., basic mathematics, algebra, geometry, and quantitative literacy) and the IRT scores computed for the three science subscales (i.e., biological, physical, and environmental sciences). The measurement model has a two-factor structure, with math subscales loading on a math factor and science subscales loading on a science factor. The model has a single cross-loading (i.e., environmental sciences has a small negative loading on the math factor). Gender and urban status are treated as known sources of heterogeneity and are incorporated as covariates in the factor mixture models. Urban status indicates whether a student lives in a suburban, urban, or rural environment. Urban and rural students do not differ significantly with respect to their math and science subscale scores (Miller et al., 2000). Both urban and rural students score lower than suburban students. Hence, a binary covariate is created: urban–rural versus suburban. The sample size in the illustration is 982 participants. As mentioned, factor mixture models are estimated conditional on covariates; therefore, the participants included in this illustration have no missing data on the covariates. Coincidentally, there are also no missing data on the math and science subscales.

The LSAY data have a multilevel structure because of the sampling design (e.g., students within classes within schools within neighborhoods). In the current study, we initially used the so-called sandwich estimator to compute parameter estimates and their standard errors (Amemiya, 1985; Zeger & Liang, 1986). This estimator accounts for the fact that observations are not independent. Because the results were very similar to those using the maximum likelihood estimator, only the latter results are presented. For more in-depth description of the design of the LSAY, the

² Suppose H_0 corresponds to a model with k classes, which is to be tested against a model with $k + 1$ classes. The k -class model can be specified to have one of the mixing proportions equal to zero. All other parameters being equal, it can also be specified by letting the means of two classes be equal. Hence, true values of parameters under H_0 are on the boundary of the likelihood and also in a nonidentifiable subset of the parameter space, which renders regularity conditions invalid (see McLachlan & Peel, 2000, p. 187).

sampling scheme, and other details, the reader is referred to the LSAY manual (Miller et al., 2000). Model fitting is carried out with Mplus 2.1 (L. K. Muthén & Muthén, 2002). Post hoc class comparisons are done in S-PLUS (S-PLUS, 2001) but may also be done in any spreadsheet program that allows for the computing of mean values.³

Factor Mixture Modeling

The distribution of the covariates gender and urban status in the total sample is shown in Table 2. The regression of factor scores on the covariates is estimated in all classes and is specified to be class specific. In other words, it is estimated how much variation in the common factors is due to gender and urban status within each of the classes. The class variable is also regressed on the covariates, which gives an estimate of the between-class variation that is due to gender and urban status.

The proportion females in the total sample is .53 and the proportion urban–rural students is .56.

Tables 3 and 4 show that gender and urban status are indeed inducing heterogeneity in the sample. The means on the math and science subscale scores are higher for the boys than for the girls and also higher for the suburban students than for the urban–rural students.

Results of latent class models, including factor mixture models, can depend on starting values for the parameters. This is due to the possibility that the likelihood has multiple local minima. Depending on the set of starting values, the estimation algorithm may stop at a local minimum and fail to detect the global minimum. The problem of multiple local minima may be especially severe if the covariance matrices are not equal across classes (McLachlan & Peel, 2000). A common way to address this problem is to estimate the model with different sets of starting values and choose the solution with the smallest minimum. In the current analysis, each model is fitted with several different sets of starting values. We provide class-specific starting values for the factor means (if these are not fixed to zero). All other parameters have default starting values, which do not differ across classes. Factor loadings have default starting values equaling 1, whereas intercept starting values are 0. None of the models displayed convergence problems during parameter estimation.

Table 2
Cross-Tabulation of Gender and Urban Status

Gender	Urban–rural	Suburban	Marginal counts
Female	308	209	517
Male	241	224	465
Marginal counts	549	433	982

Results of mixture modeling. To obtain a baseline by which mixture models can be evaluated, a two-factor single-class model was first fitted to the data. In a single-class model, the means cannot be structured. The factor scores are regressed on gender and urban status. These covariates explain 5% of the variance of the math and science factor in the total sample. The factor mixture models can provide a more detailed picture, as shown next.

Two-, three-, four-, and five-class versions of the noninvariant Model 1, the partially invariant Models 2 and 3, and the fully invariant Model 4 are fitted. The results in Table 5 pertain to Models 1, 2, and 4. Model 2 always outperforms Model 3, and the results of Model 3 are, therefore, not presented. In Model 2, the science scale is measurement invariant, but the intercepts of the indicators of the math factor are class specific. The first three columns of Table 5 show the fit indexes AIC, BIC, and aBIC. A model with a lower AIC, BIC, or aBIC is preferred over one with higher indexes. The fourth column shows the aLRT, which provides a test whether deleting a class results in a significantly worse fit (Lo et al., 2001). A significant *p* value of the test confirms, therefore, that the current model fits better than the model with one class less.

Fitting a single-class model results in worse fit measures than any of the two-class models. Equally, the *p* values of the aLRT of all two-class models confirm that two classes provide a better fit than the single-class model. Among the two-class models, the noninvariant Model 1, which is the least restrictive model, fits best. This changes when a third class is added. The indexes indicate that a three-class model should be preferred to the two-class model and that, within the three-class models, the partially invariant model provides the best fit. The results of the four-class models are consistent with respect to the best fitting model, which is again the partially invariant model. For the partially invariant model, the indexes are also in agreement with respect to the need of the fourth class. For the other two models, AIC and aBIC favor the four-class solutions, whereas BIC and aLRT favor the three-class solutions. The five-class models show almost the same pattern. The difference is that BIC and aLRT indicate that four classes suffice not only for the noninvariant and the invariant model but also for the partially invariant model. AIC and aBIC favor the five-class partially invariant model. To complete the model-fitting part of the investigation, a six-class partially invariant model is fitted to the data. The pattern of results remains unchanged; only AIC and aBIC indicate the need of the sixth class.

It is noteworthy that the factor structure is stable across models. The estimates of corresponding factor loadings are

³ An Mplus input file with comments and a file containing artificial data with a structure similar to the LSAY data are available at <http://dx.doi.org/10.1037/1082-989X.10.1.21.supp>

Table 3
Mean Subscale Scores for Gender

Gender	Algebra	Geometry	Quant. lit.	Basic	Env. sci.	Biology	Physics
Female							
<i>M</i>	9.2	8.1	7.0	7.1	6.7	6.8	6.7
<i>SD</i>	2.3	2.1	1.5	1.3	1.3	1.3	1.2
Male							
<i>M</i>	9.3	8.5	7.3	7.2	7.0	7.0	7.1
<i>SD</i>	2.6	2.3	1.6	1.4	1.7	1.5	1.4

Note. Quant. lit. = quantitative literacy; Basic = basic mathematics; Env. sci. = environmental science.

similar in value, and the loadings have the same order regardless of the number of classes a model has or how the observed means are structured. The same holds for the residual variances. The proportion of variance in the observed variables that is explained by the factors (e.g., the observed variable R^2 's) ranges very consistently between 0.4 and 0.8 across all models. Parameter values and standard errors of factor loadings, residual variances, and factor covariances for the four-, five-, and six-class partially invariant models are presented in Table 6.

The four-, five-, and six-class partially invariant models are selected for further investigation. The question whether, compared with the four-class model, the additional classes in the five- and six-class model provide useful information concerning heterogeneity is addressed in the following four ways. First, because the science scale in this model is measurement invariant, classes can be compared with respect to the science factor means. The intercepts of the indicators of the math factor are class specific; hence, class differences in the math subscales cannot be attributed to math factor differences but have to be interpreted as observed mean differences. Observed mean differences may be due to other factors in addition to the underlying math factor. Second, for each of the three models, participants are assigned to the class corresponding to their highest posterior class probability. Based on the assignments, proportions of females and urban-rural students for each class can be computed, which, in turn, can be used to characterize the classes. Third, the class assignments are used to compute the within-class means of a number of behavioral and atti-

tudinal variables that are potential sources of heterogeneity and that are included only in the post hoc analysis: thoughts of dropping out, disciplinary problems, boredom at school, and number of friends who had dropped out of high school. Finally, the class assignments are used to derive transition matrices for the four-class to five-class and five-class to six-class models. These matrices show how many participants of a given class of a model with a given number of classes migrate to each of the classes of a model with an additional class. An indication of a stable class would be if Class 1 of a 4-class model, for example, has almost the same members as one of the classes in the corresponding five-class model. Characterizing classes and tracing stable classes can provide insight whether additional classes are informative.

Table 7 provides the latent class proportions, the range of intercepts of the math subscales and the science factor mean, the within-class proportions of gender and urban status, and the means of thoughts of dropping out, disciplinary problems, boredom at school, and the number of friends who had dropped out of high school for the four-, five-, and six-class partially invariant model. Tables 8 and 9 show the transition matrices for the four-class to the five-class model and the five-class to the six-class model, respectively.

Class 1 of the four-class model has intermediate to high scores on the math subscales and an intermediate score on the science factor. It contains 38% of the participants. About 70% of the members of this class form Class 2 of the five-class model, which, in turn, remains relatively stable

Table 4
Mean Subscale Scores for Urban Status

Urban status	Algebra	Geometry	Quant. lit.	Basic	Env. sci.	Biology	Physics
Urban-rural							
<i>M</i>	8.8	7.9	6.9	6.9	6.6	6.7	6.6
<i>SD</i>	2.4	2.1	1.6	1.4	1.5	1.4	1.3
Suburban							
<i>M</i>	9.5	8.8	7.5	7.4	7.2	7.1	7.1
<i>SD</i>	2.4	2.1	1.5	1.3	1.4	1.4	1.3

Note. Quant. lit. = quantitative literacy; Basic = basic mathematics; Env. sci. = environmental science.

Table 5
Fit Indexes for One- to Six-Class Models

Variable	AIC	BIC	aBIC	aLRT (<i>p</i>)
Single class				
Unstructured means	17,493.7	17,625.7	17,540.0	N/A
Two class				
Noninvariant	17,159.7	17,359.8	17,229.6	.0
Partially invariant	17,194.5	17,385.2	17,361.4	.0
Fully invariant	17,413.0	17,589.1	17,474.7	.0
Three class				
Noninvariant	17,061.3	17,330.2	17,155.5	.0
Partially invariant	16,949.9	17,199.2	17,037.3	.0
Fully invariant	17,339.9	17,560.0	17,417.1	.0
Four class				
Noninvariant	17,032.2	17,369.6	17,150.4	.37
Partially invariant	16,881.6	17,189.7	16,989.6	.03
Fully invariant	17,233.4	17,497.5	17,326.0	.48
Five class				
Noninvariant	16,948.5	17,354.3	17,090.7	.68
Partially invariant	16,847.4	17,213.8	16,975.6	.84
Fully invariant	17,199.0	17,507.0	17,307.0	.64
Six class				
Partially invariant	16,795.6	17,221.0	16,944.7	.30

Note. Lower values of the information criteria AIC, BIC, and aBIC indicate better fitting models. AIC = Akaike Information Criteria; BIC = Bayesian Information Criteria; aBIC = adjusted BIC; aLRT = adjusted likelihood ratio test statistic.

and appears as Class 2 of the six-class model. The proportion of females and urban-rural students is higher than in the total sample.

Class 2 of the four-class model excels in the math subscales and the science factor. The class is almost identical to Class 4 of the five-class model and to Class 4 of the six-class model. This class consists of about 40% of the sample and has slightly more boys and clearly more suburban students than the general sample.

Class 3 of the four-class model can be characterized as the low-scoring class in both math and science and contains 16% of the participants. It largely contributes to Class 3 of the five-class model, and Class 6 of the six-class model draws almost entirely from this class. This class has a high percentage of urban-rural students. The proportion of females is the same as in the general sample, although it drops slightly from the four-class to the five-class to the six-class model.

Class 4 of the four-class model is a small class (6% of the sample). The class scores low on science and slightly higher on the math subscales than the low-scoring class. The class is almost identical to Class 5 of the five-class model and to Class 1 of the six-class model. The proportion of females is

low, and the proportion of urban-rural students is somewhat higher than in the total sample.

Taking into account the classes that remain relatively stable when the four-class model is extended to a five-class model, it appears that the additional class in the five-class model is Class 1. Class 1 draws 30% of the Class 1 members of the four-class model and some Class 3 members. This new class consists mainly of girls and urban-rural students, with low science scores and medium math scores. The class does not remain stable when the model is extended to a six-class model. About two thirds migrate to Class 5 of the six-class model, whereas one third goes to Class 2 (see Table 9).

The additional classes of the six-class model are Classes 3 and 5; they cannot be clearly traced back to one of the four classes of the four-class model. Class 3 is a small class that draws from Classes 2 and 4 of the five-class model. It consists mainly of boys, and the proportion urban-rural is not much lower than in the general sample. Class 5 draws from Classes 1 and 2 of the five-class model. It consists of 11% of the sample and has mainly girls and urban-rural students; scores on the math scales and the science factor are slightly higher than those of the low-scoring class.

Table 10 shows the percentage of variance of the factor scores explained by gender and urban status within each class. In the single-class model, gender and urban status explain 5% of the variance. The general result in the four-, five-, and six-class partially invariant models is that gender and urban status explain more than 10% of the variance of the math and science factor but only in low-scoring classes. These are Class 3 in the four-class model, which corresponds to Class 4 in the five- and six-class models, the additional class (e.g., Class 1) of the five-class model, and the additional classes of the six-class model (e.g., Classes 3 and 5). The pattern of covariate effects is consistent with the results of tracing classes across the four-, five-, and six-class models. Absence of similar covariate effects in the other classes does not seem to be due to a restriction of range, which can be deduced from the within-class proportions of the covariates shown in Table 7.

Considering the variables that are included only in the post hoc analysis, one can see that the scores on the covariates thoughts of dropping out, disciplinary problems, and number of friends who have dropped out are highest in the low-scoring class across all three models. The low-scoring class consists of mainly urban-rural students, and the proportion of girls is similar to that for the total sample. Hence, one might conclude that gender is less important than urban status when it comes to the three post hoc variables. Adding classes to the four-class model results in additional classes with low scores in science or both math and science. Generally, these classes have more extreme within-class proportions of gender and urban status than the classes of the four-class model. Adding classes reveals a more fine-

Table 6
Class-Invariant Parameters in the Partially Invariant Four-, Five-, and Six-Class Models

	4-class model		5-class model		6-class model	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Factor loadings on the math factor						
ALG	1.00	0.00	1.00	0.00	1.00	0.00
GEO	0.89	0.09	0.85	0.09	0.76	0.09
QLT	0.65	0.08	0.70	0.08	0.67	0.07
BAS	0.55	0.06	0.55	0.06	0.52	0.07
ENV	-0.38	0.09	-0.33	0.07	-0.33	0.08
Factor loadings on the science factor						
PHY	1.00	0.00	1.00	0.00	1.00	0.00
ENV	1.34	0.08	1.29	0.06	1.24	0.05
BIO	0.99	0.03	0.99	0.03	0.99	0.03
Covariance between the math and science factors						
	0.38	0.05	0.34	0.04	0.28	0.04
Residual variances						
ALG	0.81	0.07	0.82	0.06	0.79	0.07
GEO	0.49	0.03	0.47	0.03	0.48	0.03
QLT	0.26	0.02	0.24	0.02	0.23	0.02
BAS	0.16	0.01	0.16	0.01	0.16	0.01
ENV	0.17	0.03	0.19	0.03	0.16	0.03
BIO	0.39	0.02	0.39	0.02	0.39	0.02
PHY	0.17	0.02	0.17	0.02	0.17	0.02
Residual factor variances						
Math	0.53	0.10	0.50	0.09	0.52	0.09
Science	0.69	0.05	0.61	0.04	0.52	0.05

Note. ALG = algebra; GEO = geometry; QLT = quantitative literacy; BAS = basic mathematics; ENV = environmental science; PHY = physics; BIO = biology.

grained picture concerning the variable disciplinary problems. The characteristics of the additional classes in the five- and six-class models indicate that disciplinary problems are related to gender. The additional Class 1 of the five-class model and Class 5 of the six-class model, which draws from Class 1 of the five-class model, consist mainly of girls and urban-rural students. These classes have lower scores on three of the post hoc variables. However, Class 3 of the six-class model, which has a high proportion of boys but a proportion of urban-rural students that is only slightly lower than in the total sample, has an even higher score on disciplinary problems than the relatively stable low-scoring class in all models (e.g., Class 3 in the four-class model, Class 4 in the five- and six-class models).

In this analysis, we use different ways to gain insight concerning heterogeneity, including information criteria and the aLRT, class comparisons with respect to their means, post hoc comparisons on additional variables, and transition matrices. To summarize the results, based on BIC and aLRT, the four-class partially invariant model would be chosen. AIC and aBIC continue to decrease when classes are added. Transition matrices show that three classes of the four-class model remain largely stable when extending the

model with additional classes. Additional classes in five- and six-class models are mainly derived by splitting up the fourth class. Post hoc comparisons of the classes using additional variables show that not much information is gained by adding a fifth or sixth class, with the exception of the variable disciplinary problems. The latter piece of information may be useful in future research. Taken together, and considering the general aim of model parsimony, the four-class partially invariant model seems to provide a suitable description of the heterogeneity in the current sample.

Discussion

Factor mixture models are a tool to investigate population heterogeneity. Compared with a single-class analysis, a factor mixture analysis can provide a detailed description of clusters of participants within a sample. In the illustration, the single-class analysis showed only the very general result that gender and urban status had an effect on the math and science factors. The factor mixture analysis resulted in clusters with specific patterns of means of observed scores

Table 7
Characteristics of the Four- to Six-Class Partially Invariant Model

Class	Class proportion	Female	Urb-rur	Self-drop		Problem		Bored		Friends drop		Math	Science
				<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Total population <i>N</i> = 982													
		.53	.56	.05	.23	.07	.26	0.42	.49	1.38	.71		
Four-class model													
1	.38	.59	.60	.05	.22	.07	.25	0.40	.49	1.38	.69	6.5–8.9	0.47
2	.40	.49	.46	.03	.16	.07	.25	0.44	.50	1.19	.50	8.0–10.4	1.13
3	.16	.53	.71	.12	.32	.10	.30	0.42	.50	1.81	.99	5.0–6.1	0.03
4	.06	.38	.54	.11	.31	.11	.31	0.29	.46	1.47	.65	6.2–6.4	0
Five-class model													
1	.11	.70	.77	.04	.20	.02	.14	0.33	.47	1.37	.64	5.7–8.1	–0.44
2	.30	.57	.52	.05	.23	.08	.27	0.44	.50	1.39	.71	6.8–9.1	0.59
3	.15	.50	.70	.12	.33	.12	.32	0.41	.49	1.84	.99	5.0–6.0	–0.11
4	.38	.48	.47	.03	.17	.07	.26	0.44	.50	1.18	.50	7.8–10.5	1.00
5	.06	.35	.60	.09	.29	.09	.29	0.36	.48	1.46	.66	6.0–6.4	0
Six-class model													
1	.07	.41	.66	.04	.20	.09	.28	0.32	.47	1.53	.71	5.7–6.2	0.50
2	.28	.59	.54	.05	.22	.07	.26	0.44	.50	1.38	.70	7.0–9.4	1.13
3	.04	.30	.53	.12	.33	.15	.37	0.31	.47	1.52	.99	7.5–9.7	0.30
4	.36	.48	.44	.02	.16	.06	.25	0.45	.50	1.16	.44	7.9–10.7	1.74
5	.11	.71	.79	.06	.24	.03	.17	0.32	.47	1.36	.67	5.5–7.8	0.55
6	.14	.49	.68	.13	.34	.11	.32	0.40	.49	1.86	.99	4.8–5.7	0

Note. The science factor means should be interpreted in terms of mean differences compared with the last class, which has the science factor mean fixed at zero. Standard errors of the intercepts typically range between 0.1 and 0.5. Urb-rur = urban-rural.

and factors and with differing proportions of females and urban students.

Factor mixture modeling is a different approach to investigating heterogeneity than multigroup modeling. Multigroup models are applicable if a sample is divided into explicitly defined groups based on one or more observed variables such as gender. The purpose of multigroup models is to compare these groups with respect to their scores on a given test or questionnaire. The purpose of latent class models, including factor mixture models, is different. The clustering of the participants into the latent classes is inferred from the observed scores, such as the math and science scores, gender, and urban status in the illustration. The latent class approach is appropriate if the interest is in detecting and characterizing clusters of participants (e.g., low-scoring students in an educational study or a class of participants with high scores on an anxiety test).

Guidelines regarding the sample size needed for an exploratory analysis of population heterogeneity are difficult if not impossible to provide. An adequate sample size depends on several different factors, including the number of classes, the restrictiveness of the within-class model, the within-class model complexity, the quality of covariates, and the

reliability of the observed data within class. This situation is not different from common factor analysis, in which MacCallum, Widaman, Zhang, and Hong (1999) have shown that simple rules of thumb for the minimum sample size needed for the recovery of factors are inappropriate because an adequate sample size depends strongly on characteristics of the variables and the study design.

Table 8
Transition Matrix: Four Classes to Five Classes

Class	Class 1	Class 2	Class 3	Class 4
1	96	0	12	1
2	273	19	1	0
3	0	0	147	0
4	0	376	0	0
5	7	0	3	47

Note. Transition matrices show how many participants from a given class of a *k*-class model are assigned to each of the classes in a *k* + 1-class model and, therefore, reveal the stability of classes across solutions with increasing numbers of classes. Columns correspond to the *k*-class model and rows to the *k* + 1-class model. For instance, Class 4 in the four-class model consists of 1 + 47 = 48 participants, 1 participant is assigned to Class 1 of the five-class model, and the remaining 47 participants are assigned to Class 5 of the five-class model.

Table 9
Transition Matrix: Five Classes to Six Classes

Class	Class 1	Class 2	Class 3	Class 4	Class 5
1	2	0	8	0	49
2	35	243	0	8	6
3	0	7	0	23	0
4	0	11	0	345	0
5	72	32	2	0	1
6	0	0	137	0	1

Note. Columns correspond to the five-class model and rows to the six-class model.

Factor mixture models have the advantage over other latent class and regression models that the factor structure of a test or questionnaire is explicitly modeled and that the observed scores are decomposed into factor scores and a residual containing specific factors and measurement error. Factor mixture models are broadly applicable to data from tests or questionnaires that are designed to measure theoretical constructs that can be represented by the factors of a factor model. In the current study, factor mixture modeling is demonstrated for data collected at a single time point. The general factor mixture model described in the model section includes the growth mixture model and can, therefore, equally be applied to longitudinal data to detect latent classes that differ with respect to their growth trajectories (B. O. Muthén, 2001). To compare growth trajectories of latent classes with respect to factors that do not contain measurement error, it is necessary to analyze multivariate outcomes at each time point (Hancock, Kuo, & Lawrence, 2001). It is noteworthy that more general mixture models than the factor mixture model have been proposed. These include full structural equation models in which relations between latent variables can be investigated (see, for instance, Jedidi et al., 1997b). An extension of growth mixture models, including switching between growth curves, has been described by Dolan, Schmittmann, Lubke, and Neale (2005).

Related to modeling the factor structure is the possibility to test whether a questionnaire is measurement invariant across classes. Evidence of measurement invariance almost

certainly implies that a test or questionnaire measures the same underlying factors across classes and greatly simplifies the interpretation of results (Lubke et al., 2003; Meredith, 1993). The interpretation of partially measurement-invariant models is more cumbersome (Lubke et al., 2003). Measurement invariance with respect to the latent class variable can be investigated by comparing noninvariant models with more restrictive partially or fully invariant models. The comparisons can be guided by measures of goodness of fit such as AIC, BIC, and aLRT.

In the illustration, we focused on different models for mean differences between classes because these are often the key interest of a mixture analysis. However, it may also be of interest to compare models with different covariate effects. In an actual study of the LSAY data, one of the aims may have been an investigation of an interaction effect of gender and urban status on class membership and math and science performance. When interpreting the results of a factor mixture analysis with covariates, it is important to realize that the assigned class membership of a participant is model dependent and not an innate quality of the participant. Consider a model for a depression questionnaire in which class membership is predicted by gender and a second model for the same test data in which the class variable is predicted by social economic status. The class variable has a different interpretation in the two models: It describes a different kind of heterogeneity. In the first case, clusters are formed with respect to the depression factor and gender and in the second case with respect to the depression factor and social economic status. The class membership of a test taker is not necessarily the same in the two models.

Related to the interpretation of the class variable is the question of how to decide on the number of classes. Determining the number of classes is the subject of a still-growing body of research and has led to interesting discussions (Bauer & Curran, 2003a; Bauer & Curran, 2003b; Cudeck & Henly, 2003; B. O. Muthén, 2003; Rindskopf, 2003). Mixture modeling can serve not only to detect clusters of participants in a population but also to model distributions that are, for instance, skewed (Pearson, 1894; Pearson, 1895; see also McLachlan & Peel, 2000; Titterton,

Table 10
Variance of Math and Science Factor Scores Explained by Covariates

Class	Math c4	Science c4	Math c5	Science c5	Math c6	Science c6
1	0.03	0.05	0.02	0.15	0.00	0.03
2	0.16	0.11	0.01	0.02	0.03	0.04
3	0.04	0.06	0.05	0.08	0.23	0.47
4	0.00	0.03	0.18	0.11	0.12	0.12
5	0	0	0.03	0.02	0.16	0.00
6	0	0	0	0	0.00	0.04

Note. Columns correspond to the factors in the three models (e.g., Math c4 is the math factor of the Class 4 model). c = class.

Smith, & Makov, 1985). In the latter case, the components of the mixture serve merely to capture the skewness and do not necessarily correspond to clusters of participants within the population. The assumption that the data are multivariate and normally distributed conditional on class and the covariates is violated, and the skewed distribution may, in fact, correspond to a single homogeneous population. An interpretation on a conceptual level in terms of different subpopulations would be incorrect. The possibility of an incorrect interpretation illustrates a more general limitation inherent to exploratory analyses, namely that the selected model can only be regarded as one possible way of explaining the observed data. In the context of deciding on the number of classes, one may, therefore, place some emphasis on the question of whether an additional class is providing useful information about the heterogeneity. Characterization of the classes and tracing of stable classes as done in our illustration can be used in conjunction with statistical measures. Such an approach is not much different from a classic exploratory factor analysis carried out in a single homogeneous population in which the interpretability of additional factors is considered in addition to statistical criteria.

References

- Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control*, *AU-19*, 719–722.
- Amemiya, T. (1985). *Advanced econometrics*. Cambridge, MA: Harvard University Press.
- Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean and covariance structure models. *Psychometrika*, *64*, 475–494.
- Bartholomew, D. J. (1987). *Latent variables models and factor analysis*. New York: Oxford University Press.
- 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), Muthén (2003), and Cudeck and Henly (2003). *Psychological Methods*, *8*, 384–393.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.
- Cramér, H. (1946). *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.
- Cudeck, R., & Henly, S. J. (2003). A realistic perspective on pattern representation in growth data: Comment on Bauer and Curran. *Psychological Methods*, *8*, 378–383.
- Dolan, C. V., Schmittmann, V. D., Lubke, G. H., & Neale, M. C. (2005). Regime switching in the latent growth curve model. *Structural Equation Modeling*, *12*, 94–119.
- Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, *63*, 227–253.
- Hancock, G. R., Kuo, W., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling*, *8*, 470–489.
- Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage.
- Hershberger, S. L. (1994). The specification of equivalent models before the collection of data. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 68–105). Thousand Oaks, CA: Sage.
- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997a). Finite mixture structural equation models for response based segmentation and unobserved heterogeneity. *Marketing Science*, *16*, 39–59.
- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997b). STEMM: A general finite mixture structural equation model. *Journal of Classification*, *14*, 23–50.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426.
- Langenheine, R., & Rost, J. (1988). *Latent trait and latent class models*. New York: Plenum Press.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53–76.
- 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. (2003). Can unequal residual variances across subpopulations mask differences in residual means in the common factor model? *Structural Equation Modeling*, *10*, 175–192.
- Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the context of the common factor model. *Intelligence*, *31*, 543–566.
- Lubke, G. H., & Muthén, B. O. (2005). *Performance of factor mixture models*. Manuscript submitted for publication.
- Luijben, T. C. (1993). Equivalent models in covariance structure analysis. *Psychometrika*, *56*, 653–665.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185–199.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 89–99.
- Magidson, J., & Vermunt, J. K. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing*, *20*, 37–44.
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Meehl, P. E. (1992). Factors and taxa, traits and types, differences

- of degree and differences in kind. *Journal of Personality*, 60, 117–174.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13, 127–143.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525–543.
- Miller, J. D., Kimmel, L., Hoffer, T. B., & Nelson, C. (2000). *Longitudinal study of American youth: User's manual*. Chicago, IL: International Center for the Advancement of Scientific Literacy, Chicago Academy of Sciences.
- Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Erlbaum.
- 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. (2004). Latent variable analysis: Growth mixture modeling and related techniques. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.
- Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24, 882–891.
- Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463–469.
- Muthén, L. K., & Muthén, B. O. (2001). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2002). *Mplus 2.1*. [Computer program]. Los Angeles, CA: Muthén & Muthén.
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*, 4, 139–157.
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London (Ser. A)*, 185, 71–110.
- Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London (Ser. A)*, 186, 343–414.
- Raykov, T., & Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling*, 8, 142–149.
- 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.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229–239.
- S-PLUS. (2001). *S-PLUS 6.0* [Computer program]. Seattle: Insightful.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Titterton, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Chichester, England: Wiley.
- Uebersax, J. S. (1999). Probit latent class analysis with dichotomous or ordered category measures: Conditional independence/dependence models. *Applied Psychological Measurement*, 23, 283–297.
- Vermunt, J. K., & Magidson, J. (2000). *Latent GOLD's user's guide*. Boston: Statistical Innovations.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge, MA: Cambridge University Press.
- Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage.
- 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.
- Zeger, S. L., & Liang, K. H. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 121–130.

Received November 5, 2002

Revision received July 8, 2004

Accepted July 13, 2004 ■