Behaviormetrika
Vol.29, No.1, 2002, 81-117

BEYOND *SEM*: GENERAL LATENT VARIABLE MODELING

Bengt O. Muthén*

This article gives an overview of statistical analysis with latent variables. Using traditional structural equation modeling as a starting point, it shows how the idea of latent variables captures a wide variety of statistical concepts, including random effects, missing data, sources of variation in hierarchical data, finite mixtures, latent classes, and clusters. These latent variable applications go beyond the traditional latent variable useage in psychometrics with its focus on measurement error and hypothetical constructs measured by multiple indicators. The article argues for the value of integrating statistical and psychometric modeling ideas. Different applications are discussed in a unifying framework that brings together in one general model such different analysis types as factor models, growth curve models, multilevel models, latent class models and discrete-time survival models. Several possible combinations and extensions of these models are made clear due to the unifying framework.

1. Introduction

This article gives a brief overview of statistical analysis with latent variables. A key feature is that well-known modeling with continuous latent variables is expanded by adding new developments also including categorical latent variables. Taking traditional structural equation modeling as a starting point, the article shows the generality of latent variables, being able to capture a wide variety of statistical concepts, including random effects, missing data, sources of variation in hierarchical data, finite mixtures, latent classes, and clusters. These latent variable applications go beyond the traditional latent variable useage in psychometrics with its focus on measurement error and hypothetical constructs measured by multiple indicators.

The article does not discuss estimation and testing but focuses on modeling ideas and connections between different modeling traditions. A few key applications will be discussed briefly. Although not going into details, the presentation is statistically-oriented. For less technical overviews and further applications of new developments using categorical latent variables, see, e.g., Muthén (2001a, b) and Muthén and Muthén (2000). All analyses are performed using the Mplus program (Muthén & Muthén, 1998-2001) and Mplus input, output, and data for these examples are available at www.statmodel.com/mplus/examples/penn.html.

One aim of the article is to inspire a better integration of psychometric modeling ideas into mainstream statistics and a better use of statistical analysis ideas in latent variable modeling. Psychometrics and statistics have for too long been

Key Words and Phrases: Maximum-likelihood estimation, factor analysis, structural equation modeling, growth curve modeling, multilevel modeling, finite mixture modeling, random effects, missing data, latent classes.

* University of California, Los Angeles, Graduate School of Education & Information Studies, Moore Hall, Box 951521, Los Angeles CA 90095-1521, U.S.A.
Email: bmuthen@ucla.edu

developed too separately and both fields can benefit from input from the other. Traditionally, psychometric models have been concerned with measurement error and latent variable constructs measured with multiple indicators as in factor analysis. Structural equation modeling (SEM) took factor analysis one step further by relating the constructs to each other and to covariates in a system of linear regressions thereby purging the "structural regressions" of biasing effects of measurement error. The idea of using systems of linear regressions emanated from supply and demand modeling in econometrics and path analysis in biology. In this way, SEM consists of two ideas: latent variables and joint analysis of systems of equations. It is argued here that it is the latent variable idea that is more powerful and more generalizable. Despite its widespread use among applied researchers, SEM has still not been fully accepted in mainstream statistics. Part of this is perhaps due to poor applications claiming the establishment of causal models and part is perhaps also due to strong reliance on latent variables that are only indirectly defined. The skepticism about latent variables is unfortunate given that, as shown in this article, latent variables are widely used in statistics, although under different names and different forms.

This article argues that by emphasizing the vehicle of latent variables, psychometric modeling such as SEM can be brought into mainstream statistics. To accomplish this, it is necessary to clearly show how many statistical analyses implicitly utilize the idea of latent variables in the form of random effects, components of variation, missing data, mixture components, and clusters. To this aim, a general model is discussed which integrates psychometric latent variable models with latent variable models presented in the statistical literature. The generality of the model is achieved by considering not only continuous latent variables but also categorical latent variables. This makes it possible to unify and to extend a wide variety of common types of analyses, including SEM, growth curve modeling, multilevel modeling, missing data modeling, finite mixture modeling, latent class modeling, and survival modeling. The general model is shown schematically in Figure 1. The general framework (D) is represented by the square, while special cases (A, B, C) to be discussed in the article are shown in ellipses. The general framework is drawn from Muthén and Muthén (1998-2001; Appendix 8) as implemented in the Mplus computer program (www.statmodel.com). It should be noted that Figure 1 is a simplification. For example, the general framework includes direct effects from c to u, from c to y, and allows c to also influence regression and variance parameters in the u and y parts of the model. It is hoped that the use of a single modeling and software framework makes latent variable modeling more accessible to both statisticians and substantive researchers. Statisticians can more easily see connections between latent variable uses that they are accustomed to and psychometric uses. Substantive researchers can more easily focus on the research problem at hand rather than learning a multitude of model specification systems and software languages.

The article is structured as follows. Section 2 discusses framework A of the

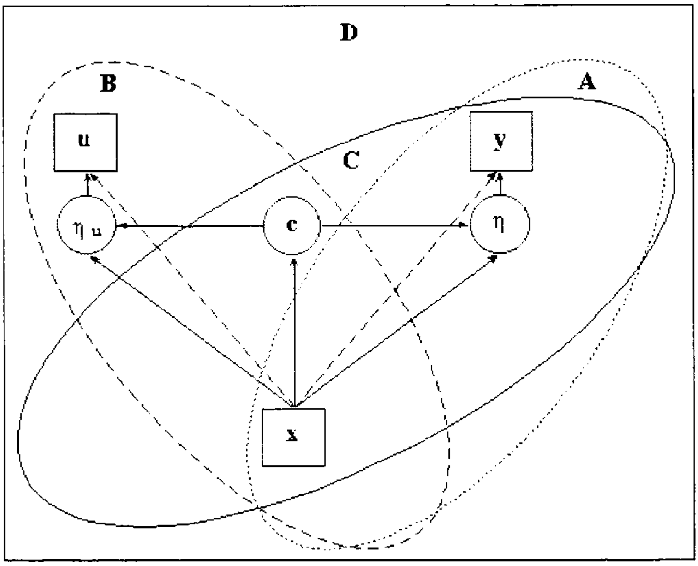


Figure 1: A general latent variable modeling framework

general model. This framework corresponds to the more well-known case of continuous latent variables. Sub-sections discuss the modeling of measurement error and measurement invariance in conventional SEM, random effects in growth modeling, and variance components in multilevel modeling. Section 3 discusses framework B introducing categorical latent variables, including latent class analysis and latent class growth analysis. A latent class analysis example is presented where individuals are classified based on their antisocial behavior. Section 4 discusses framework C, including latent profile models and models that combine continuous and categorical latent variables such as growth mixture models. A growth mixture example is presented where children are classified into a problematic class based on their reading development in Kindergarten and second grade. Section 5 discusses the general framework D, presenting new types of models, including modeling with missing data on a categorical latent variable in randomized trials. Section 6 concludes.

2. Modeling Framework A: Continuous Latent Variables

Consider the special case A of the general modeling framework shown in Figure 1. Framework A is characterized by using continuous latent variables, denoted by the vector η, shown as a circle in ellipse A in Figure 1. Here, latent variables are used to represent constructs that have fundamental substantive importance but are only measured indirectly through multiple indicators that capture different aspects of the constructs.

As a first step, a general SEM formulation of framework A is presented, followed by the key analysis areas of random effects modeling and variance component modeling.[1)]

The measurement part of the model is defined in terms of the p-dimensional continuous outcome vector $\mathbf{y}$,

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\, \boldsymbol{\eta}_i + \mathbf{K}\, \mathbf{x}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

where $\boldsymbol{\eta}$ is an m-dimensional vector of latent variables, $\mathbf{x}$ is a q-dimensional vector of covariates, $\boldsymbol{\epsilon}$ is a p-dimensional vector of residuals or measurement errors which are uncorrelated with other variables, $\boldsymbol{\nu}$ is a p-dimensional parameter vector of measurement intercepts, $\boldsymbol{\Lambda}$ is a $p \times m$ parameter matrix of measurement slopes or factor loadings, and $\mathbf{K}$ is a $p \times q$ parameter matrix of regression slopes. Usually, only a few of the elements of $\mathbf{K}$ have nonzero elements, where a non-zero row corresponds to a y variable that is directly influenced by one or more x variables. The covariance matrix of $\boldsymbol{\epsilon}$ is denoted $\boldsymbol{\Theta}$. The structural part of the model is defined in terms of the latent variables regressed on each other and the q-dimensional vector $\mathbf{x}$ of independent variables,

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\, \boldsymbol{\eta}_i + \boldsymbol{\Gamma}\, \mathbf{x}_i + \boldsymbol{\zeta}_i. \tag{2}$$

Here, $\boldsymbol{\alpha}$ is an m-dimensional parameter vector, $\mathbf{B}$ is an $m \times m$ parameter matrix of slopes for regressions of latent variables on other latent variables. $\mathbf{B}$ has zero diagonal elements and it is assumed that $\mathbf{I} - \mathbf{B}$ is non-singular. Furthermore, $\boldsymbol{\Gamma}$ is an $m \times q$ slope parameter matrix for regressions of the latent variables on the independent variables, and $\boldsymbol{\zeta}$ is an m-dimensional vector of residuals. The covariance matrix of $\boldsymbol{\zeta}$ is denoted $\boldsymbol{\Psi}$. In line with regression analysis, the marginal distribution of $\mathbf{x}$ is not modelled but is left unrestricted. This leads to the mean and covariance structures conditional on $\mathbf{x}$,

$$\boldsymbol{\nu} + \boldsymbol{\Lambda}\, (\mathbf{I} - \mathbf{B})^{-1}\, \boldsymbol{\alpha} + \boldsymbol{\Lambda}\, (\mathbf{I} - \mathbf{B})^{-1}\, \boldsymbol{\Gamma}\, \mathbf{x} + \mathbf{K}\, \mathbf{x}, \tag{3}$$

$$\boldsymbol{\Lambda}\, (\mathbf{I} - \mathbf{B})^{-1}\, \boldsymbol{\Psi}\, (\mathbf{I} - \mathbf{B})'^{-1} \boldsymbol{\Lambda}' + \boldsymbol{\Theta}. \tag{4}$$

With the customary normality assumption of $\mathbf{y}$ given $\mathbf{x}$, the parameters of the model are estimated by fitting (3) and (4) to the corresponding sample quantities. This is the same as fitting the mean vector and covariance matrix for the vector $(\mathbf{y}, \mathbf{x})'$ to the sample means, variances, and covariances for $(\mathbf{y}, \mathbf{x})'$ (Jöreskog & Goldberger, 1975). Here, the maximum-likelihood estimates of $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$ are the corresponding sample quantities.

[1)] Mplus examples of framework A models are given at
www.statmodel.com/mplus/examples/continuous.html.

Joint analysis of independent samples from multiple groups is also possible, assuming different degrees of parameter invariance across groups. In particular, full or partial invariance of the measurement parameters of $\boldsymbol{\nu}$ and $\boldsymbol{\Lambda}$ is of interest in order to study group differences with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\Psi}$.

From an application point of view, the modeling in (1), (2) is useful for purging regression relationships of detrimental effects of measurement error when multiple indicators of a construct are available. Measurement errors among the predictors are well-known to have particularly serious effects, but the modeling is also useful in examining a factor model where the measurement errors are among the outcome (indicator) variables as when using a factor analysis with covariates ("MIMIC" modeling). In this special case, $\mathbf{B} = \mathbf{0}$ in (2). A baseline MIMIC analysis assumes $\mathbf{K} = \mathbf{0}$ in (1) and a sufficient number of restrictions on $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ to make the model identified (in an exploratory analysis, this amounts to using m^2 restrictions in line with exploratory factor analysis). The covariates strengthen the factor analysis in two ways (cf. Muthén, 1989). First, by making the test of dimensionality stronger by using associations not only among the $\mathbf{y}$ variables but also between $\mathbf{y}$ and $\mathbf{x}$. Second, by making it possible to examine the extent of measurement invariance across groups defined by different values on $\mathbf{x}$. Measurement non-invariance across groups defined by x_k (e.g. $x_{ki} = 0/1$ for individual i) with respect to an outcome y_j is captured by $\kappa_{jk} \neq 0$, reflecting a group-varying intercept, $\nu_j + \kappa_{jk}\, x_k$.

The model of (1) - (4) is typically estimated by maximum-likelihood (ML) under the assumption of multivariate normality. Browne and Arminger (1995) give an excellent summary of modeling and estimation issues for this model. This is the analysis framework used for the last 20 years by conventional SEM computer programs such as AMOS, EQS, and LISREL. More recently, ML estimation assuming missing at random (MAR) in the sense of Little and Rubin (1987) has been introduced in SEM software.

Browne and Arminger (1995) also discuss the case where some or all of the $\mathbf{y}$ outcomes are categorical. The case of categorical outcomes has been further treated in Muthén (1984, 1989) with an emphasis on weighted least-squares estimation, including a new approach presented in Muthén, DuToit, Spisic (1997). Mplus includes modeling with both continuous and categorical outcomes $\mathbf{y}$.[2] Drawing on Muthén (1996) and Muthén and Christofferson (1981), Mplus provides a more flexible parameterization than conventional SEM software in terms of its categorical outcome modeling for longitudinal data and multiple-group analysis using threshold measurement parameters that allow for partial invariance across time and group. For connections with item response theory, see, e.g., Muthén (1988), Muthén, Kao and Burstein (1991), and Takane and DeLeeuw (1987).

For an overview of conventional SEM with continuous outcomes, see, e.g. Bollen (1989). For examples of SEM analysis in behavioral research, see, e.g.,

[2] Mplus examples are given at www.statmodel.com/mplus/examples/categorical.html.

MacCallum and Austin (2000).

2.1 Random effects growth modeling

The use of random effects is another example of modeling with continuous latent variables. In mainstream statistics, random effects are used to capture unobserved heterogeneity among subjects. That is, individuals differ in systematic ways that cannot be, or at least have not been, measured. Unlike the case of psychometric latent variable contexts, however, the random effects are typically not thought of as constructs of primary interest, and there is typically not an attempt at directly measuring the random effects.

2.1.1 A growth modeling example

Growth modeling is an interesting example of random effect modeling where the heterogeneity concerns individual differences in trajectories. Consider an example from reading research. The data are from a cohort-sequential reading study of 945 children in a sample of Texas schools, following them from Kindergarten through second grade. In Kindergarten a phonemic awareness score was measured as a reading precursor skill. In grades 1 and 2, word recognition scores were collected. All measures were collected at four times during the school year. At the end of grade 2, standardized reading and spelling scores were also recorded. These data will also be used to illustrate growth mixture modeling in framework C. Figure 2 shows observed individual trajectories on a phonemic awareness score for Kindergarten children divided into the upper and lower decile on the grade 2 spelling score. The individual variation in the trajectories is clearly seen with those in the lower spelling decile showing a lower initial and ending status in Kindergarten and a lower growth rate than those in the upper spelling decile.

2.1.2 Modeling issues

A modeling example shows the latent variable connections. Let the random variables η_0, η_1, and η_2 represent an intercept, a linear, and a quadratic slope, respectively. These are coefficients in the regression of the outcome on time and the fact that they vary across individuals gives rise to the term random coefficients or random effects. The random effects capture the individual differences in development over time using the Laird and Ware (1982) type of model

$$y_{it} = \eta_{0i} + \eta_{1i}\,(a_t - a) + \eta_{2i}\,(a_t - a)^2 + \kappa_t\,x_{it} + \epsilon_{it}, \tag{5}$$

$$\eta_{0i} = \alpha_0 + \gamma_0\,x_{i0} + \zeta_{0i}, \tag{6}$$

$$\eta_{1i} = \alpha_1 + \gamma_1\,x_{i0} + \zeta_{1i}, \tag{7}$$

$$\eta_{2i} = \alpha_2 + \gamma_2\,x_{i0} + \zeta_{2i}, \tag{8}$$

where a_t is a time-related variable, a is centering constant, x_t is a time-varying covariate, and x_0 is a time-invariant covariate. In multilevel terms (see, e.g., Bryk

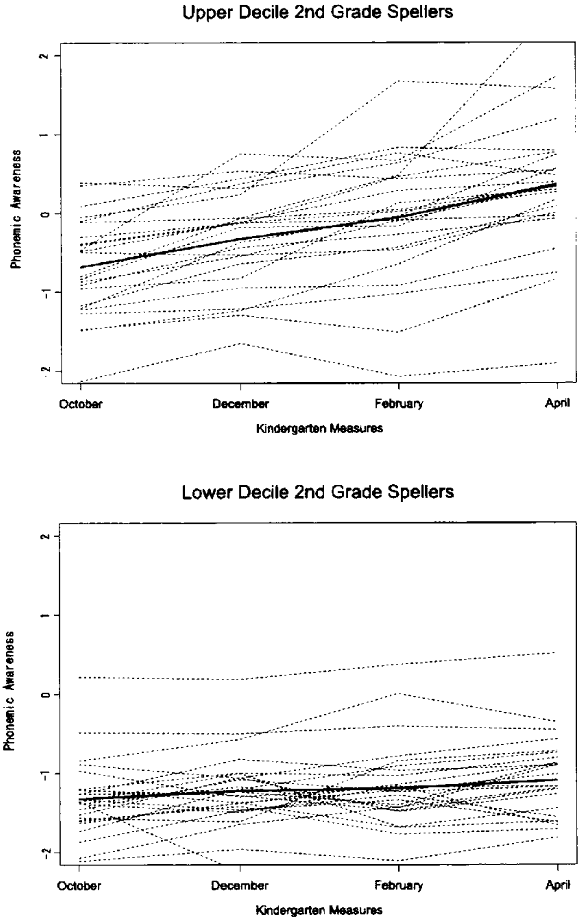


Figure 2: Phonemic awareness development in Kindergarten

& Raudenbush, 2002), (5) is referred to as the level 1 equation, while (6) - (8) are referred to as level 2 equations. In mixed linear modeling (see, e.g. Jennrich & Sluchter, 1986; Lindstrom & Bates, 1988; Goldstein, 1995), the model is expressed in terms of y_t related to a_t, x_t, and x_0, inserting (6) - (8) into (5).

It is clear that (5) - (8) can be expressed in SEM terms using (1) and (2) by letting $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $\boldsymbol{\eta}_i = (\eta_{0i}, \eta_{1i}, \eta_{2i})'$, and $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iT}, x_{i0})'$. While multilevel modeling views the analysis as a two-level analysis of a univariate outcome y, the SEM approach is a single-level analysis of the multivariate vector $\mathbf{y}$. This issue will be discussed further in Section 2.2. Furthermore, $\boldsymbol{\nu} = \mathbf{0}$ while $\boldsymbol{\alpha}$ contains three free parameters. Alternatively, the equivalent parameterization

$\nu_1 = \nu_2 = \ldots = \nu_T$ with $\boldsymbol{\alpha} = (0, \alpha_1, \alpha_2)'$ may be used. Also,

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1 & a_1 - a & (a_1 - a)^2 \\ 1 & a_2 - a & (a_2 - a)^2 \\ \vdots & \vdots & \\ 1 & a_T - a & (a_t - a)^2 \end{pmatrix}, \tag{9}$$

showing that growth modeling with random effects is a form of factor analysis with covariates. Time-invariant covariates have effects on the factors and time-varying covariates have direct effects on the outcomes. While not typically thought of as such, the repeated measures of $y_1, y_2, \ldots, y_T$ can be seen as multiple indicators of the random effects, or growth factors as they are referred to in the latent variable literature. One may wonder why a latent variable model with such a restricted $\boldsymbol{\Lambda}$ matrix as in (9), with no free parameters, would ever be realistic. But it has been found that this model often captures the essential features of growth. In the latent variable framework, however, it is easy to allow deviations from the functional growth form by estimating some of the loadings.

Multilevel and mixed linear modeling traditions consider a more general form of (5),

$$y_{it} = \eta_{0i} + \eta_{1i}\,(a_{it} - a) + \eta_{2i}\,(a_{it} - a)^2 + \kappa_i\, x_{it} + \epsilon_{it}, \tag{10}$$

where a_{it} indicates the possibility of individually-varying times of observation and the slope κ_i is yet another random effect. These traditions treat $(a_{it} - a)$ as data, whereas conventional SEM software treats $(a_t - a)$ as parameters. This is the only way that random slopes can be handled in conventional SEM. In (10), η_{1i}, η_{2i}, and κ_i are random slopes for individually-varying variables a and x. As pointed out in Raudenbush (2001) such modeling cannot be summarized in terms of mean and covariance structures. Unlike (4), the variance of y conditional on the a and x variables varies as a function of these variables. In principle, however, this points to a shortcoming of conventional SEM software, not a shortcoming of latent variable modeling. Drawing on Asparouhov and Muthén (2002), Mplus incorporates individually-varying times of observations and random slopes for time-varying covariates as in (10).[3)]

Mainstream statistics also takes an interest in what psychometricians call factor scores, i.e. estimates of η_i values to be used for estimation of individual growth curves. Both fields favor empirical Bayes estimates, referred to as the regression method in psychometrics.

It follows that there are several advantages of placing the growth model in a latent variable context. For example, the psychometric idea of a latent variable construct is not utilized in the growth model of (5) - (8). Although the y outcomes are manifestations of growth, a psychometric approach could in principle

[3)] These features are included in Version 2.1 to be released Spring 2002 as a free upgrade for Version 2 users.

seek specific indicators of the growth factors, for instance measuring indicators of growth potential at the outset of the study, an approach that does not seem to have been pursued. A more common situation is that a researcher wants to study growth in a latent variable construct measured with multiple indicators. The model specification is as follows, for simplicity shown for a linear model with a single latent variable construct η_{it}.

Let y_{ijt} denote the outcome for individual i, indicator j, and timepoint t, and let η_{it} denote a latent variable construct,

$$\text{Level-1a (measurement part):}$$

$$y_{ijt} = \nu_{jt} + \lambda_{jt}\, \eta_{it} + \epsilon_{ijt}, \tag{11}$$

$$\text{Level-1b} : \eta_{it} = \eta_{0i} + \eta_{1i}\;\; a_t + \zeta_{it}, \tag{12}$$

$$\text{Level-2a} : \eta_{0i} = \alpha_0 + \gamma_0\, x_i + \zeta_{0i}, \tag{13}$$

$$\text{Level-2b} : \eta_{1i} = \alpha_1 + \gamma_1\, x_i + \zeta_{1i}. \tag{14}$$

In line with the second parameterization given above for a single outcome, measurement invariance is specified by using time-invariant indicator intercepts and slopes:

$$\nu_{j1} = \nu_{j2} = \ldots = \nu_{jT} = \nu_j, \tag{15}$$

$$\lambda_{j1} = \lambda_{j2} = \ldots = \lambda_{jT} = \lambda_j, \tag{16}$$

setting the metric of the latent variable construct by $\lambda_1 = 1$. The intercept of the level-2a equation is fixed at zero, $\alpha_0 = 0$. $V(\epsilon_{ijt})$ and $V(\zeta_{it})$ may vary over time. Structural differences are captured by letting $E(\eta_{it})$ and $V(\eta_{it})$ vary over time. With more than one population, across-population measurement invariance would be imposed and α_0 fixed to zero only in the first population. Multiple-indicator growth modeling has the advantage that changes in measurements can be made over time, assuming measurement invariance for a subset of indicators that are maintained between adjacent time points.

Other advantages of growth modeling in a latent variable framework includes the ease with which to carry out analysis of multiple processes, both parallel in time and sequential, as well as multiple groups. Growth factors may be regressed on each other using the **B** matrix in (2), for example studying growth while controlling for not only observed covariates but also initial status. More generally, the growth model may be only a part of a larger model, including for instance a factor analysis measurement part for covariates measured with errors, or including a mediational path analysis part for variable influencing the growth factors, or including a set of variable that are influenced by the growth process.

For examples of growth modeling in a latent variable framework, see, e.g., Muthén and Khoo (1998) and Muthén and Curran (1997). The recent Collins and Sayer (2001) book gives applied contributions from several different traditions.

2.2 Components of variation in hierarchical data

Latent continuous variables are frequently used in statistical modeling of hierarchical data. Here, latent variables are used to correctly reflect the sampling procedure with latent variables representing sources of variation at different levels of the hierarchy.

It is instructive to consider a simple ANOVA model because it clearly shows relationships between factor analysis, growth modeling, and more general multilevel latent variable models.

Consider the nested, random-effects ANOVA,

$$y_{ij} = \nu + \eta_i + \epsilon_{ij} \; ; i = 1, 2, \ldots, n \; ; j = 1, 2, \ldots, J. \tag{17}$$

Here, i is the mode of variation for which an independent sample is obtained, while j is clustered within i. Typical examples are individuals observed within households and students observed within classrooms. The different sources of variation are captured by the latent variables η and ϵ. If instead $j = 1, 2, \ldots, n_i$, there is missing data on some of the J measures.

Consider the covariance and variances for $j = k$ and $j = l$,

$$cov(y_{ik}, y_{il}) = v(\eta), \tag{18}$$

$$v(y_{ik}) = v(y_{il}) = v(\eta) + v(\epsilon), \tag{19}$$

resulting in the intraclass correlation

$$\rho(y_{ik}, y_{il}) = v(\eta)/[v(\eta) + v(\epsilon)]. \tag{20}$$

The intraclass correlation is frequently considered in the context of cluster sampling. The intraclass correlation increases for increasing between-cluster variation $v(\eta)$ relative to total variation. Or, using equivalent homogeneity reasoning, the intraclass correlation increases when the within-cluster variation $v(\epsilon)$ is small. In cluster samples, the intraclass correlation is used to describe the lack of independence among observations and used when computing design effects. The simple model of (17) summarizes some key latent variable modeling issues in a nutshell, showing that factor analysis, growth modeling, and multilevel modeling are variations on the same theme.

It is clear that (17) can be seen as a special case of factor analysis in the SEM framework of (1), (2) with a single factor and $\boldsymbol{\lambda} = (1, 1, \ldots, 1)'$. Instead of thinking of the J units within each cluster as individuals as in (17), the J y variables are now multiple indicators measured on the same individual. Carrying this idea back to (17), this means that the individuals within a cluster can be seen as indicators measuring cluster characteristics.

When y_j are repeated measures over time, $j = t$, (17) represents a growth model with random intercepts. The repeated measures take the role of multiple indicators measuring the random intercept growth factor. For example, the model may represent blood pressure measurements on individuals where in a short time span there is no increasing or decreasing trend. The focus is on the construct of "long-term blood pressure level", e.g. for predicting later health outcomes. The ϵ residuals represent measurement error as well as time-specific variation, both of which may be irrelevant for the prediction.

It was noted earlier that growth modeling in the SEM framework leads to single-level analysis because a multivariate analysis of $\mathbf{y}$ is carried out. The non-independence among repeated measures within an individual indicated by the intraclass correlations is modeled by the growth factors influencing the outcome at different time points. This is analogous to factor analysis. The same multivariate analysis approach may be used for more general multilevel modeling with latent variables, for example multilevel factor analysis and multilevel growth modeling, referred to as 3-level modeling in the multilevel literature. The multivariate approach is suitable for situations where there are relatively few cluster members, such as with analysis of spouses, siblings, or analysis of twins in behavioral genetics. For a recent application to growth modeling of alcohol use among siblings, see Khoo and Muthén (2000). The multivariate approach provides very flexible modeling where the relationships among units within a cluster can be modeled. In Khoo and Muthén (2000) the growth factors of a younger sibling are regressed on those of an older sibling. The cluster units can also have different regressions on covariates. Different numbers of cluster units for different clusters can be handled via missing data, although different models may be relevant for clusters of different size (i.e. two-sibling homes may have a different dynamics than homes with many siblings). With more than a couple of cluster units, however, the multivariate approach becomes computationally cumbersome. For instance with 10 measures per student with 15 students per classrooms, a multivariate vector of length 150 would have to be analyzed. As an alternative, multilevel modeling makes a simplifying assumption of cluster units being statistically equivalent as shown below.

Assume $c = 1, 2, ..., C$ independently observed clusters with $i = 1, 2, ..., n_c$ individual observations within cluster c. Let $\mathbf{z}$ and $\mathbf{y}$ represent group- and individual-level variables. Arrange the data vector for which independent observations are obtained as

$$\mathbf{d}_c{}' = (\mathbf{z}_c{}', \mathbf{y}_{c1}{}', \mathbf{y}_{c2}{}', ..., \mathbf{y}_{cn_c}{}'),$$

where we note that the length of $\mathbf{d}_c$ varies across clusters. The mean vector and covariance matrix are

$$\boldsymbol{\mu}_{d_c}{}' = [\boldsymbol{\mu}_z{}', \mathbf{1}_{n_c}{}' \otimes \boldsymbol{\mu}_y{}'] \qquad (21)$$

$$\boldsymbol{\Sigma}_{d_c} = \begin{bmatrix} \boldsymbol{\Sigma}_{zz} & \textit{symmetric} \\ \mathbf{1}_{n_c} \otimes \boldsymbol{\Sigma}_{yz} & \mathbf{I}_{n_c} \otimes \boldsymbol{\Sigma}_W + \mathbf{1}_{n_c}\mathbf{1}'_{n_c} \otimes \boldsymbol{\Sigma}_B \end{bmatrix}. \qquad (22)$$

The covariance matrix $\boldsymbol{\Sigma}_{d_c}$ shows that the usual *i.i.d* assumption of simple random sampling is modified to allow for non-independent observations within clusters and that this non-independence is modeled by the $\boldsymbol{\Sigma}_B$ matrix in line with the nested

ANOVA model of (17). In (21) and (22) the sizes of the arrays are determined by the product $n_c \times p$ where p is the number of observed variables. McDonald and Goldstein (1989) pointed out that a great reduction in size is obtainable, reducing the ML expression

$$\sum_{c=1}^{C} \{\ln \mid \mathbf{\Sigma}_{d_c} \mid + (\mathbf{d}_c - \boldsymbol{\mu}_{d_c})' \, \mathbf{\Sigma}_{d_c}{}^{-1} \, (\mathbf{d}_c - \boldsymbol{\mu}_{d_c})\}$$

to

$$\sum_{d}^{D} C_d \, \{\ln \mid \mathbf{\Sigma}_{dd} \mid + \mathrm{tr}[\, \mathbf{\Sigma}_{dd}{}^{-1} \, (\mathbf{S}_{Bd} + n_d \, (\bar{\boldsymbol{v}}_d - \boldsymbol{\mu})(\bar{\boldsymbol{v}}_d - \boldsymbol{\mu})')]\}$$

$$+ (n - C) \, \{ \ln \mid \mathbf{\Sigma}_W \mid + \mathrm{tr}[\, \mathbf{\Sigma}_W^{-1} \, \mathbf{S}_{PW}]\}.$$

where d sums over clusters with distinct cluster sizes (for details, see Muthén, 1990).

Muthén (1989, 1990, 1994) showed how SEM software can be used for analyzing models of this type. This is referred to as 2-level modeling in a latent variable framework. Here, $\boldsymbol{\mu}$, $\mathbf{\Sigma}_W$ and $\mathbf{\Sigma}_B$ are structured in terms of SEM parameter arrays based on (1) and (2). The analysis can be carried out using the Mplus program.

In multilevel terms, this type of model may be viewed as a random intercept model in line with (17) because of the additivity $\mathbf{\Sigma} = \mathbf{\Sigma}_W + \mathbf{\Sigma}_B$. As mentioned earlier, the inclusion of random slopes leads to models that cannot be summarized in terms of mean and covariance structures as done above (Raudenbush, 2001). Nevertheless, random slopes can be incorporated into 2-level latent variable modeling (see Asparouhov & Muthén, 2002).

3-level modeling is also included in the framework of (21) and (22) when one of the levels can be handled by a multivariate representation, as in the case of growth modeling in line with Section 2.1. Latent variable growth modeling in cluster samples is discussed in Muthén (1997).

For examples, see, e.g., Muthén (1991) with an application of multilevel factor analysis and Muthén (1989) with an application to SEM. Further examples are given in Hecht (2001) and Kaplan and Elliott (1997).

3. Modeling Framework B

Consider next the special case B of the general modeling framework shown in Figure 1. Framework B is characterized by using categorical latent variables, denoted by the circle c in Figure 1 (the circle denoted η_u will be discussed later on). The choice of using a categorical latent variable instead of a continuous latent variable is more fundamental than the corresponding choice of proper scale type for observed outcomes. The addition of categorical latent variables to the general framework in Figure 1 opens up a whole new set of modeling capabilities.

In mainstream statistics, this type of modeling is referred to as finite mixture modeling. In the current article, the terms latent class and mixture modeling will be used interchangeably. As with continuous latent variables, categorical latent variables are used for a variety of reasons as will now be shown.

As a first step, a general modeling representation of framework B as used in Mplus (Muthén & Muthén, 1998-2001) is presented. This is followed by a discussion of four special cases: latent class analysis, latent class analysis with covariates, latent class growth analysis, latent transition analysis, and logistic regression mixture analysis. Methodological contributions to these areas have been made in separate fields often without sufficient connections and without sufficient connections to modeling in other frameworks. For example, until recently, modeling developments for continuous latent variables in framework A and categorical latent variables in framework B have been kept almost completely separate.[4)]

Let $\mathbf{c}$ denote a latent categorical variable with K classes, $\mathbf{c}_i = (c_{i1}, c_{i2}, \ldots, c_{iK})'$, where $c_{ik} = 1$ if individual i belongs to class k and zero otherwise. Framework B has two parts: $\mathbf{c}$ related to $\mathbf{x}$ and $\mathbf{u}$ related to $\mathbf{c}$ and $\mathbf{x}$. $\mathbf{c}$ is related to $\mathbf{x}$ by multinomial logistic regression using the $K-1$-dimensional parameter vector of logit intercepts $\boldsymbol{\alpha}_c$ and the $(K-1)\times q$ parameter matrix of logit slopes $\mathbf{\Gamma}_c$, where for $k = 1, 2, \ldots, K$

$$P(c_{ik} = 1|\mathbf{x}_i) = \frac{e^{\alpha_{c_k} + \gamma'_{c_k}\mathbf{x}_i}}{\sum_{j=1}^{K} e^{\alpha_{c_j} + \gamma'_{c_j}\mathbf{x}_i}}, \tag{23}$$

where the last class is a reference class with coefficients standardized to zero, $\alpha_{c_K} = 0$, $\gamma_{c_k} = \mathbf{0}$.

For $\mathbf{u}$, conditional independence is assumed given $\mathbf{c}_i$ and $\mathbf{x}_i$,

$$P(u_{i1}, u_{i2}, \ldots, u_{ir}|\mathbf{c}_i, \mathbf{x}_i) = P(u_{i1}|\mathbf{c}_i, \mathbf{x}_i)\; P(u_{i2}|\mathbf{c}_i, \mathbf{x}_i) \ldots P(u_{ir}|\mathbf{c}_i, \mathbf{x}_i). \tag{24}$$

The categorical variable $u_{ij} (j = 1, 2, \ldots, r)$ with S_j ordered categories follows an ordered polytomous logistic regression (proportional odds model), where for categories $s = 0, 1, 2, \ldots, S_j - 1$ and $\tau_{j,k,0} = -\infty$, $\tau_{j,k,S_j} = \infty$,

$$u_{ij} = s,\ if \quad \tau_{j,k,s} < u^*_{ij} \leq \tau_{j,k,s+1}, \tag{25}$$

$$P(u_{ij} = s|\mathbf{c}_i, \mathbf{x}_i) = F_{s+1}(u^*_{ij}) - F_s(u^*_{ij}), \tag{26}$$

$$F_s(u^*) = \frac{1}{1 + e^{-(\tau_s - u^*)}}. \tag{27}$$

where for $\mathbf{u}^*_i = (u^*_{i1}, u^*_{i2}, \ldots, u^*_{ir})'$, $\boldsymbol{\eta}_{ui} = (\eta_{u_{1i}}, \eta_{u_{2i}}, \ldots, \eta_{u_{fi}})'$, and conditional on class k,

$$\mathbf{u}^*_i = \mathbf{\Lambda}_{u_k}\ \boldsymbol{\eta}_{ui} + \mathbf{K}_{u_k}\ \mathbf{x}_i, \tag{28}$$

$$\boldsymbol{\eta}_{ui} = \boldsymbol{\alpha}_{u_k} + \mathbf{\Gamma}_{u_k}\ \mathbf{x}_i, \tag{29}$$

[4)] Mplus examples of framework B models are given at www.statmodel.com/mplus/examples/mixture.html.

where $\mathbf{\Lambda}_{u_k}$ is an $r \times f$ logit parameter matrix varying across the K classes, $\mathbf{K}_{u_k}$ is an $r \times q$ logit parameter matrix varying across the K classes, $\boldsymbol{\alpha}_{u_k}$ is an $f \times 1$ vector logit parameter vector varying across the K classes, and $\mathbf{\Gamma}_{u_k}$ is an $f \times q$ logit parameter matrix varying across the K classes. The thresholds may be stacked in the $\sum_{j=1}^{r}(S_j - 1) \times 1$ vectors $\boldsymbol{\tau}_k$ varying across the K classes.

It should be noted that (28) does not include intercept terms given the presence of τ parameters. Furthermore, τ parameters have opposite signs than u^* in (28) because of their interpretation as thresholds or cutpoints that a latent continuous response variable u^* exceeds or falls below (see also Agresti, 1990, pp. 322-324). For example, with a binary u scored 0/1 (26) leads to

$$P(u = 1|\mathbf{c}, \mathbf{x}) = 1 - \frac{1}{1 + e^{-(\tau - u^*)}}. \tag{30}$$

$$\tag{31}$$

For example, the higher the τ the higher u^* needs to be to exceed it, and the lower the probability of $u = 1$.

Mixture modeling can involve numerical and statistical problems. Mixture modeling is known to sometimes generate a likelihood function with several local maxima. The occurrence of this depends on the model and the data. It is therefore recommended that for a given dataset and a given model different optimizations are carried out using different sets of starting values.

The numerical and statistical performance of mixture modeling benefits from confirmatory analysis. The same kind of confirmatory analysis as in regular modeling is possible, using a priori restrictions on the parameters. With mixture modeling, however, there is also a second type of confirmatory analysis. A researcher may want to incorporate the hypothesis that certain individuals are known to represent certain latent classes. Individuals with known class membership are referred to as training data (see also McLachlan & Basford, 1988; Hosmer, 1973). Multiple-group modeling corresponds to the case of all sample units contributing training data so that c is in effect an observed categorical variable.

In Mplus, the training data can consists of 0 and 1 class membership values for all individuals, where 1 denotes which classes an individual may belong to. Known class membership for an individual corresponds to having training data value of 1 for the known class and 0 for all other classes. Unknown class membership for an individual is specified by the value 1 for all classes. With class membership training data, the class probabilities are renormed for each individual to add to one over the admissible set of classes. Fractional training data is also allowed, corresponding to class probabilities adding to unity for each individual. With fractional training data, the class probabilities are taken to be fixed quantities, which reduces the sampling variability accounted for in the standard error calculations. Fractional training data where each individual has a probability of 1 for one class and 0's for the other classes is equivalent to training data with class membership value 1 for only one class for each individual. Using training data with a value of 1 for one class and 0's for the other classes makes it possible to perform multinomial logistic regression with an unordered, polytomous observed dependent variable using the Mplus model part where $\mathbf{c}$ is related to $\mathbf{x}$.

3.1 Latent class analysis

In latent class analysis the categorical latent variable is used to represent unobserved heterogeneity. Here, the particular aim is to find clusters (latent classes) of individuals who are similar. It is assumed that a sufficient number of latent classes for the categorical latent variable results in conditional independence among the observed outcomes. This may be viewed as heterogeneity among subjects such that the dependence among the outcomes is obtained in a spurious fashion by mixing the heterogeneous groups. Because the latent class variable is the only cause of dependence among the outcomes, the latent class model is similar in spirit to factor analysis with uncorrelated residuals.

Latent class analysis typically considers categorical indicators $\mathbf{u}$ of the latent class variable c, using only a subset of modeling framework B. The variables of $\mathbf{u}$ are binary, ordered polytomous, or unordered polytomous. Due to the conditional independence specification, the joint probability of all u's is

$$P(u_1, u_2, \ldots, u_r) = \sum_{k=1}^{K} P(c = k)\ P(u_1|c = k)\ P(u_2|c = k) \ldots P(u_r|c = k). \tag{32}$$

The model has two types of parameters. The distribution of the categorical latent variable is represented by $P(c = k)$ expressed in terms of the logit parameters α_{c_k} in (23). The conditional u probabilities are expressed via logit parameters in line with (31) where for a binary u $logit = -\tau_k$ for class k, i.e. the u^* part of (28) is not needed. Similar to factor analysis, the conditional u probabilities provide an interpretation of the latent classes such that some activities represented by the different u's are more or less likely in some classes than others.

The latent class counterpart of factor scores is obtained by posterior probabilities for each individual belonging to all classes as computed by Bayes' formula

$$P(c = k|u_1, u_2, \ldots, u_r) = \frac{P(c = k)\ P(u_1|c = k)\ P(u_2|c = k) \ldots P(u_r|c = k)}{P(u_1, u_2, \ldots, u_r)}. \tag{33}$$

For an overview of latent class analysis, see Bartholomew (1987), Goodman (1974) and Clogg (1995). For examples, see, e.g., Muthén (2001b), Nestadt, Hanfelt, Liang, Lamacz, Wolyniec and Pulver (1994), Rindskopf and Rindskopf (1986), and Uebersax and Grove (1990).

3.1.1 *A latent class analysis example of antisocial behavior*

The National Longitudinal Survey of Youth (NLS) collected data on antisocial behavior among 16 - 23 year olds. The NLSY administers an instrument with 17 binary items. Maximum-likelihood estimation by Mplus was used. Preliminary latent class analysis of the 17 items pointed to 9 items that captured 4 different latent classes of antisocial behavior. Class 4 is a normative class (no high probability of endorsing any item). Class 3 is a drug involvement class (pot and drug items). Class 2 is a personal offense class (fight, threat items). Class 1 is a property offense class (shoplift, stealing less than 50, conning someone, stealing goods, breaking into property). The profile plot of Figure 3 shows the estimated item probabilities for each of the 4 classes. It should be noted that the classes are not ordered in the sense of increasing item probabilities, but involves different kinds of antisocial activities.

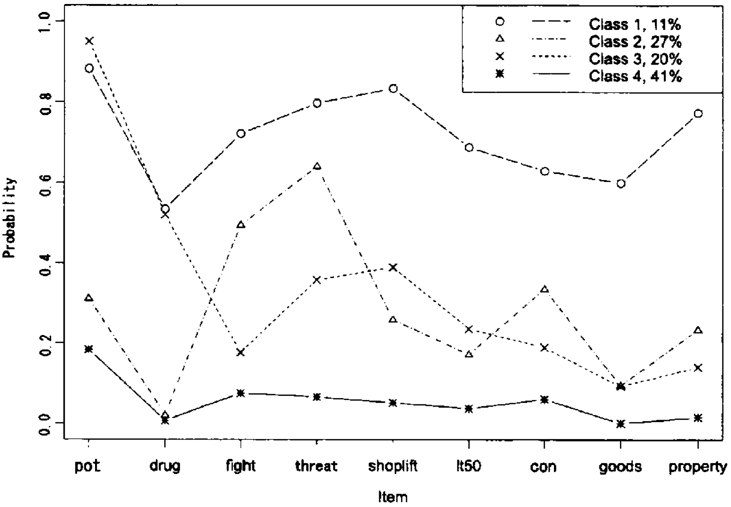


Figure 3: Profiles of antisocial behavior

Table 1 illustrates the use of the estimated posterior probabilities for each individual in each class. The rows correspond to individuals who have the highest probability for that class and the entries are the average probabilities in each class. High diagonal and low off-diagonal values are desirable for good classification. It is seen that class 2 and class 3 are the hardest to distinguish between with a relatively high average class 2 probability of 0.13 for those who have their highest probability in class 3. Class 2 is the person offense class (fight, threat) and class 3 is the drug class (pot, drug). While Figure 3 shows that the two classes have rather different item probabilities on these 4 items, they are similar on the remaining 5 items. This suggests that more items are needed to more clearly distinguish these two classes.

Table 1: Classification table for antisocial behavior latent class analysis

Most Likely Class	**Mean Posterior Probabilities**			
	Class 1	Class 2	Class 3	Class 4
Class 1	0.854	0.074	0.072	0.000
Class 2	0.042	0.810	0.082	0.066
Class 3	0.052	0.134	0.754	0.061
Class 4	0.000	0.122	0.051	0.827

3.2 Latent class analysis with covariates

Similar to factor analysis with covariates, it is useful to include covariates in the latent class analysis. The aim of the latent variable modeling is still to find homogeneous groups of individuals (latent classes), but now covariates $\mathbf{x}$ are included in order to both describe the formation of the latent classes and how they may be differently measured by the indicators $\mathbf{u}$.

The prediction of latent class membership is obtained by the multinomial regression of c on $\mathbf{x}$ in (23). This gives information on the composition of the latent classes. It avoids biases in the common ad hoc 3-step procedure: (1) latent class analysis; (2) classification of individuals based on posterior probabilities; and (3) logistic regression analysis relating classes to covariates.

The variables of $\mathbf{x}$ may also have a direct influence on the variables of $\mathbf{u}$, beyond the influence mediated by c. This is accommodated by estimating elements of $\mathbf{K}_{u_k}$ in (28). For example, with a binary u, the model forms the logistic regression of u on $\mathbf{x}$ for class k,

$$logit = -\tau_k + \kappa_k' \, \mathbf{x}, \qquad (34)$$

so that the direct influence of $\mathbf{x}$ is allowed to vary across classes.

It may be noted that all features of multiple-group analysis are included in the latent class analysis with covariates, with dummy variable covariates representing the groups. Here, τ parameters are the measurement parameters. (34) shows that conditional on class these can vary across the groups, representing for example gender non-invariance. The multiple-group examples of Clogg and Goodman (1985) can all be analyzed in this way.

For examples of latent class analysis with covariates, see, e.g., Bandeen-Roche, Miglioretti, Zeger and Rathouz (1997), Formann (1992), Heijden, Dressens and Bockenholt (1996), Muthén and Muthén (2000), and Muthén (2001b).

3.2.1 A latent class analysis example continued

Continuing the antisocial behavior latent class analysis example above, three covariates from the NLS are added: age, gender, and ethnicity. This example is drawn from Muthén and Muthén (2000). These covariates are specified to influence the probability of class membership using the multinomial regression part (23). Measurement noninvariance with respect to the three covariates can be studied by including a direct effect from a covariate to an item but was not studied here. Maximum-likelihood estimation by Mplus was used. The estimates from the multinomial regression predicting class membership can be translated into the curves of Figure 4. The estimated item profiles remain approximately the same as in Figure 3 and the class interpretation is therefore the same. For a given age, gender, and ethnicity, Figure 4 shows the probability of membership in each class (note that this is not a longitudinal study but the x axis correspond to ages represented in this cross-sectional sample). For example, it is seen that the normative class 4 is the most likely class for all ages among white women, whereas this is not true for the other three groups.

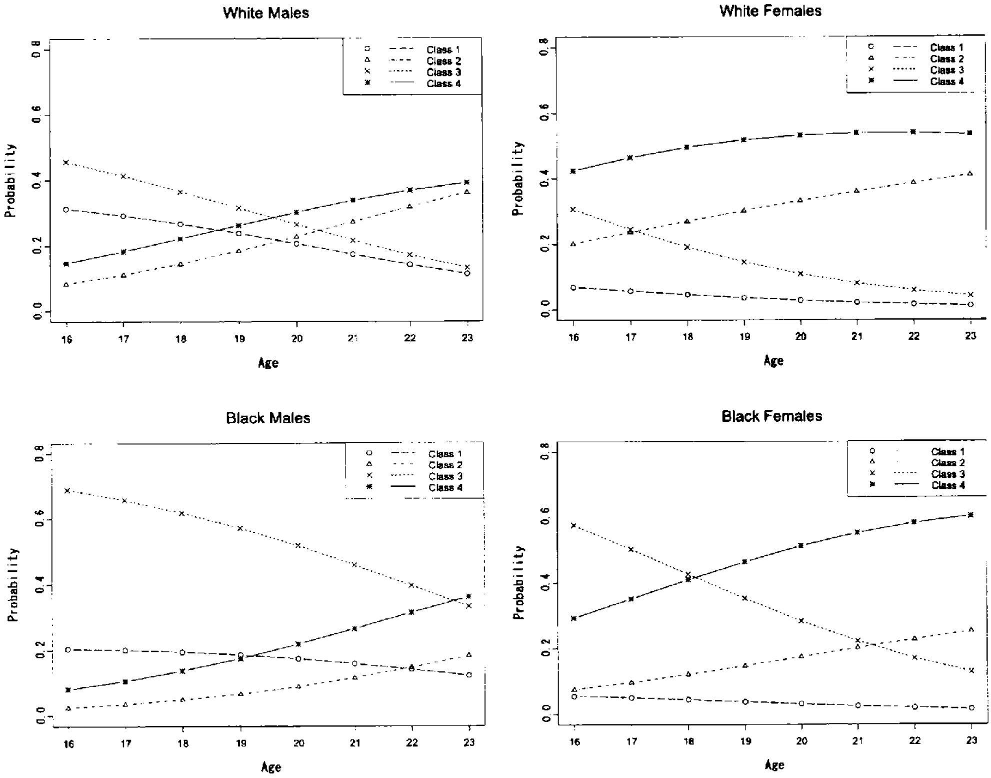


Figure 4: Influence of covariates on antisocial behavior classes

Table 2 shows the resulting classification table based on estimated posterior probabilities. It is seen that the use of covariate information improves the class 2, class 3 distinction relative to Table 1.

Table 2: Classification table for antisocial behavior latent class analysis with covariates

Most Likely Class	**Mean Posterior Probabilities**			
	Class 1	Class 2	Class 3	Class 4
Class 1	0.859	0.065	0.076	0.000
Class 2	0.047	0.808	0.087	0.058
Class 3	0.033	0.067	0.816	0.084
Class 4	0.000	0.048	0.105	0.846

3.3 Latent class growth analysis

Latent class growth analysis again uses a categorical latent variable to represent unobserved heterogeneity, but this time in a form that connects the growth modeling discussed in Section 2.1 and the latent class modeling just discussed. Here, latent classes are sought that are homogeneous with respect to development over time. The latent class growth analysis introduces the continuous latent variable $\boldsymbol{\eta}_u$ of Figure 1.

In latent class growth analysis the multiple indicators of the latent classes correspond to repeated measures over time. Individuals belong to different latent classes characterized by different types of trajectories. Assume for simplicity a single outcome at each timepoint, $\mathbf{u}_i^* = (u_{i1}, u_{i2}, \ldots, u_{it}, \ldots, u_{iT})'$, and the simple growth model corresponding to (28),

$$Level - 1 : u_{it}^* = \quad \eta_{0i} + \eta_{1i} \ \ a_t, \tag{35}$$

where a_t are fixed time scores represented in $\mathbf{\Lambda}_u$,

$$\mathbf{\Lambda}_{u_k} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{pmatrix},$$

where T is the number of time periods. Here, $\boldsymbol{\eta}_u$ in (28), (29) contains an intercept and a slope growth factor with differences across classes captured in $\boldsymbol{\alpha}_{u_k}$ and $\mathbf{\Gamma}_{u_k}\ \mathbf{x}_i$ of (29). The effects of time-varying covariates can be captured in $\mathbf{K}_{u_k}$ of (28).

It may be noted that the modeling does not incorporate continuous latent variables in the form of random effects, but that $\boldsymbol{\eta}_u$ is non-stochastic conditional on **x**. This implies that conditional on **x** there is zero within-class variation across individuals. This limitation can be relaxed in line with the growth modeling of Section 2.1.

With an ordered categorical outcome variable u_{it}, let $\tau_{t,k,s}$ be the s^{th} threshold in class k at timepoint t, $s = 0, 1, 2, \ldots, S_t - 1$, where $\tau_{t,k,0} = -\infty$, $\tau_{t,k,S_t} = \infty$. Across-time and across-class measurement invariance is imposed by the threshold specification

$$\tau_{1,1,s} = \tau_{2,1,s} = \ldots = \tau_{T,1,s} = \ldots = \tau_{1,K,s} = \ldots = \tau_{T,K,s}, \tag{36}$$

for each s value. In the level-2 equation corresponding to (29), the α mean of the intercept growth factor η_{0i} is fixed at zero in the first class for identification purposes. The mean of the intercept growth factor is free to be estimated in the remaining classes.

Latent class growth analysis has been proposed by Nagin and Land (1993); see also articles in the special issue of Land (2001). For further examples, see, e.g., Nagin (1999), Nagin and Tremblay (2001), and Muthén (2001b).

3.4 Latent transition analysis

Latent transition analysis is a form of latent class analysis where the multiple measures of the latent classes are repeated over time and where across-time transitions between classes are of particular interest. Here, latent categorical variables are used to capture fundamental latent variable constructs in a system of regression relations akin to SEM.

The latent transition model is an example of the use of multiple latent class variables c and is therefore not directly incorporated in the framework specified above. Muthén (2001b) showed how multiple latent class variables can be analyzed using a confirmatory latent class analysis with a single latent class variable including all the possible latent class combinations, applying equality restrictions among the measurement parameters. Nevertheless, this does not handle multiple time points with parameter restrictions such as first-order Markov modeling for the latent class variables. Latent transition analysis incorporated in the general is a topic for future research.

An overview of latent transition modeling issues is given in Collins and Wugalter (1992) and Reboussin, Reboussin, Liang and Anthony (1998). For examples, see, e.g., Collins, Graham, Rousculp and Hansen (1997), Graham, Collins, Wugalter, Chung and Hansen (1991), and Kandel, Yamaguchi and Chen (1992).

3.5 Logistic regression mixture analysis

Logistic regression analysis with latent classes is interesting to consider as a special case of latent class analysis with covariates. The model was proposed by Follman and Lambert (1989) and considers a single binary u. It may be expressed for class k as

$$logit = -\tau_k + \boldsymbol{\kappa}_u \, \mathbf{x}, \tag{37}$$

which is a special case of (28) where the logit intercept, i.e. the negative of the threshold τ, varies across class but the slopes do not.

In (23), $\boldsymbol{\gamma}_{c_k} = \mathbf{0}$ so that the covariates are assumed to not influence the class membership. Follman and Lambert (1989) considered an application where two types of blood parasites were killed with various doses of poison. In this application, the assumption of $\boldsymbol{\gamma}_{c_k} = \mathbf{0}$ is natural because class membership existed before the poison was administered and was not influenced by it. Follman and Lambert (1989) discuss the identification status of the model.

Even in this simple form, however, logistic regression mixture analysis is difficult to apply in practice, probably because of the limited information available with only a single binary u in addition to the covariates $\mathbf{x}$. This is most likely why the analysis has not caught on in practice. In contrast, latent class analysis with covariates using multiple u variables is typically a well-behaved analysis method.

4. Modeling Framework C

Consider next the special case C of the general modeling framework shown in Figure 1. Framework C is characterized by adding categorical latent variables, denoted by the circle c in Figure 1, to framework A. Particular models include a variety of mainstream statistical and psychometric topics. To be discussed here are finite mixture modeling, latent profile analysis, growth mixture modeling, and mixture SEM.

It is interesting to compare framework C with framework B. Framework B can be seen as containing models that use latent classes to explain relationships among observed variables. A more fundamental idea can, however, be extracted from latent class approaches. Different classes can have different parameter values and, unlike the latent class model, even different model types. In other words, the idea of unobserved heterogeneity can be taken a step further using categorical latent variables. This further step is taken in framework C, and also in the subsequent general framework D.[5)]

In framework C, the SEM parameterization is generalized to multiple latent classes, adding a subscript k. This is analogous to the multiple-group situation, except that group is unobserved. In what follows, this generalization of (1) and (2) will be understood. Here, multivariate normality of $\mathbf{y}$ conditional on $\mathbf{x}$ and class is assumed. This implies that the resulting mixture distribution, not conditioning on class, is allowed to be strongly non-normal. In the Mplus framework of Muthén and Muthén (1998-2001; Appendix 8), the mixture modeling allows

[5)] Mplus examples of framework C models are given at www.statmodel.com/mplus/examples/mixture.html.

every parameter of framework A to vary across the latent classes.

4.1 Finite mixture modeling of multivariate normals

A straightforward case of framework C is finite mixture modeling of multivariate distributions. Here, the continuous latent variables of $\boldsymbol{\eta}$ in Figure 1 are not used. It is assumed that for class k, $\mathbf{y}$ is distributed as $N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. This is a special case of the latent class generalization of (1) where there are no factors, $\boldsymbol{\mu}_k = \boldsymbol{\nu}_k$, $\boldsymbol{\Sigma}_k = \boldsymbol{\Theta}_k$. There are two different reasons why such a mixture model would be of interest, (i) to fit a non-normal distribution and (ii) to study substantively meaningful mixture components (latent classes).

The flexibility of the normal mixture model to fit highly skewed data was recognized already by Pearson (1895); for a review, see McLachlan and Peel (2000, pp. 14-17, 177-179). For example, a lognormal univariate distribution is very well fit by a 2-class mixture with equal variances. Figure 5 shows a 2-class example. At the top is shown the mixture distribution, that is the skewed distribution that would be seen in data. At the bottom are shown two normal mixture component distributions that when mixed together by the class probabilities π and $(1-\pi)$ perfectly describe the distribution at the top. If the interest is in fitting a model to data from the distribution at the top, the 2-class mixture model can be used to produce mixed maximum-likelihood estimates,

$$\hat{\mu}_m = \hat{\pi}\,\hat{\mu}_1 + (1-\hat{\pi})\,\hat{\mu}_2, \tag{38}$$

$$\hat{\sigma}_m = \hat{\pi}\,(\hat{\mu}_1^2 + \hat{\sigma}_1) + (1-\hat{\pi})\,(\hat{\mu}_2^2 + \hat{\sigma}_2) - \hat{\mu}_m^2, \tag{39}$$

using the subscript m to denote the mixed estimates for the distribution at the top. The delta method can be used to compute standard errors. The idea of using mixed estimates has for example been discussed in missing data modeling using the pattern-mixture approach, see, e.g. Little and Wang (1996), Hogan and Laird (1997), and Hedeker and Rose (2000).

In many cases, however, the mixture components have a fundamental substantive meaning, where there are theoretical reasons for individuals to behave differently and have different antecedents and consequences. Here, mixed estimates such as (38), (39) are not of interest, but the focus is on the parameters of the different mixture component distributions. There may for example be biological/genetic reasons for the existence of different mixture components, such as with the two kinds of trypanosomes in Section 3.5.

Mixture modeling in applications where there are substantive reasons to investigate different latent classes relates to cluster analysis. Cluster analysis using finite mixture modeling has been proposed as a strong alternative to conventional clustering techniques, see, e.g., McLachlan and Peel (2000). A classic example is the Fisher's iris data analyzed in Everitt and Hand (1981). Four measures corresponding to sepal and petal lengths and widths were used to classify 150 iris flowers. Here, there were three known species of iris present and the interest

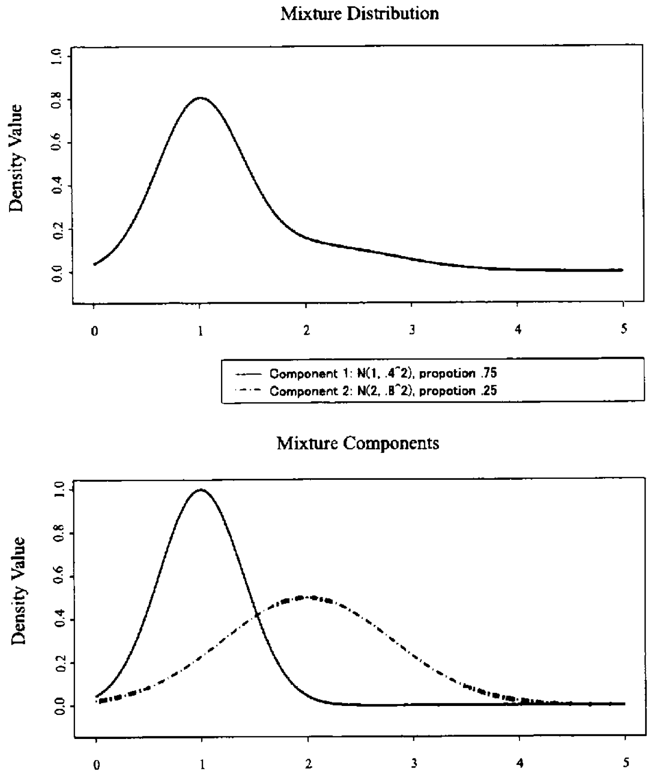


Figure 5: A mixture of two components

was in how well the classification could be recovered. This particular example also illustrates the possible difficulty in fitting mixture models with class-varying variances, with multiple maxima and possible non-convergence or convergence to singular covariance matrices depending on starting values.

An excellent overview of finite mixture modeling is given in McLachlan and Peel (2000). This source also gives a multitude of examples. The iris data example is available at the Mplus web site given above.

4.2 Latent profile analysis

In contrast to the analysis of the iris example above, latent profile analysis applies a structure to the covariance matrices, assuming uncorrelated outcomes conditional on class,

$$\mathbf{\Sigma}_k = \mathbf{\Theta}_k = \begin{pmatrix} \theta_{11_k} & 0 & 0 & 0 \\ 0 & \theta_{22_k} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \theta_{pp_k} \end{pmatrix}. \tag{40}$$

With class-varying means $\boldsymbol{\mu}_k$, latent profile analysis is therefore analogous to latent class analysis. In actual analyses, models with class-invariant variances in (40) are better behaved in terms of convergence. It is interesting to note that the latent class analysis does not face this choice given that means and variances of the categorical variables of $\mathbf{u}$ are not represented by separate parameters. Relationships among latent class, latent profile, and factor analysis models are described in Bartholomew (1987), Gibson (1959), and Lazarsfeld and Henry (1968).

4.3 Growth mixture modeling

The growth modeling of Section 2.1 uses continuous latent variables in the form of random effects. The continuous latent variables capture unobserved heterogeneity in terms of individual differences in growth over time. In many applications, however, there are more fundamental forms of unobserved heterogeneity that cannot be well captured by continuous latent variables but require categorical latent variables. The classes of the categorical latent variable can represent latent trajectory classes. Substantive theories motivating latent trajectory classes are common in many different fields, such as with normative and non-normative development in behavioral research and disease processes in medicine.

As for latent profile analysis, growth mixture modeling imposes a structure on the covariance matrix for each class. Unlike latent profile analysis, however, growth mixture modeling does not assume uncorrelated outcomes given class. Instead, further heterogeneity within class is represented by random effects that influence the outcomes at all time points, causing them to be correlated.

Assume for example the following quadratic growth model for individual i in class k $(k = 1, 2, \ldots, K)$,

$$y_{it} = \eta_{0i} + \eta_{1i}\, a_{kt} + \eta_{2i}\, a_{kt}^2 + \epsilon_{it}, \tag{41}$$

where y_{it} $(i = 1, 2, \ldots, n;\ t = 1, 2, \ldots, T)$ are outcomes influenced by the random effects η_{0i}, η_{1i}, and η_{2i}. In line with Section 2.1, the time scores of a enter into the $\mathbf{\Lambda}_k$ matrix. The residuals ϵ_{it} have a $T \times T$ covariance matrix $\mathbf{\Theta}_k$, possibly varying across the trajectory classes $(k = 1, 2, \ldots, K)$. The random effects are related to the covariates $\mathbf{x}$,

$$\eta_{0i} = \alpha_{0k} + \boldsymbol{\gamma}_{0k}'\, \mathbf{x}_i + \zeta_{0i}, \tag{42}$$

$$\eta_{1i} = \alpha_{1k} + \boldsymbol{\gamma}_{1k}'\, \mathbf{x}_i + \zeta_{1i}, \tag{43}$$

$$\eta_{2i} = \alpha_{2k} + \boldsymbol{\gamma}_{2k}'\, \mathbf{x}_i + \zeta_{2i}. \tag{44}$$

The residuals ζ_i have a 3×3 covariance matrix $\mathbf{\Psi}_k$, possibly varying across classes k ($k = 1, 2, \ldots, K$). It is clear that this model fits into framework C in line with how the growth model fit into framework A.

The growth mixture model offers great flexibility in across-class parameter differences. The different shapes of the latent trajectory classes are typically characterized by the class-varying α_k parameters holding $\mathbf{\Lambda}_k$ class-invariant. Certain classes may require class-specific variances $\mathbf{\Psi}_k$ and $\mathbf{\Theta}_k$. In addition, different classes may have different relations to $\mathbf{x}$ corresponding to class-varying γ_k coefficients.

A special case of the growth mixture model is obtained as a continuous-outcome version of the latent class growth analysis presented in Section 3.3. This type of modeling, proposed by Nagin and introduced into PROC TRAJ in SAS specifies $\mathbf{\Psi}_k = \mathbf{0}$, $\mathbf{\Theta}_k = \theta\ \mathbf{I}$. In contrast, growth mixture modeling allows for individual variation within each class through $\mathbf{\Psi}_k$. The latent class growth analysis typically requires many more classes to fit the same data and often several of the classes represent only minor variations in trajectories and not fundamentally different growth forms.

Muthén et al. (in press) present a growth mixture model suitable for randomized trials. In conventional growth modeling the treatment effects can be modeled as affecting the trajectories after the treatment has started. The Muthén et al. generalization addresses the common situation that treatment effects are often different for different kinds of individuals. It allows treatment effects to vary across latent trajectory classes for the repeated measures.

For a technical description of growth mixture modeling, see Muthén and Shedden (1999) and Muthén and Muthén (1998-2001: Appendix 8). For examples, see, e.g. Muthén and Shedden (1999), Muthén and Muthén (2000), Muthén (2001a, b), Muthén, Brown, Masyn, Jo, Khoo, Yang, Wang, Kellam, Carlin and Liao (in press), and Li, Duncan, Duncan and Acock (2001).

4.3.1 A growth mixture modeling example of reading failure

An example clarifies the analysis opportunities presented by growth mixture modeling. Section 2.1.1 introduced a reading data example with phonemic awareness development in Kindergarten related to end of grade 2 spelling performance. Figure 2 suggests heterogeneity in the phonemic awareness development, with a group of children having a close to zero growth rate in Kindergarten. Reading research points to a subgroup of children who experience reading failure by third grade. It is therefore of interest to see if early signs of a failing group can be found earlier, and perhaps as early as end of Kindergarten. Two analyses are presented here as illustration (see also Muthén, Khoo, Francis, & Boscardin, in press). First, a growth mixture analysis with 1 to 5 classes was made of the four phonemic awareness outcomes. Second, this growth mixture model was extended to include in the same analysis the spelling test outcome from the end of second grade, letting the mean and variance of this outcome vary as a function of the latent trajectory classes. Both models clearly fit in framework C. Maximum-likelihood estimation by Mplus was used.

A conventional linear, single-class random effects growth model fits well in this case ($\chi^2(5) = 7.49$, $n = 582$) and shows significant variation in the intercept and slope growth factors. Such a good mean and covariance structure fit can, however, be obtained even when the true model is a growth mixture model with more than one class (see, e.g. Muthén, 1989). Fitting linear models with 2, 3, 4, and 5 latent classes pointed to a steady improvement of the Bayesian information criterion that rewards a high log likelihood and a low number of parameters. Given the particular interest in a low, failing class, a choice does not have to be made between the 3-, 4-, and 5-class solutions since they all resulted in the same formation of a lowest class of 56% of the children. Figure 6 shows the estimated growth (solid line) and the corresponding observed trajectories, where the latter are obtained by using "pseudo-classes", i.e. the selection of individuals are obtained by random draws from their estimated posterior probabilities as suggested in Bandeen-Roche et al. (1977) and Muthén et al. (in press).

Adding the second-grade spelling test to the growth mixture model shows the predictive power of the Kindergarten information from two years earlier. The extended growth mixture model analysis showed that the means of the spelling test were significantly different across the 3 classes. Box plots of the spelling test scores based on pseudo-class assignments into the 3 classes are given in Figure 7.

4.4 Mixture SEM

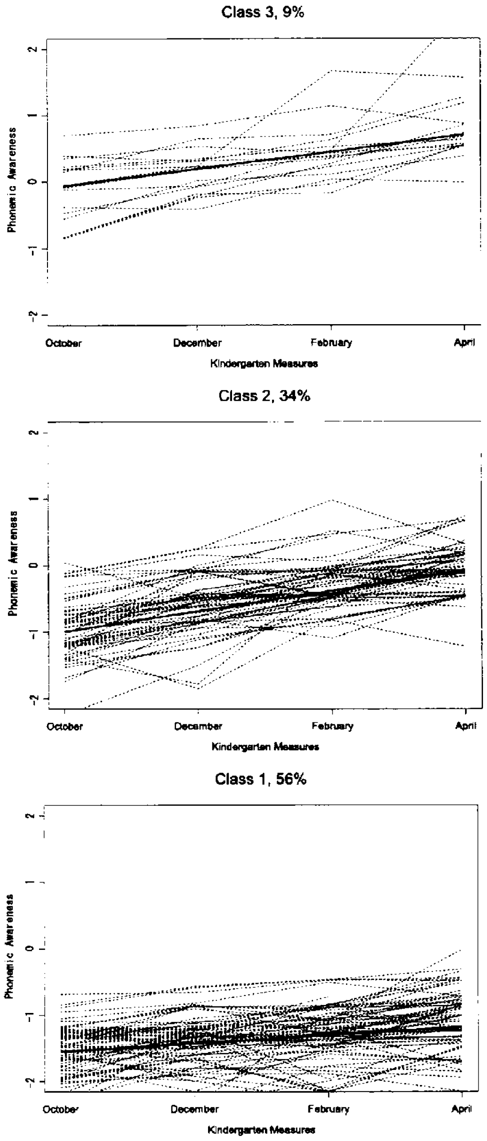
Mixture SEM will be mentioned only briefly in this article. It follows from the discussion in Section 2 that mixture SEM and growth mixture modeling fit into the same modeling framework. Mixture SEM includes mixture linear regression, mixture path analysis, factor mixture analysis, and general mixture SEM. Consider as an example factor mixture analysis, where for class k

$$E(\mathbf{y}_k) = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \, \boldsymbol{\alpha}_k, \tag{45}$$

$$V(\mathbf{y}_k) = \boldsymbol{\Lambda}_k \, \boldsymbol{\Psi}_k \, \boldsymbol{\Lambda}'_k + \boldsymbol{\Theta}_k. \tag{46}$$

Analogous to multiple-group analysis, a major interest is in across-class variation in the factor means, variances, and covariances of $\boldsymbol{\alpha}_k$, $\boldsymbol{\Psi}_k$. The model is similar to growth mixture analysis in that continuous latent variables, i.e. the factors, are used to describe correlations among the outcomes conditional on class as in (46). Lubke, Muthén and Larsen (2001) studied the identifiability of the factor mixture model. The special case of measurement invariance for all the outcomes, i.e. no class variation in $\boldsymbol{\nu}$, $\boldsymbol{\Lambda}$, is of particular interest because it places the factors in the same metric so that $\boldsymbol{\alpha}_k$, $\boldsymbol{\Psi}_k$ comparisons are meaningful. However, Lubke, Muthén and Larsen (2001) point to analysis difficulties with near-singular information matrix estimates when fitting such full invariance models. These difficulties are not shared by the growth mixture model, which typically imposes equality of ν parameters across time and across class and has few if any free parameters in Λ.

For overviews and examples of factor mixture analysis and mixture SEM, see, e.g., Arminger and Stein (1997), Arminger, Stein and Wittenberg (1998), Blåfield (1980), Dolan and van der Maas (1998), Hoshino (2001), Jedidi, Jagpal and DeSarbo(1997), Jedidi, Ramaswamy, DeSarbo and Wedel (1996), McLachlan and Peel (2000), and Yung (1997).


Class 3, 9%

Class 2, 34%

Class 1, 56%

Figure 6: 3-class growth mixture model for phonemic awareness

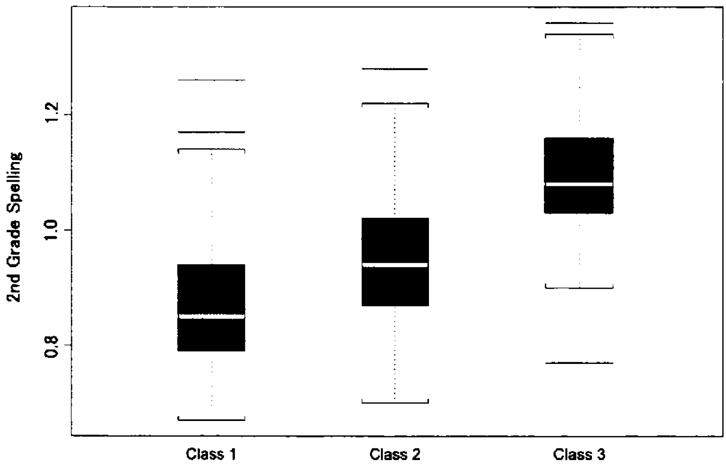


Figure 7: Box plots for second-grade spelling scores in three phonemic awareness classes

5. Framework D

Consider next the most general case D of the modeling framework shown in Figure 1. Framework D is characterized by adding categorical latent variable indicators **u** to framework C. Framework D clearly shows the modeling generality achieved by a combination of continuous and categorical latent variables. This unified framework is an example of the whole being more than the sum of its parts. It is powerful not only because it contains many special cases, but also because it suggests many new modeling combinations. Particular models include a wide variety of statistical and psychometric topics. To be discussed here are complier-average causal effect modeling, combined latent class and growth mixture modeling, prediction of distal outcomes from growth shapes, discrete-time survival mixture analysis, non-ignorable missing data modeling, and modeling of semicontinuous outcomes.[6]

5.1 Complier-average causal effect modeling

Complier-average causal effect (CACE) modeling is used in randomized trials where a portion of the individuals randomized to the treatment group choose

[6] Mplus examples of framework D models are given at www.statmodel.com/mplus/examples/mixture.html as well as at www.statmodel.com/mplus/examples/penn.html

to not participate ("noncompliers"). Although developed for this specialized application, CACE modeling involves interesting general latent variable modeling issues. In particular, CACE modeling illustrates how latent variables are used in mainstream statistics to capture missing data on categorical variables. Here, the mixture modeling focuses on estimating parameters for substantively meaningful mixture components, where these mixture components are inferred not only from the outcomes but also from auxiliary information. CACE modeling represents a transition from framework C to framework D, where in addition to the framework C observed data information, a minimal amount of information on class membership is added in the form of a single u variable observed for part of the sample.

In a randomized trial, it is common to have noncompliers among those invited to treatment, that is, some individuals do not show up for treatment or do not take the medication. Because of randomization, a equal-sized group of noncompliers is also present among control group individuals, although the non-compliance status does not manifest itself. The noncomplier and complier groups are typically not similar, but may differ with respect to several characteristics such as age, education, motivation, etc. The assessment of treatment effects with respect to say the mean of an outcome is therefore complicated. Four main approaches are common. First, "intent-to-treat" analysis makes a straightforward comparison of the treatment group to the control group. This may lead to a diluted treatment effect given that not everyone in this group has received treatment. Second, one may compare compliers in the treatment group with the controls. Third, compliers in the treatment group may be compared to the combined group of noncompliers in the treatment group and everyone in the control group. Fourth, compliers in the treatment group may be compared to compliers in the control group. Only the last approach compares the same subset of people in the treatment and control groups, but presents the problem that this subset is not observed in the control group. This problem is solved by CACE mixture modeling.

CACE modeling can be expressed by the framework D combination of (1), (2) generalized to include the latent class addition of (23) - (29). The probability of membership in the two latent classes as a function of covariates may be expressed by the logistic regression (23), while the outcome y is expressed by the mixed linear regression,

$$y_{ik} = \alpha_k + \gamma_k \, I_{ik} + \zeta_{ik}, \tag{47}$$

where I denotes the 0/1 treatment/control dummy variable. Here, α_k captures the different y means for individuals in the absence of treatment. CACE modeling typically takes $\gamma_k = 0$ for the noncomplier class.

In statistical analysis this situation is viewed as a missing data problem. Data are missing on the binary compliance variable for individuals in the control group, while data on this variable are present for the treatment group. The framework D conceptualization is that non-compliance status is a latent class variable, where this latent class variable becomes observed for treatment group individuals. Hence. the latent class variable captures missing data on a categorical variable. Although the choice between the two conceptualizations may seem as only a matter of semantics. as described below the latent variable approach suggests extensions of CACE modeling using connections with psychometric modeling that have potential value in randomized trials.

The fact that latent class status is known for treatment group individuals can be handled in two equivalent ways in the Mplus analysis. First. training data may be used to indicate that membership in the non-compliance class is impossible for complying individuals in the treatment group and that membership in the compliance class is impossible for non-complying individuals in the treatment group. Second. a binary latent class indicator u defined to be identical to the latent class variable may be introduced in line with the latent class analysis of framework B. With 0 representing noncompliance and 1 representing compliance, the u variable has fixed parameter values. $P(u = 1|\text{compliance class}) = 1$. $P(u = 1|\text{non-compliance class}) = 0$. The variable u is observed for treatment group individuals and missing for control group individuals. This second approach shows that CACE modeling belongs in framework D and also suggests a generalization. In psychometrics. the typical approach is to seek observed indicators for latent variables. An attempt could be made to measure the variable u also among controls. for example by asking individuals before randomization how likely they are to participate in the treatment if chosen. Several different measures $\mathbf{u}$ could be designed and used as latent class analysis indicators in line with framework B modeling.

For background on CACE modeling. see, e.g., Angrist. Imbens and Rubin (1996) and Frangakis and Baker (2001). For examples, see, e.g., Little and Yau (1998). Jo (2001a, b, c), and Jo and Muthén (2001).[7]

5.2 *Combined latent class and growth mixture analysis*

Figure 1 shows clearly that framework D can combine the framework B latent class analysis with the framework C growth mixture modeling. As an example, Muthén and Muthén (2000) analyzed the NLSY data discussed in Section 3, where it was of interest to relate latent classes of individuals with respect to antisocial behavior at age 17 to latent trajectory classes for heavy drinking ages 18-30. Here. a latent class variable was used for each of the two sets of variables, the latent class measurement instrument for antisocial behavior and the repeated measures of heavy drinking. Using the confirmatory latent class analysis technique described in Muthén (2001b). these two latent class variables can be analyzed together. This gives estimates of the relationships between the two classifications. To the extent that the two classifications are highly correlated, a latent class

[7] Data and Mplus input for the Little and Yao (1998) example is available on the Mplus web site.

analysis measurement instrument can improve the classification into the latent trajectory classes. This approach is of potential importance, e.g. using the latent class measurement instrument as a screening device in a treatment study, where different treatments are matched to different kinds of trajectory classes.

5.3 Prediction from growth shapes

Muthén and Shedden (1999) used the latent trajectory classes in a growth mixture model of heavy drinking as predictors of distal outcomes in the form of binary u variables, such as indicators of alcohol dependence. Predicting from the heavy drinking growth factors faces the potential problem of a highly non-linear relationship given that a growth factor aquires its meaning in conjunction with other growth factors. For example, in a study of problematic behavior such as heavy drinking, a low slope growth factor value has a different meaning if the intercept factor value is high ("chronic" development) than when it is low ("normal" development). Given that the latent trajectory classes can represent different shapes of development, prediction from the latent classes is a powerful approach.

5.4 Special uses of **u** *indicators*

The framework D addition of $\mathbf{u}$ to framework C not only adds latent class analysis type features but also provides several unexpected additional modeling possibilities. Muthén and Masyn (2001) show how $\mathbf{u}$ can be used as event history indicators in discrete-time survival analysis. This model corresponds to a single-class latent class analysis, but Muthén and Masyn (2001) also explore different types of mixture survival models. Muthén and Brown (2001) show how $\mathbf{u}$ can be used as missing data indicators for missingness on $\mathbf{y}$. This leads to an approach to study non-ignorable missing data in mixture modeling, for example where missingness is related to latent trajectory classes. Muthén (2001) shows how $\mathbf{u}$ can be used to indicate zero or "floor" values for $\mathbf{y}$, that is values that represent absence of an activity. Such data are frequently seen in behavioral research given that time of onset varies across individuals. It is the strength of framework D that these seemingly disparate models can be integrated and used in new combinations to provide answers to more probing research questions.

6. Conclusions

This article has provided an overview of statistical analysis with latent variables. In psychometrics it is typical to use latent variables to represent theoretical constructs. The constructs themselves are of key interest and a focus is on measuring different aspects of the constructs. In statistics, latent variables are more typically used to represent unobserved heterogeneity, sources of variation, and missing data. The latent variables are often not of key interest but are included to more correctly model the data. Unobserved heterogeneity is typically represented by random effects, i.e. continuous latent variables, a common example being growth modeling in the form of the mixed linear modeling (multilevel modeling) to capture individual differences in growth. Continuous latent variables are also used to represent sources of variation in hierarchical cross-sectional data, to let the model properly reflect a cluster sampling scheme and to estimate variance components. Cluster analysis considers unobserved heterogeneity in the form of categorical latent variables, i.e. latent classes, in order to find homogeneous groups of individuals. Finite mixture modeling with categorical latent variables is a rigorous approach to such cluster analysis. Missing data corresponds to latent variables that are either continuous or categorical.

The article discussed a general latent variable modeling framework that uses a combination of continuous and categorical latent variables to give a unifying view of psychometric and statistical latent variable applications. This framework shows connections between different modeling traditions and suggests interesting extensions. It is the hope that this general latent variable modeling framework stimulates a better integration of psychometric and statistical development. Also, it is hoped that this framework provides substantive researchers with an analysis tool that is both powerful and easy to understand in order to more readily respond to the complexity of their research questions. Ongoing research by the author aims at further extensions of the modeling framework.

Acknowledgements

The research was supported by grant K02 AA 00230-01 from NIAAA, by grant 40859 from NIMH, and by a grant from NIDA. The author thanks Tihomir Asparouhov for valuable discussions and computational assistance. Katherine Masyn provided helpful assistance with displaying the results. The work has also benefitted from discussions in Bengt Muthén's Research Apprenticeship Course and Hendricks Brown's Prevention Science Methodology group.

References

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley & Sons.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, **91**, 444-445.

Arminger, G. & Stein, P. (1997). Finite mixtures of covariance structure models with regressors. *Sociological Methods & Research*, **26**, 148-182.

Arminger, G., Stein, P., & Wittenberg, J. (1998). Mixtures of conditional mean- and covariance-structure models. *Psychometrika*, **64**, 475-494.

Asparouhov, T. & Muthén, B. (2002). Full-information maximum-likelihood estimation of general two-level latent variable models. Draft.

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., & Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, **92**, 1375-1386.

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

Blåfield, E. (1980). Clustering of observations from finite mixtures with structural information. Unpublished doctoral dissertation, Jyväskyla studies in computer science, economics, and statistics, Jyväskyla, Finland.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley.

Browne, M. W. & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311-359). New York: Plenum Press.

Bryk, A. S. & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods, 2nd ed.* Newbury Park, CA: Sage Publications.

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311-359). New York: Plenum Press.

Clogg, C. C. & Goodman, L. A. (1985). Simultaneous latent structural analysis in several groups. In Tuma, N. B. (Ed.), *Sociological Methodology* (pp. 81-110). San Francisco: Jossey-Bass Publishers.

Collins, L. M. & Sayer, A. (Eds.) (2001). *New methods for the analysis of change.* Washington, D.C.: APA.

Collins, L. M. & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, **27**, 131-157.

Collins, L. M., Graham, J. W., Rousculp, S. S., & Hansen, W. B. (1997). Heavy caffeine use and the beginning of the substance use onset process: An illustration of latent transition analysis. In K. Bryant, M. Windle, & S. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance use research* (pp. 79-99). Washington DC: American Psychological Association.

Dayton, C. M. & Macready, G. B. (1988). Concomitant variable latent class models. *Journal of the American Statistical Association*, **83**, 173-178.

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

Everitt, B. S. & Hand, D. J. (1981). *Finite mixture distributions.* London: Chapman and Hall.

Follmann, D. A. & Lambert, D. (1989). Generalizing logistic regression by nonparametric mixing. *Journal of the American Statistical Association*, **84**, 295-300.

Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, **87**, 476-486.

Frangakis C. E. & Baker, S. G. (2001). Compliance subsampling designs for comparative research: Estimation and optimal planning. *Biometrics*, **57**, 899-908.

Gibson, W. A. (1959). Three multivariate models: factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika*, **24**, 229-252.

Goldstein, H. (1995). *Multilevel statistical models.* London: Edward Arnold.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, **61**, 215-231.

Graham. J. W., Collins, L. M., Wugalter, S. E., Chung, N. K., & Hansen, W. B. (1991). Modeling transitions in latent stage- sequential processes: A substance use prevention example. *Journal of Consulting and Clinical Psychology*, **59**, 48-57.

Hecht, R. H. (2001). Multilevel modeling with SEM. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 89-127). Mahaw, NJ: Lawrence Erlbaum Associates.

Hedeker, D. & Rose, J. S. (2000). The natural history of smoking: A pattern-mixture random-effects regression model. In J. Rose, L. Chassin, C. Presson, & J. Sherman (Eds.), *Multivariate applications in substance use research* (pp. 79-112). Hillsdale, NJ: Erlbaum.

Heijden, P. G. M., Dressens, J., & Bockenholt, U. (1996). Estimating the concomitant-variable latent-class model with the EM algorithm. *Journal of Educational and Behavioral Statistics*, **21**, 215-229.

Hogan, J. W. & Laird, N. M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, **16**, 239-258.

Hoshino, T. (2001). Bayesian inference for finite mixtures in confirmatory factor analysis. *Behaviormetrika*. **28**, 37-63.

Hosmer, D. W. (1973). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics*, **29**, 761-770.

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

Jedidi, K., Ramaswamy, V., DeSarbo, W. S., & Wedel, M. (1996). On estimating finite mixtures of multivariate regression and simultaneous equation models. *Structural Equation Modeling*, **3**, 266-289.

Jennrich, R. I. & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, **42**, 805-820.

Jo, B. (2001a). Estimation of intervention effects with noncompliance: Alternative model specifications. Accepted for publication in the *Journal of Educational and Behavioral Statistics*.

Jo, B. (2001b). Model misspecification sensitivity analysis in estimating causal effects of interventions with noncompliance. Accepted for publication in *Statistics in Medicine*.

Jo, B. (2001c). Statistical power in randomized intervention studies with noncompliance. Forthcoming in *Psychological Methods*.

Jo, B. & Muthén, B. (2001). Modeling of intervention effects with noncompliance: A latent variable modeling approach for randomized trials. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 57-87). Mahaw, NJ: Lawrence Erlbaum Associates.

Jöreskog, K. G. & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, **70**, 631-639.

Kandel, D. B., Yamaguchi, K., & Chen, K. (1992). Stages of progression in drug involvement from adolescence to adulthood: Further evidence for the gateway theory. *Journal of Studies of Alcohol*, **53**, 447-457.

Kaplan D. & Elliott, P. R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling*, **4**, 1-23.

Khoo, S. T. & Muthén, B. (2000). Longitudinal data on families: Growth modeling alternatives. In J. Rose, L. Chassin, C. Presson, & J. Sherman (Eds.), *Multivariate applications in substance use research* (pp. 43-78). Hillsdale, NJ: Erlbaum.

Laird, N. M. & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.

Land, K. C. (2001). Introduction to the special issue on finite mixture models. *Sociological Methods & Research*, **29**, 275-281.

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

Li, F., Duncan, T. E, Duncan, S. C., & Acock, A. (2001). Latent growth modeling of longitudinal data: A finite growth mixture modeling approach. *Structural Equation Modeling*, **8**, 493-530.

Little, R. J. & Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, **52**, 98-111.

Little, R. J. & Yau, L. H. Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods*, **3**, 147-159.

Lindstrom, M. J., & Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, **83**, 1014-1022.

Lubke, G., Muthén, B., & Larsen, K. (2001). Global and local identifiability of factor mixture models. Submitted for publication. University of California, Los Angeles.

MacCallum, R. C. & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, **51**, 201-226.

McDonald, R. & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology*, **42**, 215-232.

McLachlan, G. J. & Krishnan, T. (1997). *The EM algorithm and extensions.* New York: John Wiley & Sons.

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

Muthén, B. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 213-238). Hillsdale, NJ : Erlbaum Associates.

Muthén, B.(1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, **54**, 557-585.

Muthén, B. (1990). Mean and covariance structure analysis of hierarchical data. UCLA Statistics Series #62, August 1990.

Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, **28**, 338-354.

Muthén, B. (1994). Multilevel covariance structure analysis. In J. Hox & I. Kreft (Eds.), Multilevel Modeling, a special issue of *Sociological Methods & Research*, **22**, 376-398.

Muthén, B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological Methodology* (pp. 453-480). Boston: Blackwell Publishers.

Muthén, B. (2001a). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In Collins, L. M. & Sayer, A. (Eds.), *New methods for the analysis of change* (pp. 291-322). Washington, D.C.: APA.

Muthén, B. (2001b). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1-33). Mahaw, NJ: Lawrence Erlbaum Associates.

Muthén, B. (2001c). Two-part growth mixture modeling. Draft. University of California, Los Angeles.

Muthén, B. & Brown, C. H. (2001). Non-ignorable missing data in a general latent variable framework. Draft. University of California, Los Angeles.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., Wang, C. P., Kellam, S., Carlin, J., & Liao, J. (in press). General growth mixture modeling for randomized preventive interventions. Forthcoming in *Biostatistics*.

Muthén, B. & Curran, P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, **2**, 371-402.

Muthén, B., Kao, C. F., & Burstein, L. (1991). Instructional sensitivity in mathematics achievement test items: Applications of a new IRT-based detection technique. *Journal of Educational Measurement*, **28**, 1-22.

Muthén, B. & Khoo, S. T. (1998). Longitudinal studies of achievement growth using latent variable modeling. In John B. Cooney (Ed.), Latent Growth Curve Analysis, a special issue of *Learning and Individual Differences*, **10**, 73-101.

Muthén, B., Khoo, S. T., Francis, D., & Kim Boscardin, C. (in press). Analysis of reading skills development from Kindergarten through first grade: An application of growth mixture modeling to sequential processes. In S. R. Reise & N. Duan (Eds), *Multilevel modeling: Methodological advances, issues, and applications*. Mahaw, NJ: Lawrence Erlbaum Associates.

Muthén, B. & Masyn, K. (2001). Discrete-time survival mixture analysis. University of California, Los Angeles.

Muthén, B. & Muthén, L. (2000). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, **24**, 882-891.

Muthén, L. & Muthén, B. (1998-2001). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*, **4**, 139-157.

Nagin, D. S. & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, **31**, 327-362.

Nagin, D. S. & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: a group-based method. *Psychological Methods*, **6**, 18-34.

Nestadt, G., Hanfelt, J., Liang, K. Y., Lamacz, M., Wolyniec, P., & Pulver, A. E. (1994). An evaluation of the structure of schizophrenia spectrum personality disorders. *Journal of Personality Disorders*, **8**, 288-298.

Pearson, K. (1895). Contributions to the theory of mathematical evolution, II: skew variation. *Philosophical Transactions of the Royal Society of London A*, **186**, 343-414.

Raudenbush, S. W. (2001). Toward a coherent framework for comparing trajectories of individual change. In Collins, L. M. & Sayer, A. (Eds.), *New methods for the analysis of change* (pp. 35-64). Washington, D.C.: APA.

Reboussin, B. A., Reboussin, D. M., Liang, K. Y., & Anthony, J. C. (1998). Latent transition modeling of progression of health-risk behavior. *Multivariate Behavioral Research*, **33**, 457-478.

Rindskopf, D. & Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. *Statistics in Medicine*, **5**, 21-27.

Takane, Y. & DeLeeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, **52**, 393-408.

Uebersax, J. S., & Grove, W. M. (1990). Latent class analysis of diagnostic agreement. *Statistics in Medicine*, **9**, 559-572.

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

Received December 2001. Final version accepted January 2002.