Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (eds.), New Developments and Techniques in Structural Equation Modeling (pp. 1-33). Lawrence Erlbaum Associates. (#86)

1

Latent Variable Mixture Modeling

Bengt O. Muthén
University of California, Los Angeles

This chapter discusses models with latent variables that are continuous and/or categorical. It also gives an overview of modeling issues related to cross-sectional analysis using latent class models, modeling of longitudinal data using latent class models, and modeling of longitudinal data using a combination of continuous and categorical latent variables (growth mixture models). A series of examples are presented. The analyses are carried out within a general latent variable modeling framework shown in the appendix using the Mplus program (Muthén & Muthén, 1998). Mplus input specifications for these analyses can be obtained from www.statmodel.com. To introduce the analyses, a brief overview of modeling ideas is presented in Figs. 1.1 to 1.3.

The top left part of Fig. 1.1 shows three distributions for a continuous outcome variable y. The idea is that the data consist of different groups of individuals, but the group membership is not observed. The two broken curves represent the distribution

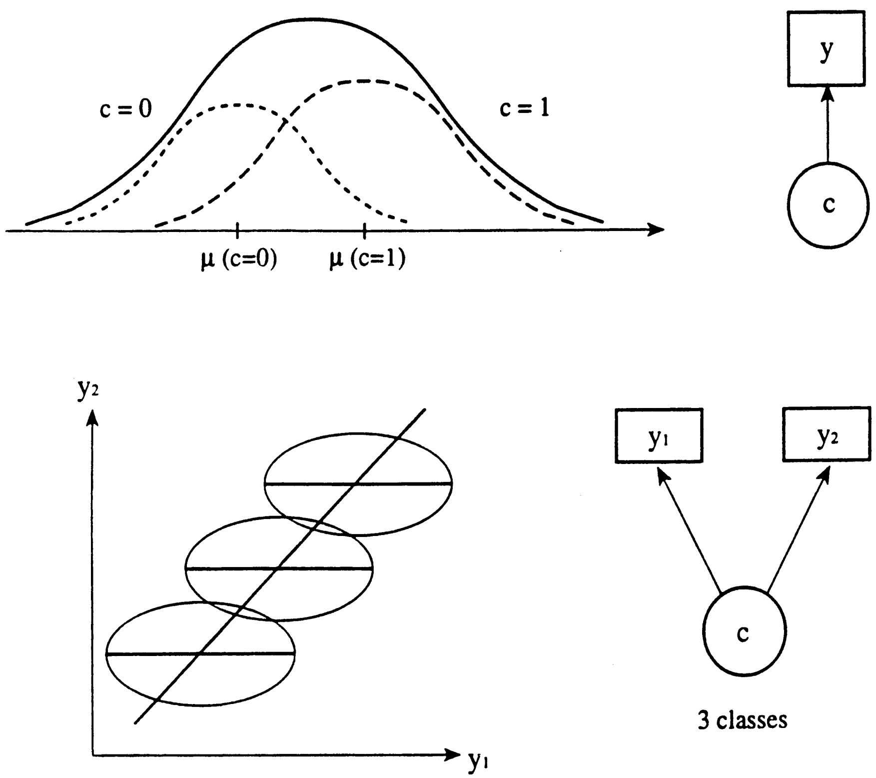


FIG. 1.1. Mixture modeling.

of y for two latent classes, $c = 0$ and $c = 1$, which have different means. These two distributions are not observed, but only the mixture of the two, shown by the solid curve. There are many examples of such unobserved heterogeneity. An example often used in the statistic literature considers the length of fish in a stream. The analysis task is to find out how many cohorts of fish there are in the stream. In Fig. 1.1, there are two cohorts of fish, where the older fishes are longer. Alcohol researchers may consider brain wave responses to stimuli, measuring a P300 wave amplitude that is assumed to differ between individuals susceptible to alcohol dependence versus those who are not, with the interest in classifying individuals. Reading researchers may consider a latent class corresponding to reading disability and a class of normal readers with the interest in estimating the mean difference and classifying individuals as early as possible. The top right part shows the corresponding path diagram, using c to denote the latent categorical variable with two classes.

The bottom left part of Fig. 1.1 shows unobserved heterogeneity with respect to two continuous outcomes, y_1 and y_2. The line indicates a strong relationship between the outcomes, but this relationship is due to mixing three different

classes of individuals, each having unrelated outcomes. The corresponding path diagram is shown to the right, viewing y_1 and y_2 as indicators of the latent categorical variable c. This type of modeling is referred to as *latent profile analysis* or *latent class analysis* when the outcomes are categorical. The modeling has features similar to factor analysis in that it is assumed that a latent variable accounts for the association between the outcomes. This is also referred to as a conditional independence assumption, with the idea that if a sufficient number of classes is introduced, the independence is more and more likely to hold.

There is, however, a more profound aspect of the latent profile/class model that has interesting possibilities for model generalizations. This is that each latent class has different parameter values and possibly a different model. In latent profile/class analysis, the model is the same across classes—namely, an independence model. The parameter values differ across classes. For latent profile analysis, the mean for each outcome variable changes over classes, and in latent class analysis with binary outcomes, the probability of each outcome variable changes over classes. More complex class-specific models and changes in parameter values across classes are, however, possible. This realization leads to a huge set of new modeling opportunities indicated in Figs. 1.2 and 1.3.

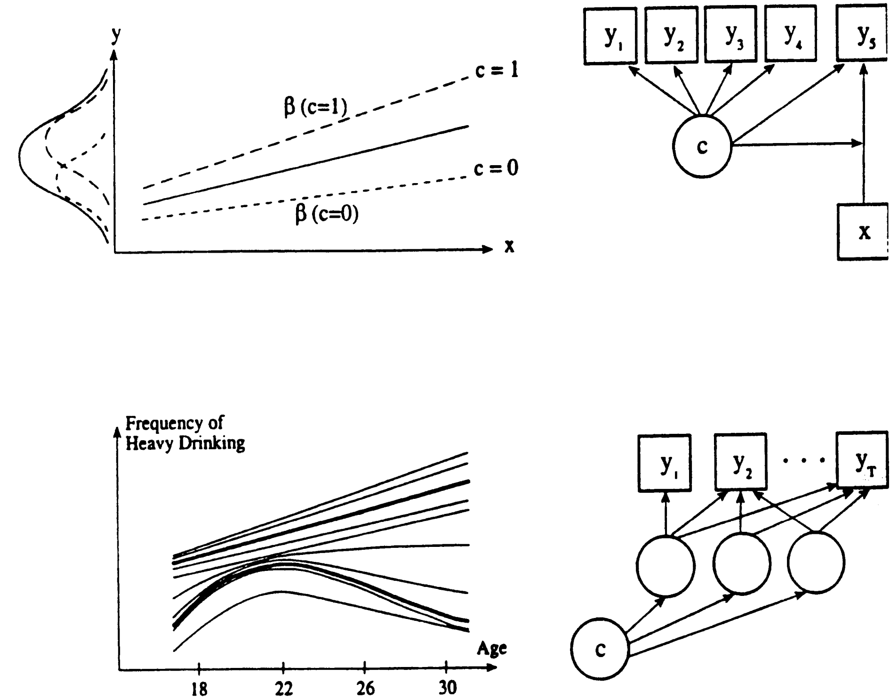


FIG. 1.2. Mixture modeling.

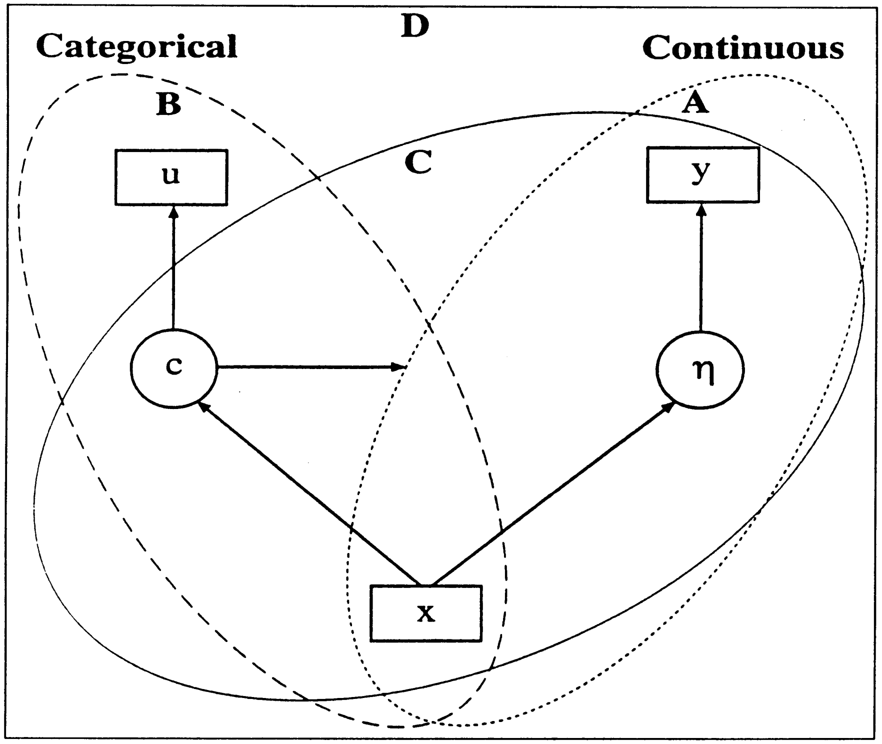


FIG. 1.3. General modeling framework.

The top part of Fig. 1.2 shows a regression analysis with unobserved heterogeneity. The solid line gives the regression for the mixture, which is not correct for either class. The right part of the picture shows a generalization of the latent profile/class modeling, where $y_1 - y_4$ are class indicators, but where the key interest is in capturing the variation across class in the regression of y_5 on x. Here the arrow from c to y_5 indicates that the intercept in y_5 differs across classes, whereas the arrow from c to the arrow from x to y_5 indicates that the slope in the regression differs across classes. In this way, the most important class variation in parameter values is for the regression model, and the latent profile part is merely a vehicle for making it easier to identify the classes.

The bottom part of Fig. 1.2 shows latent variable modeling where different classes have different growth models. An example is development of the frequency of heavy drinking, ages 18 to 30. The general population is likely to be quite heterogeneous with respect to this development. The left part of the figure shows two average growth curves as solid curves. A normative class shows a typical increase in this behavior in the early 20s, with a subsequent decline.

However, less prevalent classes are present, such as the class of individuals whose heavy drinking does not decline in their late 20s. Within each class, there is further heterogeneity as indicated by the thinner curves. The path diagram on the right shows three continuous latent variables corresponding to the three growth factors of quadratic growth influencing the repeated measures on y. The arrows from the categorical latent variable c to these growth factors indicate that their means vary across the latent classes as seen in the graph on the left.

The modeling ideas in Figs. 1.1 and 1.2 may be summarized as in Fig. 1.3. Here, the modeling framework labeled D is the Mplus framework given in the appendix. A fuller discussion of this framework with references is given in Muthén (in press). Three different ellipses represent various special cases of the general framework.

On the right, an ellipse labeled A represents a framework where the latent variables η are continuous, including exploratory and confirmatory factor analysis, structural equation modeling (SEM), and latent growth curve modeling. This is the framework of conventional SEM as it has been practiced for the last couple of decades, using software such as AMOS, EQS, and LISREL.

On the left, an ellipse labeled B represents a framework where the latent variables c are categorical, including latent class analysis with or without covariates. Typically in the past, modeling using Framework B has been separate from modeling in Framework A.

The ellipse labeled C represents a framework where a combination of categorical and continuous latent variables is used. This includes latent profile analysis and mixture cluster analysis, both excluding continuous latent variables. Complier-average causal effect estimation in randomized trials is another application, discussed by Jo

TABLE 1.1
Summary of Techniques Using Latent Classes

Class	*Outcome/ Indicator Scale*	*Number of Time Points*	*Number of Outcomes/ Time Points*	*Within-Class Variation*
LCA	Categorical (u)	Single	Multiple	No
LPA	Continuous (y)	Single	Multiple	No
LCGA	Categorical (u) Continuous (y)	Multiple	Single	No
LTA	Categorical (u)	Multiple	Multiple	No
GMM	Continuous (y)	Multiple	Single Multiple	Yes
GGMM	Categorical (u) Continuous (y)	Multiple	Single Multiple	Yes

LCA – latent class analysis, LPA – latent profile analysis, LCGA – latent class growth analysis, LTA – latent transition analysis, GMM – growth mixture modeling, GGMM – general growth mixture modeling.

and Muthén (chap. 3, this volume). Growth mixture modeling is an example where both categorical and continuous latent variables are used.

The square labeled D represents the general framework, adding direct indicators u for the categorical latent variables.

Because latent variable modeling with categorical latent variables is an emerging methodology, the summary of techniques using latent classes given in Table 1.1 may be useful. The techniques are defined by the characteristics given in the three columns. Here, LCA and LTA fall into Framework B, LPA falls into Framework C, LCGA falls into Framework B or C, GMM falls into Framework C, and GGMM falls into Framework D.

LATENT CLASS ANALYSIS

Latent class analysis (LCA) was introduced by Lazarsfeld and Henry (1968), Goodman (1974), Clogg (1995), and others. The setting is cross-sectional data with multiple items measuring a construct represented as a latent class variable. The aims are to identify items that indicate classes well, estimate class probabilities, relate class probabilities to covariates, and classify individuals into classes.

Consider the LCA model for the special case of binary outcomes u. Letting the categorical latent variable c have K classes ($c = k$; $k = 1, 2, \ldots, K$), the marginal probability for item $u_j = 1$ is

$$P(u_j = 1) = \sum_{k=1}^{K} P(c = k)\, P(u_j = 1 \mid c = k), \qquad (1)$$

while the joint probability of all us, assuming conditional independence, is

$$P(u_1, u_2, \ldots, u_r) = \sum_{k=1}^{K} P(c = k)\, P(u_1 | c = k) P(u_2 = 1 | c = k) \ldots P(u_r = 1 | c = k). \quad (2)$$

There are two types of parameters—the conditional item probabilities for each class and the class probabilities. In the Mplus framework, LCA parameters are expressed in logit form, where

$$P = \frac{1}{1 + e^{-L}}, \qquad (3)$$

$$L = \text{logit}[P] = \ln[P/(1-P)], \qquad (4)$$

for example, $L = 0$ gives $P = 0.50$, $L = -1$ gives $P = 0.27$, $L = 1$ gives $P = 0.73$, $L = -3$ gives $P = 0.05$, and $L = -10$ gives $P = 0.00005$.

LCA Estimation and Testing

A by-product of LCA is estimated class probabilities for each individual, analogous to factor scores in factor analysis. These are estimates of

$$P(c = k \mid u_1, u_2, \ldots, u_r) = \frac{P(c = k) P(u_1 | c = k) P(u_2 | c = k) \ldots P(u_r | c = k)}{P(u_1, u_2, \ldots, u_r)}. \qquad (5)$$

Note that each individual is allowed fractional class membership and may have nonzero values for several classes.

In Mplus, parameters are estimated by maximum-likelihood estimation via the EM algorithm, where c is seen as missing data. The EM algorithm maximizes the expected complete-data log likelihood conditional on $(u_{i1}, u_{i2}, \ldots, u_{ir})$ with respect to the parameters. The E step computes $E(c_i | u_{i1}, u_{i2}, \ldots, u_{ir})$ as the posterior probability for each class and $E(c_i\, u_{ij} | u_{i1}, u_{i2}, \ldots, u_{ir})$ for each class and u_j. The M step estimates $P(u_j | c_k)$ and $P(c_k)$ parameters by regression and summation over individual posterior probabilities, respectively. Multiple starting values are strongly recommended because the likelihood may have several different local maxima.

As an overall test, the likelihood-ratio χ^2 with H_1 as the unrestricted multinomial may be used, although with many items the chi-square approximation is poor due to small cell sizes. Models with different number of classes can be compared using the Bayesian information criterion (Schwartz, 1978)

$$BIC = -2 \log L + r \ln n, \qquad (6)$$

where r is the number of free parameters in the model. A low BIC value indicates a better fitting model.

LCA of Alcohol Dependence

Consider the example in Table 1.2, where in the National Longitudinal Survey of Youth (NLSY), nine diagnostic criteria for alcoholism were analyzed in a sample of 8,313 young adults (Muthén & Muthén, 1995). A three-class solution fit well as measured by the chi-square test against the unrestricted multinomial. In this solution, Class 1 is the most prevalent, with 75% showing low probabilities of endorsing the criteria. Class 2 has 21% of the individuals and has high probabilities for Larger and Major Role-Hazard, having to do with drinking larger amounts than planned and drinking while driving. Class 3 has 3% and has high probabilities for most criteria. Loosely speaking, one may think of Class 3 as an alcohol dependence class, Class 2 as an alcohol abuse class, and Class 1 as a problem-free class.

In this application, the classes appear to be ordered in the sense that the item probabilities increase from Class 1 to Class 2 to Class 3. However, this is not always the case in LCA. (For an example with unordered classes for antisocial

TABLE 1.2
NLSY 1989: Latent Class Analysis of DSM–III–R Alcohol Dependence Criteria ($n = 8{,}313$):

	Latent Classes				
	Two-Class Solution[1]		*Three-Class Solution*[2]		
	I	*II*	*I*	*II*	*III*
Prevalence	*0.78*	*0.22*	*0.75*	*0.21*	*0.03*
DSM–III–R Criterion	*Conditional Probablities of Fulfilling a Criterion*				
Withdrawal	0.00	0.14	0.00	0.07	0.49
Tolerance	0.01	0.45	0.01	0.35	0.81
Larger	0.15	0.96	0.12	0.94	0.99
Cut down	0.00	0.14	0.01	0.05	0.60
Time spent	0.00	0.19	0.00	0.09	0.65
Major role-hazard	0.03	0.83	0.02	0.73	0.96
Give up	0.00	0.10	0.00	0.03	0.43
Relief	0.00	0.08	0.00	0.02	0.40
Continue	0.00	0.24	0.02	0.11	0.83

Note. Source: Muthén and Muthén (1995).
[1]Likelihood ratio chi-square fit = 1,779, with 492 degrees of freedom.
[2]Likelihood ratio chi-square fit = 448, with 482 degrees of freedom.

behavior, see, Muthén & Muthén, 1999.) With ordered classes, one may ask what advantage LCA has versus doing regular factor analysis of binary outcomes using continuous latent variables (see e.g., Muthén, 1989). The answer is that LCA helps find clusters of individuals who are similar, whereas this is difficult in factor analysis. For these data, factor analysis suggested two factors. The estimated factor scores from the two-factor solution are plotted in Fig. 1.4.

Figure 1.4 shows that there is no natural cut points on the factors by which to divide individuals into having different levels of alcohol problems. However, the figure also includes the three classes found by the LCA. The three classes appear to be arranged along the principal axis of the two factors, the two factors being correlated around 0.7 in this example. This analysis shows that LCA is a vehicle for finding clusters of individuals, thereby complementing a regular factor analysis. A related observation is that data that fit well by a K-class model often fit well by a $K-1$-dimensional factor analysis model. For a related proof of an exact relationship for latent profile analysis, see Bartholomew (1987).

In terms of the alcohol problem diagnosis, discussions often center around how many criteria need to be fulfilled to give a certain diagnosis. Here, the LCA solution can be used as guidance as shown in Table 1.3 (see also Nestadt et al., 1994, for a similar analysis of schizophrenia criteria). Each individual can be classified into the class with largest posterior probability, and the classes can then

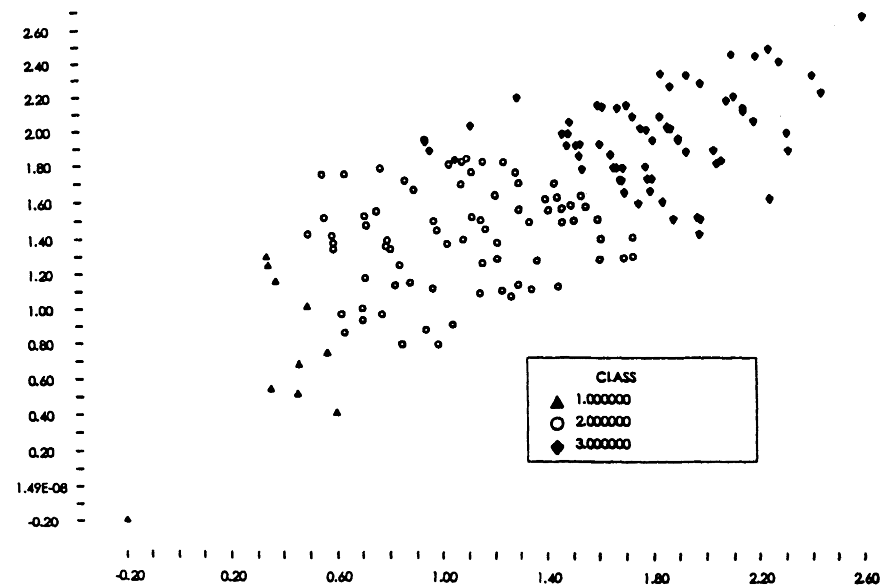


FIG. 1.4. Estimated factor scores from two-factor solution.

TABLE 1.3
Latent Class Membership by Number of DSM–III–R Alcohol Dependence Criteria Met ($n = 8{,}313$)

		Latent Classes				
		Two-Class Solution		*Three-Class Solution*		
Number of Criteria Met	*%*	*I*	*II*	*I*	*II*	*III*
0	64.2	5335	0	5335	0	0
1	14.0	1161	1	1161	1	0
2	10.2	0	845	0	845	0
3	5.6	0	469	0	469	0
4	2.6	0	213	0	211	2
5	1.4	0	116	0	19	97
6	0.8	0	68	0	0	68
7	0.5	0	42	0	0	42
8	0.5	0	39	0	0	39
9	0.3	0	24	0	0	24
%	100.0	78.1	21.9	78.1	18.6	3.3

be cross-classified with the number of criteria met. Table 1.3 shows that Class 1 membership supports requiring ≤1 criteria, Class 2 membership supports requiring 2–4 criteria, and Class 2 membership supports requiring ≥5 criteria fulfilled.

LATENT CLASS ANALYSIS WITH COVARIATES

LCA with covariates (concomitant variables) has been considered by Bandeen-Roche, Miglioretti, Zeger, and Rathouz (1997), Dayton and Macready (1988), Formann, (1992), and Heijden, Dressens, and Bockenholt (1996). This modeling considers a covariate x, where the probability that individual i falls in Class k of the latent class variable c is expressed through multinomial logistic regression as

$$P(c_i = k \mid x_i) = \frac{e^{\alpha_k + \gamma_k x_i}}{\sum_{k=1}^{K} e^{\alpha_k + \gamma_k x_i}}, \tag{7}$$

where $\alpha_K = 0$, $\gamma_K = 0$ so that $e^{\alpha_K + \gamma_K x_i} = 1$, implying that the log odds of comparing Class k to the last Class K is

$$log[P(c_i = k \mid x_i)/P(c_i = K \mid x_i)] = \alpha_k + \gamma_k x_i. \tag{8}$$

In addition,

$$logit\ [P(u_{ij} = 1 \mid c_i = k, x_i)] = \alpha_{u_k} + \kappa_{jk} x_i, \tag{9}$$

where the α_{u_k} s are the logit counterparts to the conditional item probabilities discussed earlier and κ_j is a direct effect parameter for the influence of x on u_j. Muthén and Muthén (1999) gave an example of LCA with covariates applied to antisocial behavior classes related to age, gender, and ethnicity.

The model in Eqs. (7) and (9) relates to those considered in Clogg and Goodman (1985), studying invariance across groups of individuals similar to multiple-group analysis in SEM. A multiple-group analysis is not needed because the model in Eqs. (7) and (9) is sufficient for capturing across-group differences in parameters when the groups are represented by dummy x variables. For example, the direct effect of a group dummy variable x on a certain u implies that measurement invariance does not hold, but that the groups differ in their conditional item probabilities within class. The direct effect may vary across classes.

The LCA model with covariates also allows for direct dependencies among the us conditional on class (i.e., violations of the conditional independence assumption). A certain u variable may be changed into an x variable, which allows for a

direct effect of this u on another u, conditional on class. The other u variables are still conditionally independent given class.

CONFIRMATORY LATENT CLASS ANALYSIS WITH SEVERAL LATENT CLASS VARIABLES

Confirmatory latent class analysis (CLCA) with several latent class variables was introduced in Goodman (1974), considering a panel study with two waves. Figure 1.5 shows an example based on the antisocial behavior analysis of Muthén and

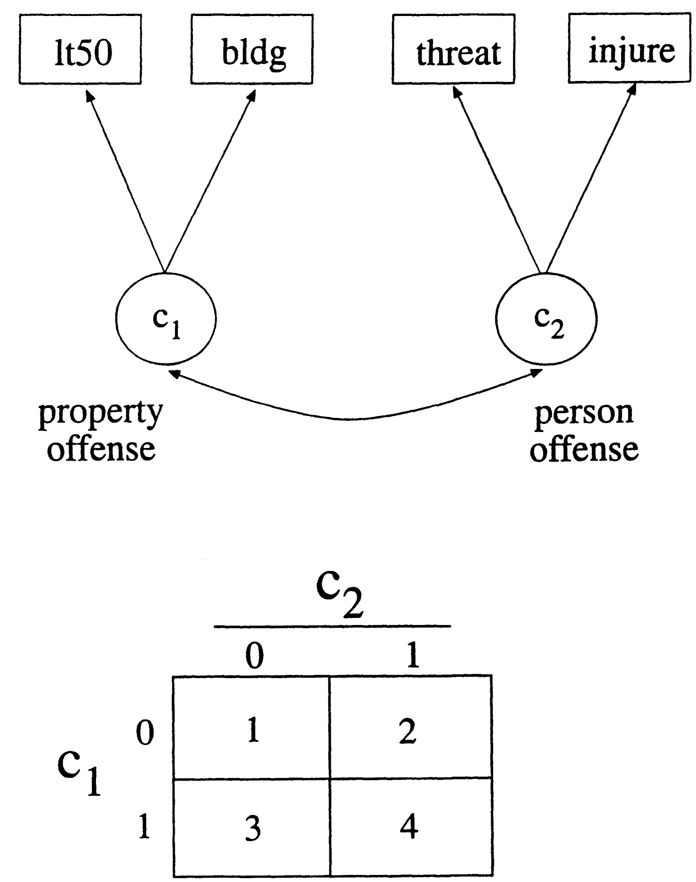


FIG. 1.5. Confirmatory latent class analysis with several latent class variables.

		Class 1	Class 2	Class 3	Class 4
C1	Lt50	1	1	2	2
	Bldg	3	3	4	4
C2	Threat	5	6	5	6
	Injure	7	8	7	8

FIG. 1.6. Restriction on the conditional probabilites for the items.

Muthén (1999). Among the three dimensions found in a factor analysis of 17 binary antisocial behavior items, items measuring two factors interpreted as property offense and person offense are considered. The property offense factor was well measured by the items *stole less than 50* and *broken into a building*, whereas the person offense factor was well measured by the items *seriously threaten* and *intent to injure*. The intent of the CLCA was to consider a dichotomized latent distribution for each of the two factors to capture non-normality of each factor and to divide individuals into classes based on each factor dimension. It is also possible to relate the corresponding two dichotomous latent class variables c_1 and c_2 to each other.

The way the CLCA model is drawn in Fig. 1.5 implies that the item probabilities for the threat and injure items should not vary across the c_1 classes, and the item probabilities for the lt50 and bldg items should not vary across the c_2 classes. In the Mplus framework, the analysis indicated by Fig. 1.5 is carried out by creating a latent class variable that combines the dichotomous c_1 and c_2 variables into one latent class variable with four classes. Using the numbering of classes shown at the bottom of Fig. 1.5, Fig. 1.6 shows the implied equality restrictions on the item probabilities, where each number corresponds to one parameter and repeated numbers indicate equalities.

LATENT CLASS GROWTH ANALYSIS

In longitudinal data, the multiple indicators of latent classes may correspond to repeated univariate outcomes at different time points. This is the situation considered in LCGA (see Nagin, 1999). Here the classes define different trends over

time in the item probabilities. For example, using a linear trend with an intercept and a slope,

$$\Lambda_{u_k} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T-1 \end{bmatrix}, \qquad (11)$$

the logits for the u items may be expressed as

$$\Lambda_{u_k} \eta_{ui} + \mathrm{K}_{u_k} x_i \qquad (11)$$

where η_{ui} contains the intercept and slope growth factors expressed as

$$\eta_{ui} = \alpha_{u_k} + \Gamma_{u_k} x_i. \qquad (12)$$

Here the K_{u_i} parameters capture effects on the *u*s of time-varying covariates, varying across time but not across individuals, and the Γ_{u_k} parameters capture effects of time-invariant covariates on the growth factors. The growth factors have fixed values conditional on x (for the random counterpart, see next section).

As an example, data from Jackson, Sher, and Wood (1999) were reanalyzed. This analysis considers the co-occurrence of alcohol and tobacco use disorders—that is, the *u* variables correspond to two processes. In a college sample of 450 students, Jackson et al. found five classes as shown in Fig. 1.7. The classes are defined by five repeated measures of alcohol disorder and, concurrently, five repeated measures of tobacco disorder. For example, Class 4 shows tobacco use disorder (see bottom panel), but no alcohol use disorder (see top panel). However, the Jackson et al. analysis does not take into account the time ordering for the measures, but use a regular LCA.

As an alternative analysis, LCGA was carried out for the two processes as shown in the path diagram of Fig. 1.8. For each process, two growth factors are used corresponding to linear growth, I and S. Each of the growth factors is influenced by a latent class variable specific to the process, so that the means of the growth factors change over classes. Three classes are used for each process, and the two sets of three-class variables c_1 and c_2 are related to each other.

The trends for each process are shown in Fig. 1.9. For each process, there is a chronic class with high probabilities throughout and a low class with low probabilities throughout. For the alcohol use disorders, there is also a declining class, whereas for the tobacco use disorders, there is an increasing class. These trends are approximations of those seen in Fig. 1.8, although the class probability curves in Fig. 1.8 for each process have been combined into fewer trend classes in Fig. 1.9. Figure 1.8

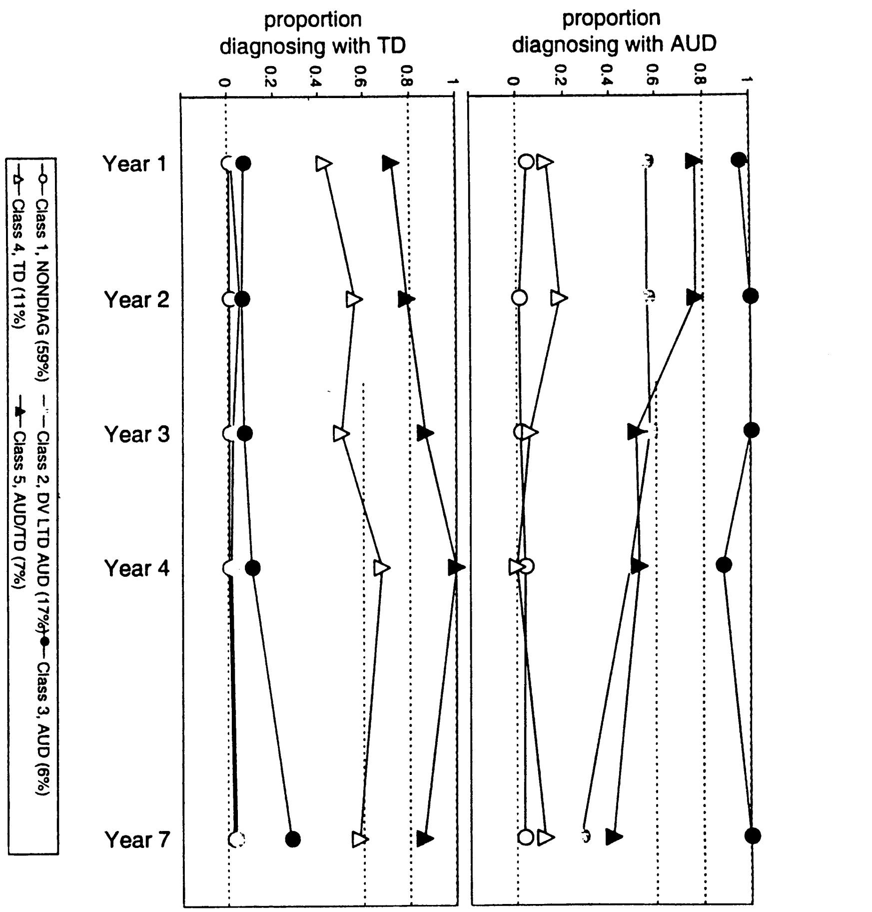
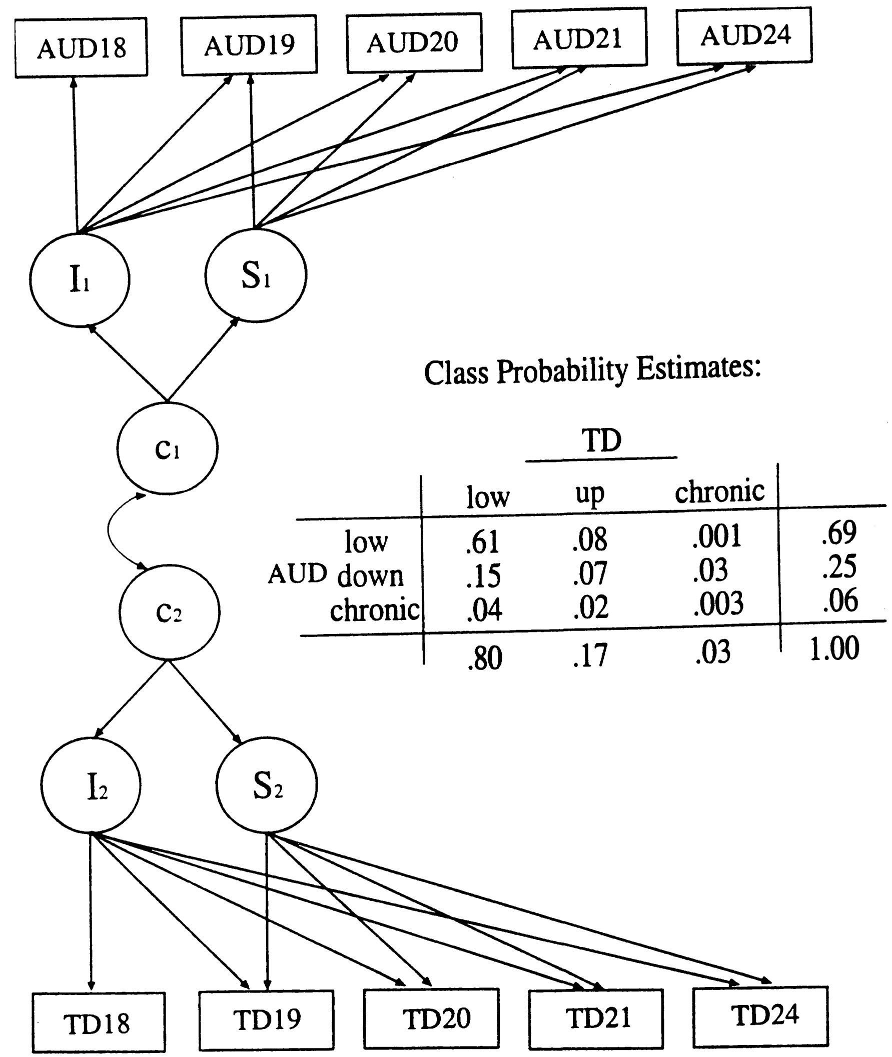


FIG. 1.7. Latent class analysis solution.

also shows the estimates of the joint probability table for c_1 and c_2. The four classes with the smallest probabilities were not included in the Jackson et al. analysis.

GROWTH MIXTURE MODELING

The rest of this chapter considers growth mixture modeling of repeated measures data. In analyzing such data, individual differences in development are typically captured by random effects using mixed linear modeling or multilevel modeling. These random effects represent continuous variation across individuals in growth features such as initial status and rate of change. Often, however, more fundamental individual differences in development are present and need to be allowed for to make the



		TD low	TD up	TD chronic	
AUD	low	.61	.08	.001	.69
	down	.15	.07	.03	.25
	chronic	.04	.02	.003	.06
		.80	.17	.03	1.00

FIG. 1.8. Co-occurrence of alcohol use and tobacco disorder.

modeling realistic. Such fundamental differences in development can be described by latent trajectory classes, where each class has a different random effect growth model. Random effects and trajectory classes are latent variables. Random effects are continuous latent variables, and trajectory classes are categorical latent variables. Growth mixture modeling (Muthén, in press; Muthén & Shedden, 1999; Muthén, Brown et al., 2000) uses both types of latent variables to represent individual differences in development, resulting in a very flexible repeated measures analysis.

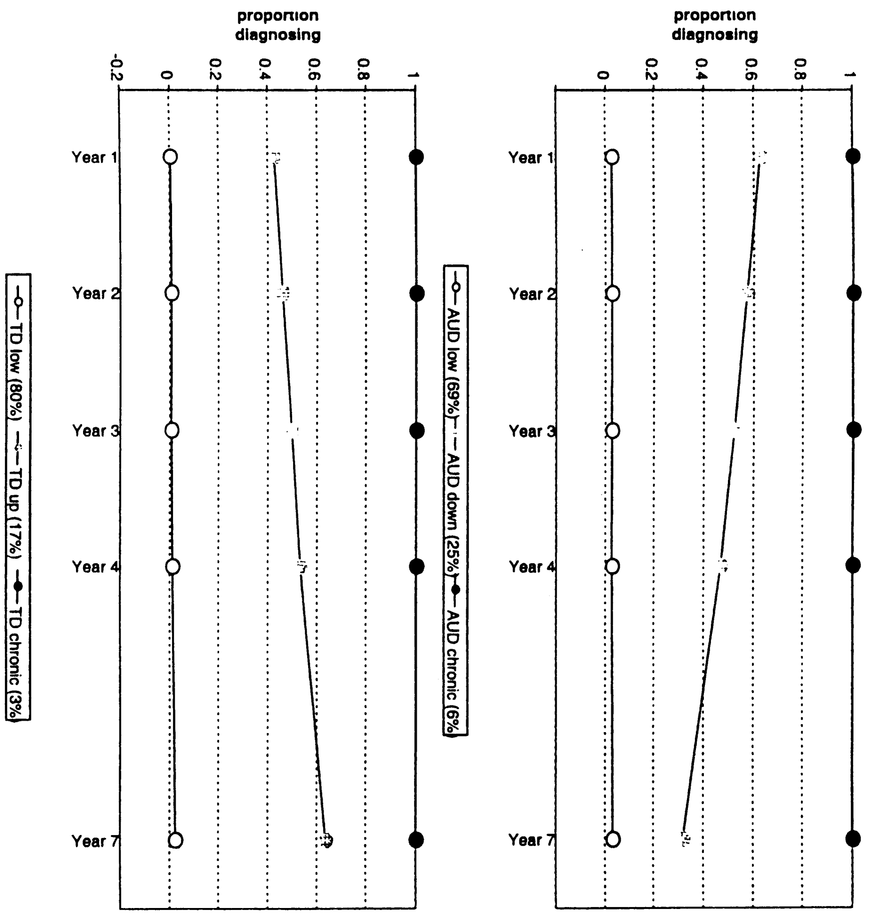


FIG. 1.9. Latent class growth analysis solution.

Two examples clarify the ideas. Children in early school grades may be on a developmental path of reading disability, others may show mild forms of reading problems, whereas still others progress normally. Children in school may exhibit serious aggressive/disruptive behavior in the classroom, others may show more common forms of such behavior, whereas others show no such problems. The average trajectories of the three classes in these examples are different, and there is individual variation around the average trajectories. It is important to distinguish among individuals in the different classes because membership in different classes may have different antecedents and consequences. Growth mixture modeling provides estimates of the class probabilities, the average trajectory for each class, the trajectory variation in each class, and estimates of each individual's most likely class membership. The probabilities of class membership can be related to background variables, and class membership can be used to predict other outcomes.

Conventional Random Effects Modeling in a Latent Variable Framework

Consider a quadratic growth model for continuous outcomes $y_{it}(i = 1, 2, \ldots, n; t = 1, 2, \ldots, T)$ that can be described by three random effects η_{0i}, η_{1i}, and η_{2i}, and time-specific residuals ϵ,

$$y_{it} = \eta_{0i} + \eta_{1i} x_{it} + \eta_{2i} x_{it}^2 + \kappa_t w_{it} + \epsilon_{it}. \qquad (13)$$

In the latent variable framework, the random effects are referred to as growth factors (i.e., continuous latent variables). Here it is assumed that individuals are measured at the same time points so that the time scores $x_{it} = x_t$ (deviations from this can be handled via missing data techniques). Assume that for substantive reasons it is of interest to define η_0 as an initial status growth factor, setting the time score $x_1 = 0$. Also for identification purposes, $x_2 = 1$. With equidistant times of observation, the model would typically have $x_t = 0, 1, 2, \ldots, T-1$. The time-specific residuals have zero means and covariance matrix Θ, typically with different variances and often with some off-diagonal elements to represent residual correlation across time.

The variation in the three-growth factors is expressed as,

$$\eta_{0i} = \alpha_0 + \gamma_0 w_i + \zeta_{0i}, \qquad (14)$$

$$\eta_{1i} = \alpha_1 + \gamma_1 w_i + \zeta_{1i}, \qquad (15)$$

$$\eta_{2i} = \alpha_2 + \gamma_2 w_i + \zeta_{2i}, \qquad (16)$$

where the αs are mean parameters and the ζs are residuals with zero means and covariance matrix Ψ and w is a time-variant covariate. Growth factors may be fixed or random. For example, with a fixed quadratic factor,

$$\Psi = \begin{pmatrix} \psi_{00} & & symm. \\ \psi_{10} & \psi_{11} & \\ 0 & 0 & 0 \end{pmatrix}, \qquad (17)$$

so that, conditional on w, there is no variation in η_2.

Conventional Growth Modeling of Reading Data. The reading data set is from the Early Assessment of Reading Skills (EARS) study a multiple-cohort study design repeatedly measuring children from kindergarten through third grade in a suburb of Houston. One key aim of the EARS study was to investigate if early identification of children at risk for poor academic outcomes could be made using longitudinal data. The outcome variable considered here is word-recognition skills repeatedly measured four times in Grade 1 and four times in Grade 2. In the current analyses, a subset of 411 children is considered. One cohort consisting of about half the children has data on all eight outcomes, whereas a second cohort has data on only the first four outcomes. The measurement occasions were October, December, February, and May of Grades 1 and 2. The children were also measured during kindergarten, and the current analyses use a measure of phonemic awareness at the end of kindergarten as a predictor of word-recognition development.

Initial exploration of the reading data suggests that a linear growth model is suitable for the eight time points. The growth model in Eqs. (13) to (16) without the η_2 term and without covariates is therefore used. The model allows the variances of the time-specific residuals ϵ to vary across time. The estimated model has a rather poor model fit [$n=411$, number of parameters $=13$, $\chi^2(31)=470.767$ ($p=0.0000$), log L $=-1145.785$, BIC $=2369.812$, CFI $=0.897$, RMSEA $=0.186$ (CI: .171, .201)]. Modification indexes point to a covariance between the residuals for time point 3 and 4 as by far the most important source of misfit, but freeing this parameter does not improve the model fit in important ways. The estimated mean line for this model is shown in Fig. 1.10 as a solid dark line together with observed data for a random sample of children. The individual observations suggest a considerable amount of heterogeneity in the word-recognition development, possibly including a separate low-achieving group of children marked by darker lines.

Conventional Growth Modeling of Aggression Data.

The aggression data set is from a school-based preventive intervention study carried out by the Johns Hopkins Prevention Center in Baltimore public schools, Grades 1 to 7. Here only the control group is analyzed. The outcome variable of interest is teacher ratings of each child's aggressive behavior in the classroom from Grade 1 to Grade 7. Teacher ratings of a child's aggressive behavior were made from fall and spring for the first two grades and every spring in Grades 3 to 7. The ratings were made using the TOCA-R instrument, using an average of 10 items, each rated on a 6-point scale from *almost never* to *almost always*. Information was also collected on other concurrent and distal outcomes, including school removal and juvenile court records. The current analyses focus on 80 boys in the control group.

The quadratic growth model of Eqs. (13) to (16) without covariates is applied to the nine time points for Grades 1 to 7 of the aggression data. The estimated model shows a slightly negative variance for the quadratic growth factor. Restricting the covariance matrix as in Eq. (17) gave a rather poor model fit [$n=80$, number of parameters $=16$, $\chi^2(38)=73.54$ ($p=0.0005$), log L $=-627.28$, BIC $=1324.67$, CFI $=0.887$, RMSEA $=0.108$ (CI: .070, .145)]. Modification indices point to correlations between time-specific residuals, but this does not improve the fit in important ways. A large modification index value is observed for the covariance between the intercept and quadratic growth factors, but this covariance cannot be included given the zero variance of the quadratic growth factor. It is not clear how to improve the fit of the model, although a suspicion is that the data contain more fundamental heterogeneity than can be captured by the conventional growth model used here. The estimated mean growth curve for the model with 16 parameters is shown in Fig. 1.11.

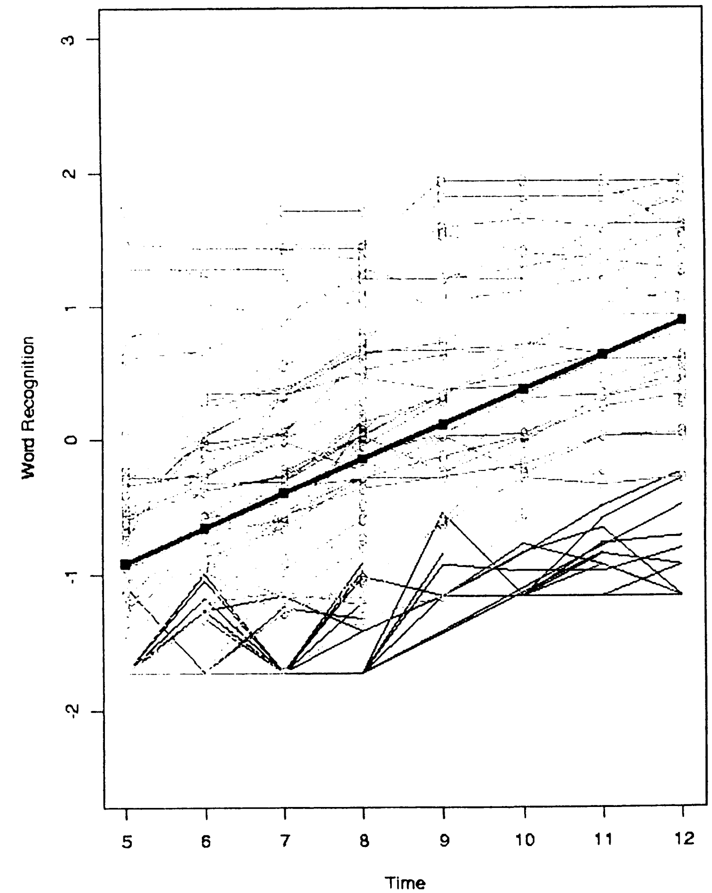


FIG. 1.10. Word-recognition development.

A Simple Growth Mixture Model

Model Specification.

The quadratic growth model of Eqs. (13) to (16) can be extended to a growth mixture model for K latent trajectory classes, where in Class k ($k=1,2,\ldots,K$),

$$\eta_{0i}=\alpha_{0k}+\gamma_{0k}w_i+\zeta_{0i}, \quad (18)$$

$$\eta_{1i}=\alpha_{1k}+\gamma_{1k}w_i+\zeta_{1i}, \quad (19)$$

$$\eta_{2i}=\alpha_{2k}+\gamma_{2k}w_i+\zeta_{2i}, \quad (20)$$

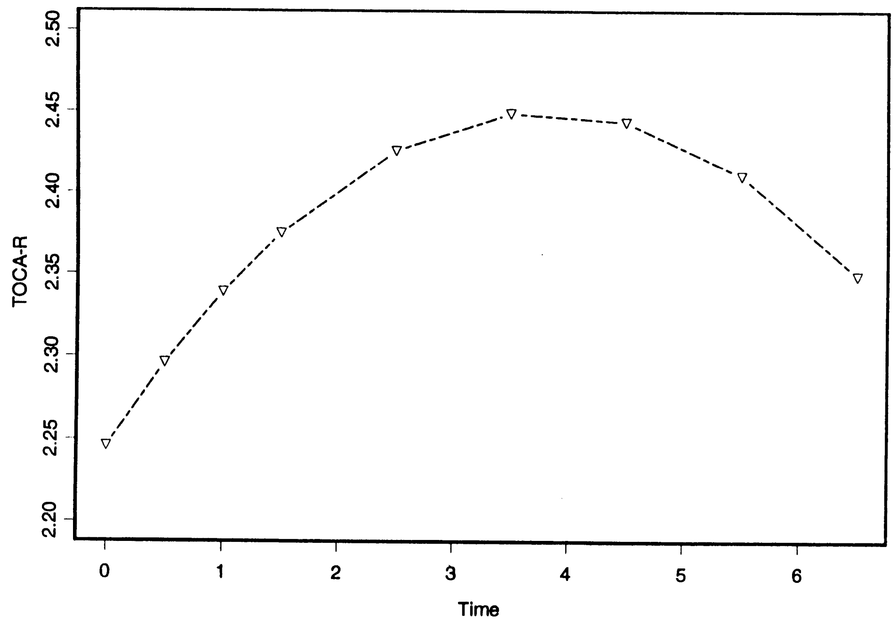


FIG. 1.11. Estimated mean growth curve of aggression.

Here α_k parameters vary across classes to capture different types of trajectories. If there is no covariate w, the αs are the means of the growth factors. For example, a class with a trajectory that is low and flat has a low α_0 value, a zero α_1 value, and a zero α_2 value, whereas a class with a trajectory that accelerates from an early low level and then decelerates has a low low α_0 value, a positive α_1 value, and a negative α_2 value. The γ_k parameters allow variation across class in how a covariate influences the growth factors. Class-specific covariance matrices Ψ_k are allowed for ζ. Class-specific covariance matrices Θ_k for ε in Eq. (13) are also allowed for. The growth curve shape can also vary across classes through class-specific x_{it} values in Eq. (13).

Model Fit. Tests of model fit require special attention in growth mixture modeling. It should be noted that a test against a completely unrestricted mean vector and covariance matrix, as in conventional structural equation modeling, is not used with mixture modeling. This is because the mixture modeling does not rely on normality assumptions where such summaries are natural. The set of sufficient statistics is nothing less than the raw data because the (mixture) distribution for the observed variables is not normal, but can be distinctly non-normal as a function of the mixture of normals.

However, the model can be evaluated based on the fit of first- and second-order moments in the following sense. For an estimated growth mixture model, estimated posterior probabilities of each individual's membership in each class are

obtained as shown in the appendix. These probabilities can be used to classify an individual into the class that he or she most likely belongs to. For each class, the raw data can be multiplied by the individual probabilities of that class to compute weighted sample mean vectors and covariance matrices for each class that can be compared to the corresponding model-estimated quantities.

The quality of a growth mixture model can also be evaluated based on the precision of the classification. For individuals classified into a given class, the average posterior probability of belonging to this class should be high, and the average posterior probability of belonging to each of the other classes should be low.

A key issue in growth mixture modeling is to determine the number of classes. For comparison of fit of models that have the same number of classes and are nested, the usual likelihood-ratio chi-square difference test of twice the difference in log likelihood values can be used. Comparison of models with different numbers of classes, however, cannot be done by likelihood-ratio chi-square. Instead this is accomplished by a Bayesian information criterion (BIC; Schwartz, 1978) as mentioned in connection with LCA.

Growth Mixture Modeling of Reading Data

The next analysis step for the reading data is to try to account for heterogeneity in development using growth mixture modeling. The first task is to decide on the number of latent trajectory classes. A useful procedure for exploring the number of classes is to first fit a series of models that have zero growth factor covariance matrices Ψ (i.e., assuming that individuals are homogeneous with respect to their growth). Variation is still allowed for across individuals through time-specific variances in Θ. This type of modeling has been proposed by Nagin (1999) and was referred to as latent class growth mixture analysis (LCGA) earlier, given that the within-class homogeneity specification is analogous to LCA. Here LCGA is used to derive starting values for a growth mixture model. In particular, the estimated LCGA growth factor means are used as starting values, letting Ψ be free.

The analysis of the number of classes relies to a large degree on the BIC values. Plotting the BIC values against the number of classes, the lowest point in the BIC curve is sought. For this part of the analysis, it is important to make a special investigation of the degree of class-invariance of the covariance matrices Ψ_k and Θ_k. Different degrees of invariance give different sets of BIC curves with different minima. In the reading data, a class with problematic word-recognition development is found. This class can be seen as a class at risk for reading failure in that it has almost zero growth rate. This class needs a class-specific growth factor covariance matrix Ψ that shows larger intercept and slope variance than the other classes. Only in the two-class model is this not needed because the lowest class is less clearly a failing class.

Figure 1.12 shows a plot of the BIC values for the reading data, using 1 to 6 classes. BIC values are shown for both LCGA and growth mixture modeling. The

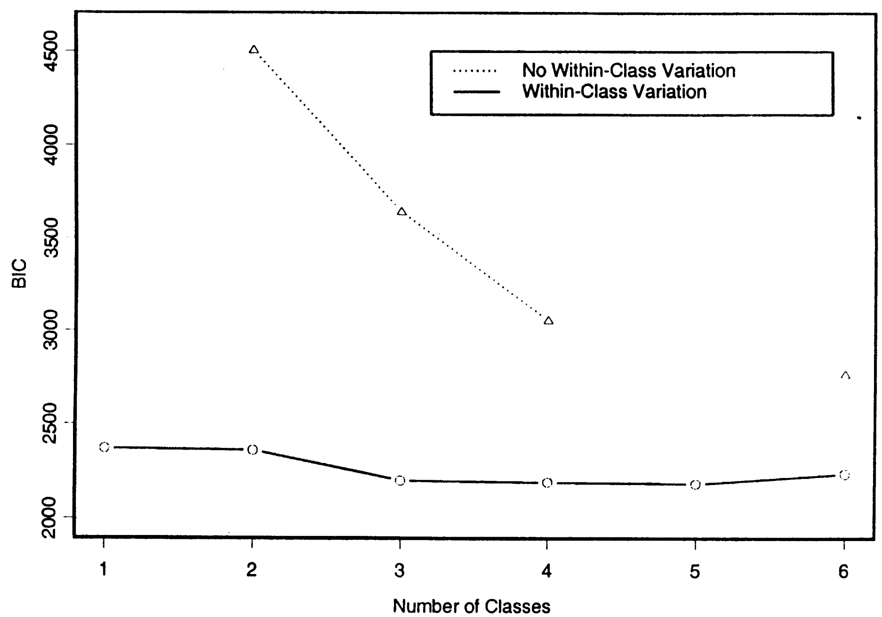


FIG. 1.12. BIC values for reading data.

one-class BIC value is for the conventional growth model. It is seen that BIC is improved by using more than one class, and that five classes seems optimal in the growth mixture modeling. The LCGA BIC values are considerably worse for any given number of classes, showing that it is important to allow for within-class growth heterogeneity for this application.

For the five-class growth mixture model, the hypothesis of class-invariant growth factor covariance matrix Ψ is strongly rejected in favor of allowing Class 1 to have a different growth factor covariance matrix [$\chi^2(3)=80.13$, $p<.0000$]. The estimated mean curves for the five-class growth mixture model are shown in Fig. 1.13. Classes 1 to 5 have class probabilities 0.14, 0.34, 0.30, 0.13, and 0.10. This means that the problematic Class 1 contains 56 children.

Figure 1.14 shows the quality of the classification using average posterior probabilities from the five-class model.

The posterior-probability-weighted sample means and the estimated means for the outcomes are shown in Fig. 1.15.

Muthén, Francis, Khoo, and Boscardin (in press) investigated the question of how early it was possible to classify children into the problematic Class 1. As was done in their investigation, it is possible to take the estimated model parameters as a given and study the posterior probabilities for a certain individual, varying the number of repeated measures available. The Muthén et al. investigation indicated that a good classification was already possible at the end of first grade.

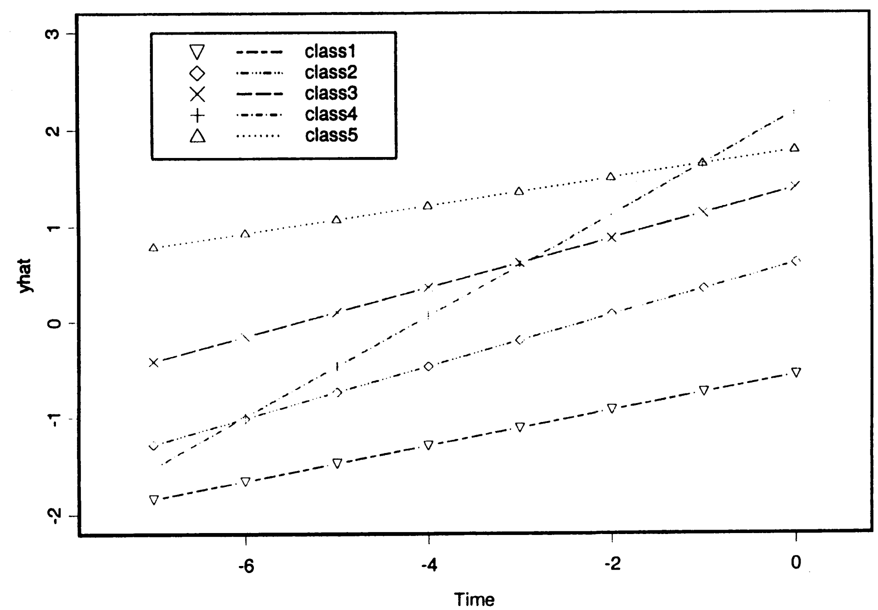


FIG. 1.13. Estimated mean curves for the five-class growth mixture model for reading data.

	Class 1	*Class 2*	*Class 3*	*Class 4*	*Class 5*
Class 1	**0.872**	0.121	0.000	0.008	0.000
Class 2	0.026	**0.850**	0.075	0.049	0.000
Class 3	0.000	0.080	**0.850**	0.022	0.048
Class 4	0.001	0.147	0.058	**0.794**	0.000
Class 5	0.000	0.000	0.055	0.000	**0.945**

FIG. 1.14. Average posterior probabilities from the five-class model.

Growth Mixture Modeling of Aggression Data

The BIC results for the aggression data using Classes 1 to 5 are shown in the top curve of Fig. 1.16. The BIC values indicate a better fit when allowing more than one class. It is seen that three classes is favored by BIC.

Inspection of the three-class growth mixture model shows that the likelihood could be significantly improved by allowing class-specific variances for the class with the lowest trajectory. This class shows considerably less fluctuation over

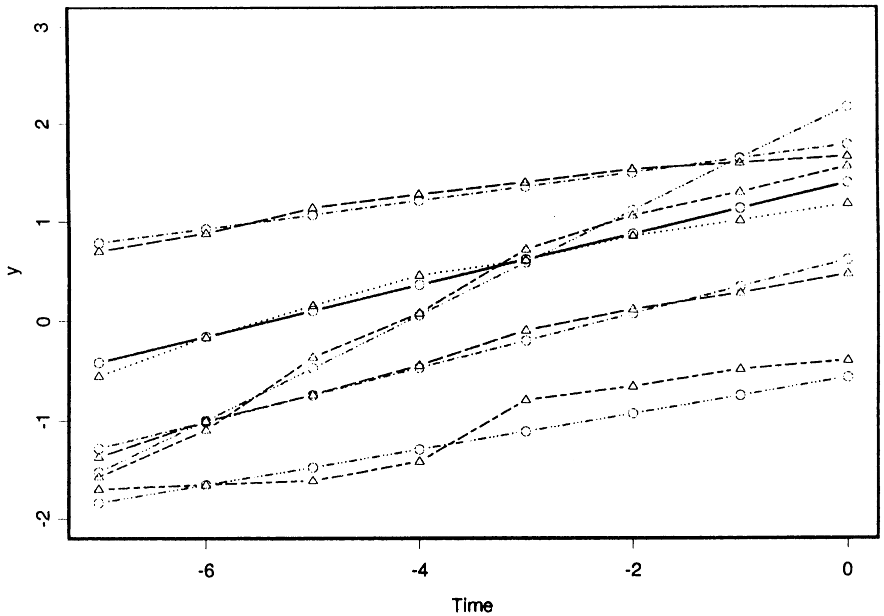


FIG. 1.15. Posterior-probability-weighted sample means and estimated means for reading data.

time in aggression than the other two classes. In particular, the intercept variance and the time-specific residual variances are lower for this class. The markedly lower BIC values in the bottom curve of Fig. 1.16 show the superior fit when allowing noninvariant variances for these models. With noninvariant variances, the lowest values are at Classes 3 and 5. The five-class solution has class probabilities 0.08, 0.45, 0.06, 0.09, 0.32, while the three-class solution has class probabilities 0.09, 0.52, and 0.39. The first and last classes are very similar in the two solutions.

For reasons of parsimony, and to have higher class counts, the three-class model is chosen here. Fig. 1.17 shows the estimated mean growth curves for the three-class model for Grades 1 to 7.

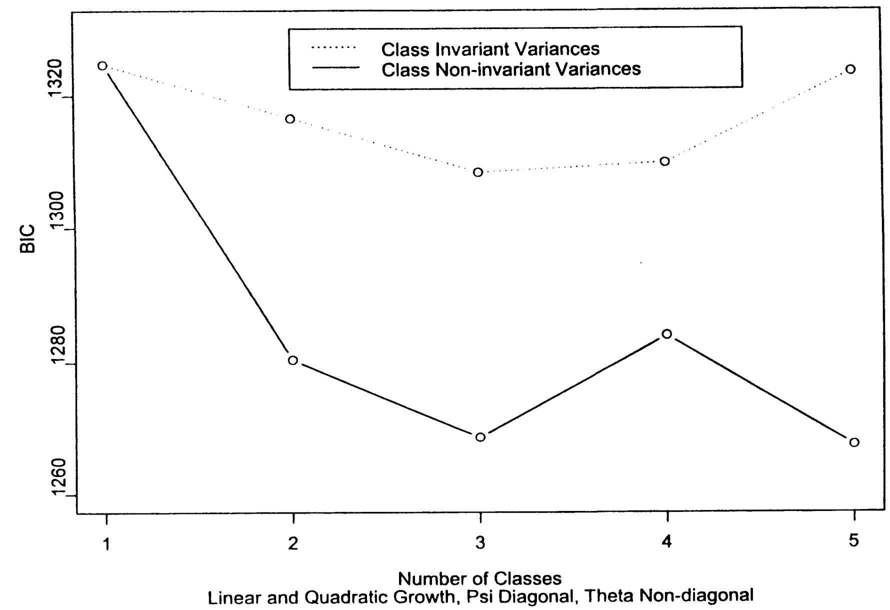


FIG. 1.16. BIC plot for STD/GBG group.

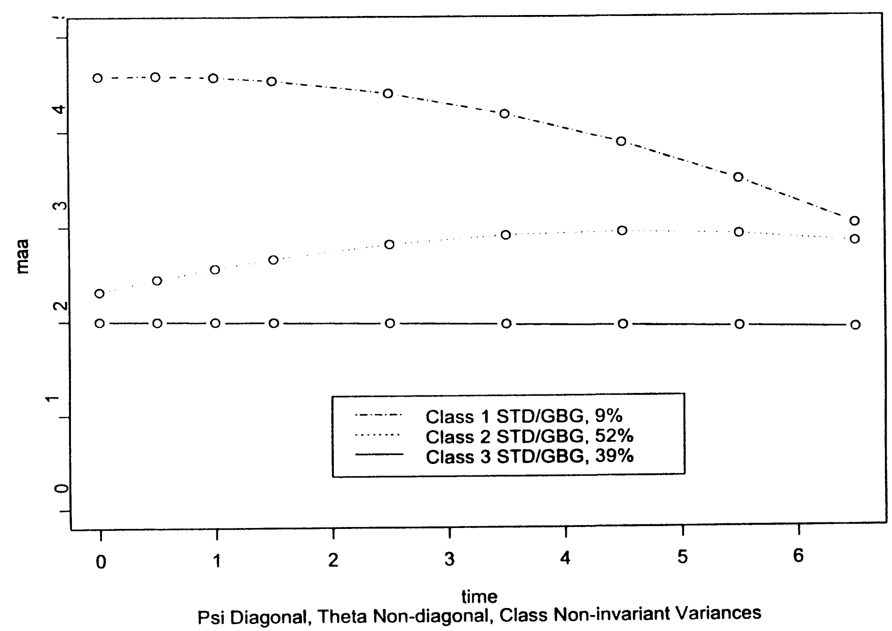


FIG. 1.17. Three-class exploration for STD/GBG.

Class 1 has the lowest probability, showing a high aggression level in early grades that decreases over time. Class 2 has the highest probability and shows a slightly increasing aggression trajectory. Class 3 consists of children showing very low, flat, and stable aggression trajectories. Figure 1.18 shows the quality of the classification using average posterior probabilities from the three-class model. The posterior-probability-weighted sample means and the estimated means for the outcomes are shown in Fig. 1.19.

	Class 1	*Class 2*	*Class 3*
Class 1	**0.860**	0.140	0.000
Class 2	0.052	**0.926**	0.022
Class 3	0.000	0.076	**0.924**

FIG. 1.18. Average posterior probabilites from the three-class model.

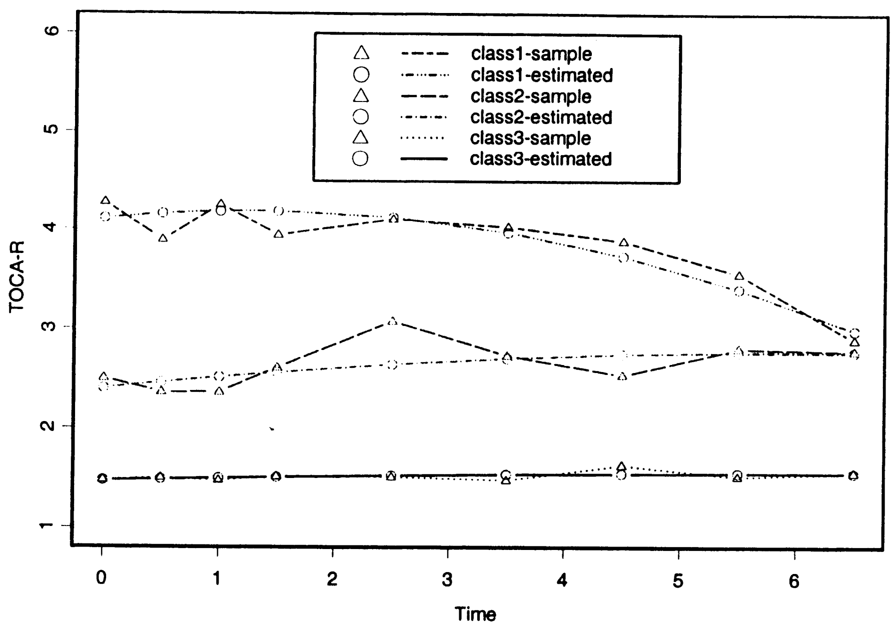


FIG. 1.19. Posterior-probability-weighted sample means and estimated means for aggression.

Growth Mixture Modeling With Antecedents and Consequences

Growth mixture modeling allows for class variation in how covariates influence the growth factors and also in how the growth classes influence variables other than the repeated measures. The former aspect is illustrated by the reading data and the latter aspect by the aggression data.

A Growth Mixture Model for Reading With a Covariate Predicting Class Membership. As discussed earlier, the growth mixture modeling produces posterior probabilities of class membership for each individual, and these can be used to classify individuals into their most likely class. Often the researcher wants to explore the profile of individuals in the different classes in terms of means of a set of background variables. This can be done using the individuals' classifications, but a more powerful analysis is to bring the background variables directly into the growth mixture analysis.

The five-class growth mixture model for word-recognition development is now expanded to include a predictor of class membership. A phonemic awareness measure taken at the end of kindergarten is used as a predictor. This variable is a proxy for some of the important prerequisites that a child needs to fully benefit from the instruction in Grade 1.

The modeling of the influence of phonemic awareness on class membership can be expressed as in multinomial logistic regression,

$$P(c_{ik} = 1 \mid A) = \frac{e^{\alpha_{c_k} + \gamma_{c_k} A_i}}{\sum_{k=1}^{K} e^{\alpha_{c_k} + \gamma_{c_k} A_i}}, \tag{21}$$

where $c_{ik}=1$ if Individual i belongs to Class k, A stands for phonemic awareness, $\alpha_{c_K} = 0$, $\gamma_{c_K} = 0$. Here γ_{c_k} $(k = 1, 2, \ldots, K-1)$ express the effect of phonemic awareness on the log odds of being in Class k versus Class K. For a two-class model, this is a regular logistic regression except that the dependent variable is latent.

The estimates of the extended five-class growth mixture model showed a similar picture for the trajectory class shapes and the class probabilities, indicating a desirable stability in the five-class model. The estimates of the multinomial logistic regression show that the probability of being in a high class increases as a function of increasing phonemic awareness value. A plot of the class probabilities as a function of phonemic awareness is given in Fig. 1.20.

A Growth Mixture Model for Aggression With a Distal Outcome Predicted by Class Membership. Many research questions related to growth mixtures concern the consequences of being in a certain trajectory class. For example, in the aggression data, it may be asked whether

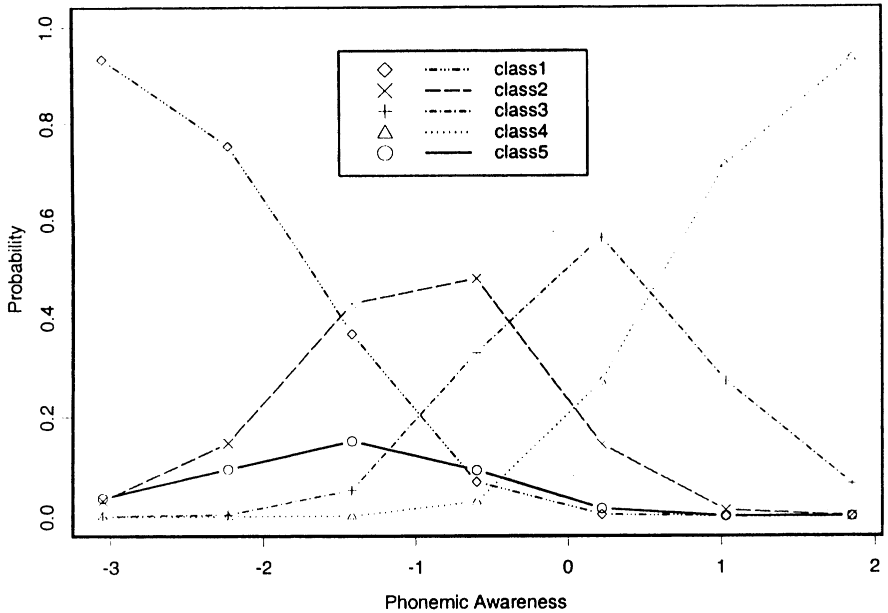


FIG. 1.20. Class probabilities as a function of phonemic awareness.

members of the high class have a higher risk for obtaining a juvenile court record. The juvenile court record variable is scored as u=1 versus u=0 for having a record before age 18 or not.

With a binary distal outcome, the class influence is described as the logit regression

$$logit\, P(u_i = 1 \mid c_{ik} = 1) = log[P(u_i = 1 \mid c_{ik} = 1)/P(u_i = 0 \mid c_{ik} = 1)] = \alpha_{u_k}. \quad (22)$$

Here α_{u_k} is the log odds for u_i=1 versus u_i=0 for individual i in Class k. An odds estimate and a corresponding confidence interval are obtained by exponentiating the α_k estimate and confidence limits.

The estimated growth mixture model for aggression and juvenile court record shows the same three classes as found earlier (see Fig. 1.8). The estimated odds for having a juvenile court record are 4.81 for Class 1, 0.61 for Class 2, and 0.49 for Class 3. A likelihood-ratio test of no differential effect of class on the juvenile court record probability did not, however, give a strong rejection [$\chi^2(2)=5.22$, $.05<p<.10$] perhaps due to low power associated with the low sample size.

COMPUTATIONAL ISSUES

Modeling with categorical latent variables presents the same potential computational problems as finite mixture analysis. Finite mixture analysis sometimes presents convergence problems and results in multiple maxima (see e.g., Titterington, Smith, & Makov, 1985). The degree of such complications is related to the information about the latent classes available in the data and the particular model applied. For example, growth mixture modeling of repeated measures data with clear trajectory classes may be less prone to such complications than a latent profile model for cross-sectional data. Models that allow for a larger degree of across-class variation in parameters are more likely to show such complications. Particularly sensitive are models with a large degree of across-class variation in variance–covariance parameters. For example, a latent profile model with across-variation in within-class variance is typically more difficult to fit than a latent profile model that only allows the means to vary across classes. In some cases, class-specific variances can lead to a small class with a singular covariance matrix giving an infinite likelihood value. The possibility of multiple maxima is well known in LCA, where random starting points are often used. For all models, the analyst is urged to search for multiple maxima to find the solution with the highest log likelihood value. In some cases, multiple maxima may be an indication of the need for more classes, as was observed in Muthén and Shedden (1999).

The identification status of a finite mixture model is difficult to assess, and general rules do not seem to be available in the literature. There is also a possible difference between theoretical and empirical nonidentification. Goodman (1974) observed theoretical nonidentification for a latent class model with four binary outcomes and three classes. Although this model has one parameter less than the unrestricted multinomial model, there is one indeterminacy among the parameters of the model that holds for any parameter values. Other models may be empirically nonidentified (i.e., in a region of the parameter space, the information matrix used to compute the estimated standard errors may be singular). In some cases, saddle points are found with a Hessian that has both positive and negative diagonal elements. These problems can possibly be avoided by using other starting values. It is recommended that mixture models be built up from relatively simple models, adding parameters stepwise and checking to which extent the log likelihood value improves.

CONCLUSIONS

This chapter has presented a series of latent variable models that introduce categorical latent variables in the form of clusters of individuals and in the form of latent trajectory classes. The models are special cases of a general latent variable modeling framework offering a unified view of seemingly disparate models. Many other models not discussed here fit into this framework, including noncompliance modeling in randomized trials (see Jo & Muthén, chap. 3, this volume), mixture cluster analysis, mixture factor analysis, and mixture structural equation modeling. The general modeling framework offers a large set of new analysis opportunities going far beyond the conventional SEM of the last few decades. This chapter is offered as a stimulus for further methodological developments and applications using this general framework.

APPENDIX: A GENERAL LATENT VARIABLE FORMULATION

This section gives the statistical specification of the general latent variable mixture model used in Mplus, drawing on Muthén, Shedden, and Spisic (1999). Related technical descriptions are presented in Muthén and Muthén (1998; Appendix 8) and Muthén and Shedden (1999). Applications are given in Muthén (in press), Muthén and Muthén (1999), and Muthén, Francis, Khoo, and Boscardin (in press).

GENERAL MODEL FORMULATION

The observed variables are x, y, and u, where x denotes a $q \times 1$ vector of covariates, y denotes a $p \times 1$ vector of continuous outcome variables, and u denotes an $r \times 1$ vector of binary and ordered polytomous categorical outcome variables. The latent variables are η denoting an $m \times 1$ vector of continuous variables and c denoting a latent categorical variable with K classes, $c_i = (c_{i1}, c_{i2}, \dots c_{iK})'$, where $c_{ik} = 1$ if Individual i belongs to Class k and zero otherwise.

The model relates c to x by multinomial logistic regression using the $K-1$-dimensional parameter vector of logit intercepts α_c and the $(K-1) \times q$ parameter matrix of logit slopes Γ_c, where for $k = 1, 2, \dots, K$

$$P(c_{ik} = 1 \mid \mathrm{x}_i) = \frac{e^{\alpha_{c_k} + \gamma'_{c_k} x_i}}{\sum_{k=1}^{k} e^{\alpha_{c_k} + \gamma'_{c_k} x_i}}, \tag{23}$$

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

The latent classes of c influence both y and u. Consider first the y part of the model. Conditional on Class k,

$$y_i = \nu_k + \Lambda_k \eta_i + \mathrm{K}_k \mathrm{x}_i + \epsilon_i, \tag{24}$$

$$\eta_i = \alpha_k + \mathrm{B}_k \eta_i + \Gamma_k \mathrm{x}_i + \zeta_i, \tag{25}$$

where the residual vector ε_i is $N(0, \Theta_k)$ and the residual vector ζ_i is $N(0, \Psi_k)$, both assumed to be uncorrelated with other variables. For u, conditional independence is assumed given c_i and x_i,

$$P(u_{i1}, u_{i2}, \dots, u_{ir} \mid c_i, \mathrm{x}_i) = P(u_{i1} \mid c_i, \mathrm{x}_i) P(u_{i2} \mid c_i, \mathrm{x}_i)\, P(u_{ir} \mid c_i, \mathrm{x}_i). \tag{26}$$

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

$$u_{ij} = s, \text{ if } \tau_{j,k,s} < u^*_{ij} < \tau_{j,k,s+1}, \tag{27}$$

$$P(u_{ij} = s \mid c_i, \mathrm{x}_i) = F_{s+1}(u^*_{ij}) - F_s(u^*_{ij}), \tag{28}$$

$$F_s(u^*) = \frac{1}{1 + e^{-(\tau_s - u^*)}}, \tag{29}$$

where for $u^*_i = (u^*_{i1}, u^*_{i2}, \dots, u^*_{ir})'$, $\eta_{ui} = (\eta_{u_{i1}}, \eta_{u_{i2}}, \dots, \eta_{u_{if}})'$, and conditional on Class k,

$$\mathrm{u}^*_i = \Lambda_{uk} \eta_{ui} + \mathrm{K}_{uk} \mathrm{x}_i, \tag{30}$$

$$\eta_{ui} = \alpha_{uk} + \Gamma_{uk} \mathrm{x}_i, \tag{31}$$

where Λ_{uk} is an $r \times f$ logit parameter matrix varying across the K classes, K_{uk} is an $r \times q$ logit parameter matrix varying across the K classes, α_{uk} is an $f \times 1$ vector logit parameter vector varying across the K classes, and Γ_{uk} is an $f \times q$ logit parameter matrix varying across the K classes. The thresholds may be stacked in the $(\Sigma^r_{j=1}(S_j - 1) \times 1)$ vectors τ_k varying across the K classes.

Mplus uses maximum-likelihood with the EM algorithm, viewing c_i as missing data. In the E step, the posterior probability of Individual i belonging to Class k is evaluated as

$$p_{ik} = P(c_{ik} = 1 \mid y_i, \mathrm{u}_i, \mathrm{x}_i) = P(c_{ik} = 1 \mid \mathrm{x}_i)[y_i \mid c_i, \mathrm{x}_i][\mathrm{u}_i \mid c_{ik} = 1]/[y_i, \mathrm{u}_i \mid \mathrm{x}_i]. \tag{32}$$

For certain individuals, prior or auxiliary information may restrict the admissible class membership to a subset of all the classes. This includes having individuals with known class membership. In this case, the posterior probabilities in Eq. (32) are renormed for each individual to add to one over the admissible set of classes. In Mplus, this is referred to as having training data.

The M step consists of three separate optimizations: for the y, x part using quasi-Newton; for the u, c, x part using Newton–Raphson and quasi-Newton; and for the c, x part using Newton–Raphson and quasi-Newton. Each M step need not be reaching an optimum, but often a few steps are sufficient.

In Mplus, missing data assuming MAR (Little & Rubin, 1987) is allowed for with respect to y and u.

ACKNOWLEDGEMENT

The research was supported by grant K02 AA 00230–01 from NIAAA and by grant 40859 from NIMH. The work has benefited from discussions in Hendricks Brown's Prevention Science Methodology group and Muthén's Research Apprenticeship Course. The e-mail address is bmuthen@ucla.edu.

REFERENCES

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.

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.

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

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

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

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

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.

Jackson, K. M., Sher, K. J., & Wood, P. K. (1999). Trajectories of conjoint substance use disorders: A developmental, typological approach to comorbidity. Forthcoming in *Alcoholism: Clinical and Experimental Research*.

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

Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Muthén, B. (1989). Dichotomous factor analysis of symptom data. In W. Eaton & G. Bohrnstedt (Eds.), *Latent variable models for dichotomous outcomes: Analysis of data from the epidemiological catchment area program* (pp. 19–65). A special issue of *Sociological Methods & Research, 18*, 19–65.

Muthén, B., & Muthén L. (1995). Tailoring Psychometric Techniques for Epidemiological and Clinical Applications. Technical report.

Muthén, B. (in press). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L. M. Collins, & A. Sayer (Eds.), *New methods for the analysis of change*. Washington, DC: American Psychological Assoction.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S., Yang, C., Wang, C. P., & Kellam, S. (2000). *General growth mixture modeling in randomized preventive interventions*. Unpublished manuscript, University of California, Los Angeles.

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. *Multilevel modeling: Methodological advances, issues, and applications*. In S. R. Reise & N. Duan (Eds.), NJ: Lawrence Erlbaum Associates.

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

Muthén, B., & Muthén, L. (1999). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882–891.

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

Muthén B., Shedden, K., & Spisic, D. (1999). General latent variable mixture modeling. Technical report.

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

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.

Schwartz G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Chichester, England: Wiley.