CHAPTER 1

LATENT VARIABLE HYBRIDS

Overview of Old and New Models

Bengt Muthén[1]
University of California, Los Angeles

LATENT VARIABLE HYBRIDS: OVERVIEW OF OLD AND NEW MODELS

The conference that this book builds upon contained many different special topics within the general area of modeling with categorical latent variables, also referred to as *mixture modeling*. The many different models addressed at that conference and within this volume may overwhelm a newcomer to the field. In fact, however, there are really only a small number of variations on a common theme. This chapter aims to distinguish the different themes, show how they relate to each other, and give some key references for further study. Some new mixture models are also proposed.

Table 1.1 gives a summary of different types of latent variable models. An overview discussion of the models of Table 1.1 was presented in Muthén (2002). The entries of the table are types of models, with the rows dividing the models into cross-sectional and longitudinal and the columns dividing models into traditional models with continuous latent variables, models

Advances in Latent Variable Mixture Models, pages 1–24

TABLE 1.1 Model Overview

	Continuous latent variables	Categorical latent variables	Hybrids
Cross-sectional models	Factor analysis, SEM	Regression mixture analysis, Latent class analysis	Factor mixture analysis
Longitudinal models	Growth analysis (random effects)	Latent transition analysis, Latent class growth analysis	Growth mixture analysis

with categorical latent variables, and newer hybrids using both types of latent variables. The upper left cell includes conventional psychometric models such as factor analysis (FA) and structural equation models (SEMs). The bottom left cell contains the generalization to longitudinal settings, where the continuous latent variables appear in the form of random effects describing individual differences in development over time. The categorical latent variable column includes cross-sectional models such as latent class analysis (LCA), which in longitudinal settings generalizes to latent transition analysis (LTA). LTA is a longitudinal model in the class of auto-regressive models (also including "hidden Markov" models), where the status at one time point influences the status at the next time point. Another LCA-related model is latent class growth analysis (LCGA), where the outcomes are influenced by growth factors analogous to conventional random effects growth modeling. The current chapter gives an overview that emphasizes the last column of hybrid models, with the typical examples of factor mixture analysis (FMA) and growth mixture modeling (GMM). As will be discussed, these models present useful generalizations of the models in the other columns, allowing for both classification of subjects in the form of latent classes and determination of continuous latent scores within these classes. All analyses to be discussed can be carried out using maximum-likelihood estimation in the Mplus program (Muthén & Muthén, 1998-2007).

Figure 1.1 gives a diagrammatic overview of hybrid latent variable models. The following sections will discuss the different branches of this diagram. A key distinction is made between models that specify measurement invariance and those that do not. In this case, invariance refers to measurement parameters being equal across the latent classes of the categorical latent variable(s). Measurement invariance with respect to observed groups such as gender is a well-known topic in psychometrics (see, e.g., Meredith, 1964, 1993). Simultaneous confirmatory factor analysis in several groups to study measurement invariance and group comparisons of latent variable distributions has been discussed in Jöreskog (1971) and Sörbom (1974). Measurement invariance is an important prerequisite for valid across-group comparisons of continuous latent variable constructs, giving a latent construct the same meaning and scale for proper comparisons across groups.

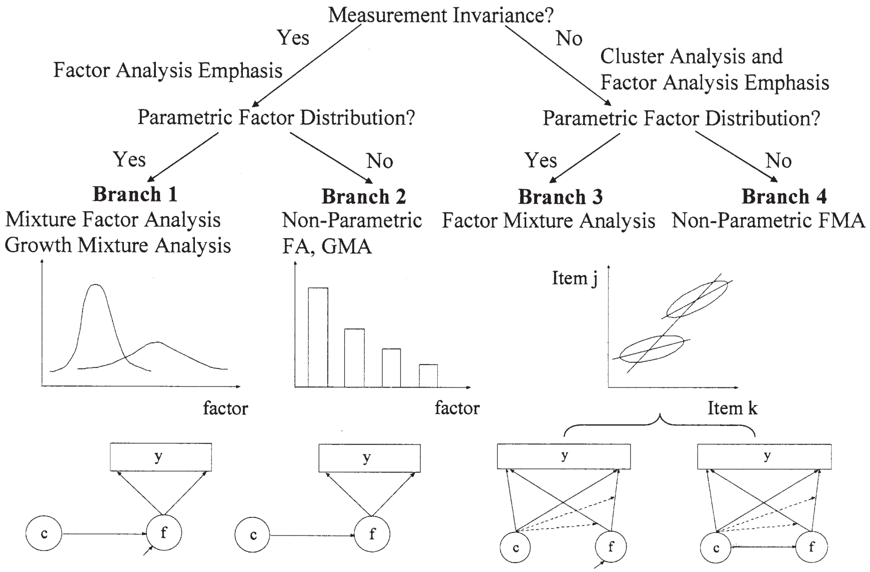

Figure 1.1 Overview of cross-sectional hybrids: Modeling with categorical and continuous latent variables.

Measurement invariance issues for latent groups are analogous and mixture analysis can be thought of as a multiple-group analysis, except with groups determined by the data and the model. With both continuous and categorical outcomes, the key factor analysis measurement parameters are the intercepts and the slopes in the regression of the outcome on the respective continuous latent variable. For example, with binary outcomes, a 2-parameter logistic regression model is typically used where in the item response theory (IRT) formulation the terms difficulty and discrimination are used for the two parameters, respectively. In IRT, measurement non-invariance is referred to as differential item functioning (DIF) and often focuses on the intercepts (difficulties) as in Rasch modeling.

As shown in the Figure 1.1 diagram, models with measurement invariance typically have a factor analysis (or IRT) focus. Here, the latent classes are used to describe heterogeneity among individuals in their continuous latent variable distributions. Separating heterogeneous classes of individuals is important when studying antecedents and consequences. For example, a covariate may have different influence on a factor for one class compared to another or a distal outcome may have different means or probabilities in different classes. As the diagram of Figure 1.1 shows, new branches are created by the choice of how to represent the continuous latent variable distribution. The typical approach is to make a parametric assumption such as normality as in the left-most branch, referred to as branch 1.

Branch 1: Hybrid Modeling With Measurement Invariance and Parametric Factor Distribution

The bottom of branch 1 displays two graphs. The top graph shows two factor distributions for two latent classes, differing in means and variances. The model diagram below denotes the categorical latent class variable as *c*, the continuous factor as *f*, and the observed items as *y*. Here, *c* influences *f* and *f* influences *y*. The regressions of the *y* items on *f* are either linear or non-linear (logit/probit) depending on the *y* scale. The regression of *f* on *c* is a linear regression. In line with dummy variables in linear regression, the different classes of *c* have different means for *f*. The short arrow pointing to *f* is a residual, indicating that *c* does not explain all the variation in *f*, but that there is also unaccounted for within-class variance. The measurement invariance specification is shown in the model diagram in that *c* neither influences *y*, nor changes the slopes in the regression of *y* on *f*.

Cross-Sectional Analysis

The following are some references to work in cross-sectional studies for branch 1, referring to the modeling as mixture factor analysis to emphasize the factor analysis aspects. Articles by McDonald (1967, 2003) on factor analysis represent pioneering work. Yung (1997) specifically studied measurement invariant mixture factor analysis and its maximum-likelihood (ML) estimation. Lubke and Muthén (2005) applied mixture factor analysis to continuous achievement data using ML via the Mplus program. Lubke and Muthén (2003) did Monte Carlo studies of how well mixture factor model parameters could be recovered under different degrees of factor mean separation across latent classes. It was found that it is more difficult to recover a mixture solution if only the factor means change across classes than if the measurement intercepts change as well. In other words, the measurement invariant hybrid can be more difficult to work with in practice.

Longitudinal Analysis

Turning to longitudinal examples, measurement invariant models are far more commonly used. In the branch 1 model diagram, the *y* box now represents repeated measures over several time points of a univariate *y* variable and *f* would correspond to intercept and slope growth factors (random effects). The growth factor means change over the latent classes, and thereby give rise to different trajectory shapes. In the SEM approach to growth modeling, the time points at which the *y* items are measured are captured by fixed factor loadings and zero *y* intercepts. Measurement invariance is natural because the time points are the same across the latent classes. Compared to the mixture factor analysis model, such growth mixture models appear more successful in recovering parameter values. Figure 1.2 shows a hypo-

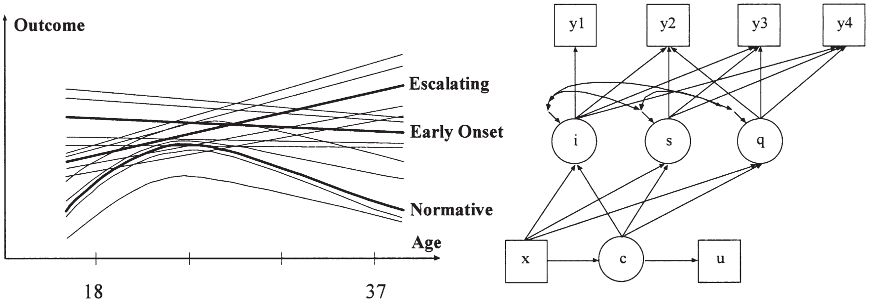

Figure 1.2 Growth mixture modeling of developmental pathways.

thetical example of three trajectory classes for an outcome studied over ages 18 to 37. In the graph on the left, the thick curves represent mean curves for each of the classes, while the thin curves represent individual curves within each class. The individual curves are variations on the curve shape themes represented by the mean curves. Defining the intercept growth factor *i* as the status at age 18, it is seen that the intercept *i*, the linear slope *s*, and the quadratic slope *q* have different means for the three classes. The model diagram of Figure 1.2 shows the mean differences of *i*, *s*, and *q* as arrows from *c* to *i*, *s*, and *q*.

Key references to growth mixture modeling include Verbeke and Lesaffre (1996) with applications to the development of prostate-specific antigen, Muthén and Shedden (1999) with application to the development of heavy drinking and alcohol dependence, Muthén et al. (2002) with application to intervention effects varying across trajectory classes for aggressive-disruptive behavior among school children, Lin, Turnbull, McCulloch, and Slate (2002) with application to prostate-specific antigen and prostate cancer, and Muthén (2004) with application to achievement development. Dolan, Schmittman, Lubke, and Neal (2005) modify the model to study regime (latent class) switching.

The Muthén (2004) analysis concerned mathematics achievement development in grades 7–10 in U.S. public schools. It was argued that poor development in this challenging topic was predictive of high school dropout, with antecedents of poor math development and dropout being found among variables capturing disengagement from school. In Figure 1.2 terms, *x* contains the antecedents and *u* is high school dropout. Figure 1.3 shows that the 20% classified as developing poorly in math have a drastically higher dropout rate than other students.

The Muthén (2004) analysis was carried out as two-level growth mixture modeling shown in diagram form in Figure 1.4. The top part of the figure labeled "Within" shows student variation, while the bottom part labeled "Between" shows variation across schools. This is a 3-level model with varia-

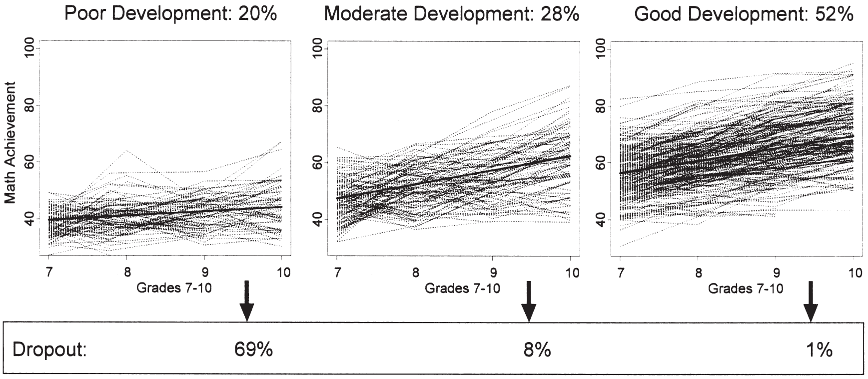

Figure 1.3 Growth mixture modeling: LSAY math achievement trajectory classes and the prediction of high school dropout.

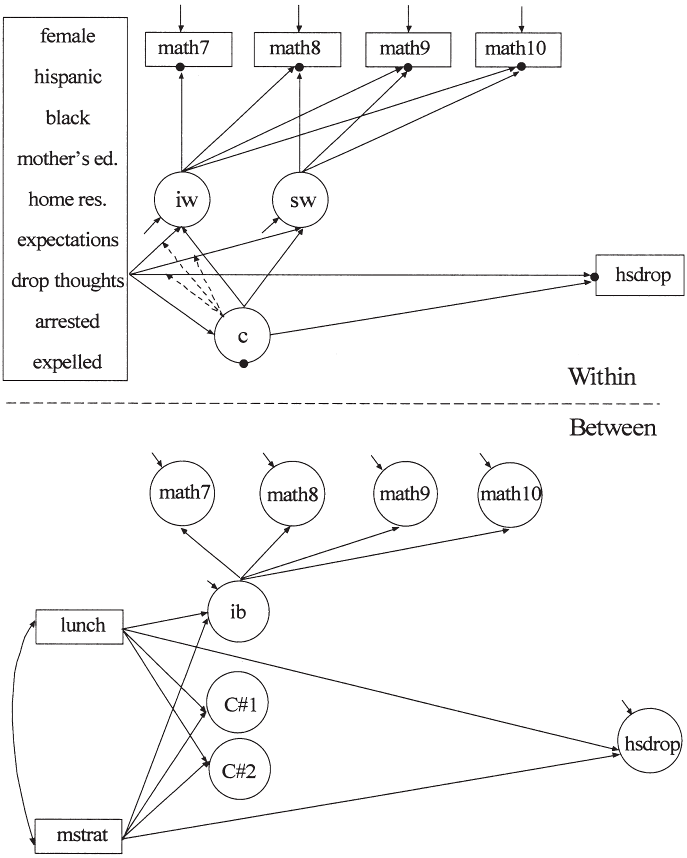

Figure 1.4 Two-level growth mixture modeling.

tion across time as level 1, variation across student as level 2, and variation across school as level 3. The figure shows the Mplus representation in "wide," multivariate form, transforming the model to two levels, Within and Between, where Within combines level 1 and level 2. The list of antecedents are shown in the box to the left in the Within part of the picture. Grade 7 measures of low schooling expectation and dropout thoughts were strongly related to both poor math development and dropout probability. The arrows from c to the intercept growth factor *iw* and the slope growth factor *sw* indicate that latent trajectory class membership influences the values of these growth factors. The arrow from *c* to high school dropout indicates that latent trajectory class membership influences the probability of this binary variable. The broken arrows from *c* to the arrows from the antecedents to the growth factors indicate that the antecedent influence varies across the latent trajectory classes. The math outcomes, the dropout outcome, and the latent class variable have filled circles attached to their boxes/circle indicating random intercepts, which vary across schools. On the Between level, these random intercepts are continuous latent variables. The math development on the Between level is captured by the random intercept *ib*, while the slope variance across schools is set to zero for simplicity. The latent class variable gives rise to two random intercepts, *c#1* and *c#2*, due to there being three classes. The regression of *ib* on the antecedents is a linear regression, the regression of *c#1*, *c#2* on the antecedents is a multinomial logistic regression, and the regression of *hsdrop* is a linear regression. Muthén (2004) found that a school-level covariate indicating quality of math teaching had a significant negative influence on being in the class of poor math achievement development and a positive influence on the within-class math achievement level. A school-level covariate indicating school neighborhood poverty had a positive influence on the probability of dropout.

Growth mixture modeling is also useful for outcomes with more complex distributions. The middle, left part of Figure 1.5 shows a commonly seen distribution of an outcome in longitudinal studies. A large portion of individuals is at the lowest point of the scale. A common reason is that these individuals at this time point have not yet started to engage in the activity studied. Examples of such outcomes include drinking and smoking among middle school students. A growth model approach that takes into account the large portion at zero was presented in Olsen and Schafer (2001). Figure 1.5 shows that the idea behind this modeling is to split the outcome in two parts. One part, labeled *u*, refers the binary outcome obtained by considering whether or not the individual engaged in the activity at the time point in question. The other part, labeled *y*, represents the amount of activity for those who engaged in the activity. At a time point where the person is not engaged in the activity, y is coded as missing. A parallel process growth

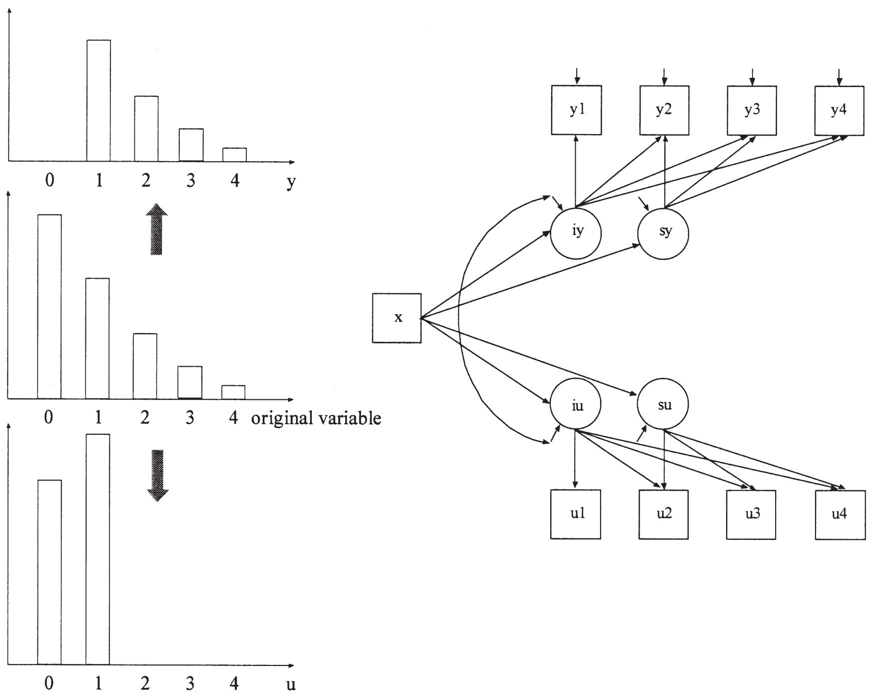

Figure 1.5 Two-part (semi-continuous) growth modeling.

model analyzes the two parts where the growth factors are correlated. The two parts may have different covariate influence.

The two-part growth model assumes that at a given time point individuals who just started to engage in the activity are at the same point in the growth process as individuals who started earlier. Individuals starting earlier may, however, be at a higher point in the growth process. To accommodate this, a mixture two-part model may be introduced as shown in Figure 1.6. A latent class variable *cu* influences the *u* part of the model, while the latent class variable *cy* influences the *y* part. In conclusion, it is clear that the models shown in Figures 1.4–1.6 can be seen as variations on the branch 1 theme of Figure 1.1.

Branch 2: Hybrid Modeling With Measurement Invariance and Non-Parametric Factor Distribution

In branch 2, the parametric latent variable distribution is replaced by a non-parametric approach using a flexible discretized representation of the distribution. This is illustrated at the bottom of branch 2 in the form of a

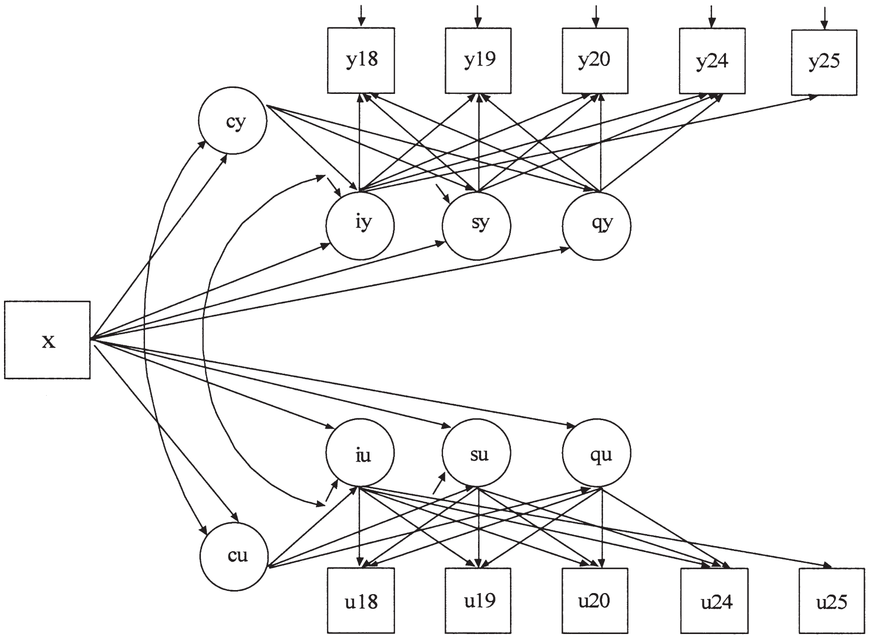

Figure 1.6 Two-part growth mixture modeling.

bar chart with four bars indicating a skewed distribution. The model diagram below the histogram shows that a mixture model can represent this where *c* influences *f*, but *f* has no within-class variability (there is no residual arrow pointing to *f*). Four classes of *c* results in four factor means for *f*. The positions of the four bars in the bottom graph represent scores on the latent variable distribution and are captured by the factor means in the four classes. The heights of the bars represent the class probabilities.

The relationship between the non-parametric approach and numerical integration is instructive and is illustrated in Figure 1.7 below. Numerical integration is necessary in maximum-likelihood estimation when a continuous latent variable has categorical indicators. With numerical integration, the latent variable distribution is also discretized, but the scores and the heights (called points and weights) are fixed, not estimated quantities. Figure 1.7 shows an example of a normal and a non-normal distribution, each with five points of support.

Cross-Sectional Analysis

Non-parametric estimation of latent variable distributions has both cross-sectional and longitudinal applications. In IRT applied to multiple-choice educational testing, Bock and Aitkin (1981) discussed the possibility of re-

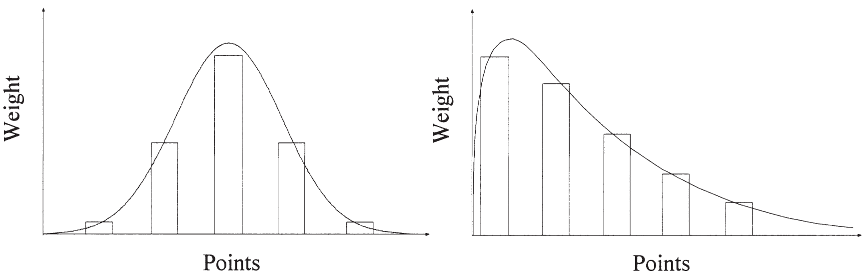

Figure 1.7 Non-parametric estimation of the random effect distribution using mixtures.

estimating the model with points and weights obtained from the posterior distribution, but suggested that this may not make much of a difference for the model parameters. This is also the experience of others (Robert J. Mislevy, personal communication), suggesting that the data commonly does not carry enough information about the particular form of the latent variable distribution. The normality assumption for the latent variable distribution may therefore be harmless in many applications, but perhaps not in cases where there is a strongly skewed or multimodal distribution. The non-parametric distribution shown in the histogram of branch 2 may be suitable for mental health applications where it is plausible that a large percentage of the population is unaffected. It is useful to try out such an alternative form and see if the likelihood improves to an important degree. An application to diagnostic criteria for alcohol dependence and abuse was studied in Muthén (2006), using the term latent class factor analysis.

Longitudinal Analysis

In longitudinal settings, Aitkin (1999) studied distributions of random effects in growth models, arguing that there it is hard empirically to find support for a normal distribution. He found that a few latent classes offered an adequate representation in several applications. Nagin and Land (1993) and Roeder, Lynch, and Nagin (1999) similarly argued for a non-parametric distribution of growth factors using the term "group-based" analysis, with application to groups of trajectories of criminal offenses. In some of Nagin's writings, however, the latent trajectory classes are given an interpretation as substantively meaningful subpopulations rather than seen as a mechanical way to non-parametrically represent a single population distribution (see Nagin, 2005). Figure 1.8 illustrates a unifying approach that does not seem to have been pursued by Nagin, namely using a combination of substantive and non-parametric classes. This example concerns analyses of the Cambridge data used in Nagin's research. Extending the analyses in Muthén, (2004), counts of biannual criminal convictions ages 11–21 scored

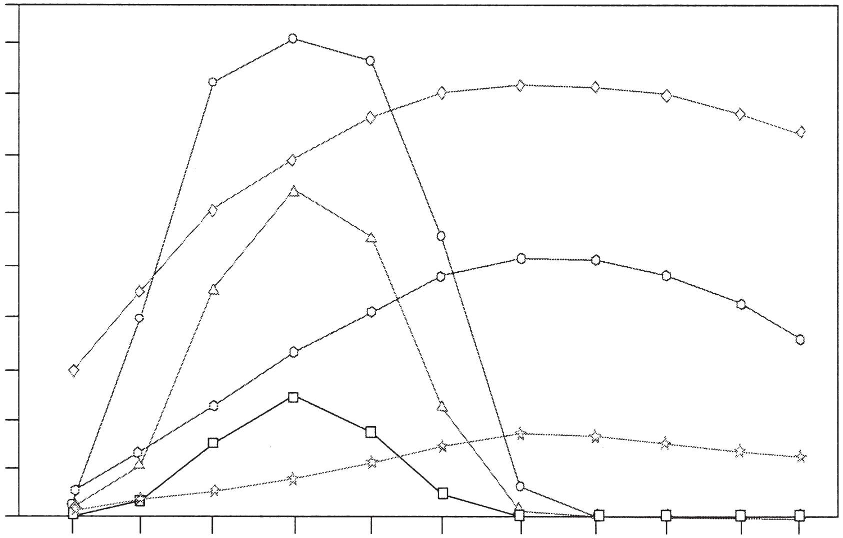

Figure 1.8 Three non-parametric classes within each of two substantive classes.

as 0, 1, and 2 for zero, one, or more convictions are analyzed using a quadratic growth model and six latent classes. The linear and quadratic growth factors were considered fixed, with zero variances. Figure 1.8 shows two substantively different latent classes of crime curves: early-peaking and late-peaking. Within each of these two classes, there are three variations: low, middle, and high. For each of the two substantive classes, the three variations are arrived at by using a three-class, non-parametric representation of the intercept growth factor in line with the branch 2 model diagram. Analogous to having the intercept factor random with a parametric intercept growth factor distribution, the linear and quadratic means were both held equal across the three non-parametric intercept classes. The non-parametric approach resulted in a skewed distribution for the intercept factor with more individuals in the low class as expected. An LCGA in line with Nagin's work would use six classes with no restrictions across classes on the linear and quadratic growth factor means.

Figure 1.9 compares three major approaches to growth modeling; hierachical linear modeling (HLM; see, e.g., Raudenbush & Bryk, 2002), growth mixture modeling (GMM; Muthén, 2004), and latent class growth analysis (LCGA; Nagin, 2005). LCGA and HLM are special cases of GMM. LCGA is a special case where there is no within-class variation so that the growth factor variances are all zero. In other words, there are no thin, individual curves in the graph implying that all individuals are the same within class. This in turn means that the within-class correlations across time are zero as

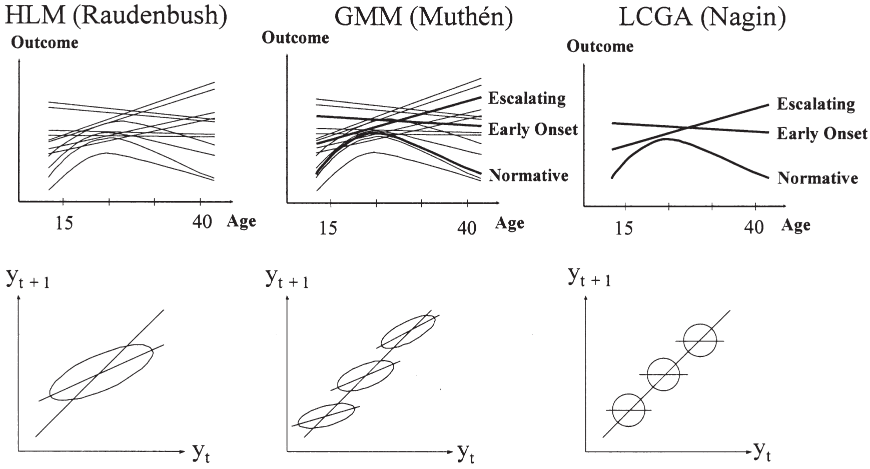

Figure 1.9 Growth modeling paradigms.

shown by the graph at the bottom right. HLM is a special case where there is only a single trajectory class. The thin, individual curves vary due to the growth factor variation. GMM allows more than one trajectory class as well as within-class variation, allowing non-zero within-class correlations across time. As mentioned above, GMM can be combined with a non-parametric representation of the growth factor distribution so that some latent classes have substantive meaning whereas others merely represent variation on the theme. A criminology application bridging the HLM, GMM, non-parametric GMM, and LCGA approaches is given in Kreuter and Muthén (2006).

Branch 3: Hybrid Modeling With Measurement Invariance and Parametric Factor Distribution

Going back up to the top of Figure 1.1, the first branching concerns measurement invariance or not. The non-invariant measurement branch often has a cluster analysis focus. Here, comparability of factor metrics across latent classes is not of importance, but the aim is to group individuals using a within-class model that is flexible. In some applications, however, the factor analysis focus is also present, although with no attempt to compare scores in different latent classes. The choice between a parametric and non-parametric latent variable distribution is available also here. The following discussion will focus on the parametric case of branch 3 given that there appears to be no literature on the non-parametric approach.

The bivariate graph of branches 3 and 4 shows two continuous items on the *x* and *y* axes, with the bivariate distribution of those two items displayed for two latent classes as two ellipses. Latent class analysis (LCA) specifies that the items are uncorrelated within each latent class, which means that the ellipses for the two classes in the figure would both have zero slopes. In contrast, the hybrid model of factor mixture analysis (FMA) allows the slopes to be estimated as non zero. The branch 3 model diagram shows how this within-class correlation is represented. The factor *f* represents unobserved heterogeneity among individuals and because *f* influences all items the items have non-zero correlation within each class. Consider for example the measurement of depression by a set of diagnostic criteria. Here, f may represent environmental influence such as stress in a person's life, whereas *c* may represent genetically determined categories of depression. Therefore, f influences all items (say, diagnostic criteria) to varying degrees, such that the measurement slopes vary across items. Furthermore, the variance of the factor and the measurement parameters may be different in different classes.

The branch 3 measurement non-invariance is indicated in the model diagram by the arrows from c to the items, representing intercept differences, and by the broken arrows from *c* to the arrows from *f* to the items, representing slope differences. Using the depression example, the normative group of individuals who are not depressed may have smaller or a different pattern of measurement slopes (loadings) across the items due to being less influenced by stress or because stress has a different meaning for such individuals. Note that the model diagram does not include an arrow from *c* to *f*. This implies that the factor means in all classes can be standardized to zero, instead representing mean/probability differences across classes for an item by a direct arrow from c to the item. In cross-sectional analysis, the model in the diagram is referred to as an FMA model. LCA is a special case of FMA where *f* is absent, in other words has zero variance, so that only the arrows from *c* to the items are present.

Cross-Sectional Analysis

There are many cross-section examples of branch 3, referred to here as FMA models. The earliest application appears to be Blafield (1980), studying factor mixture analysis applied to Fisher's Iris data. Measurement parameters of slopes (factor loadings) were allowed to vary across classes to improve the classification, while measurement intercepts, factor means, and covariance matrices were class-invariant. Yung (1977) studied factor mixture models where all parameters were allowed to be class-varying. An application to the classic Holzinger-Swineford mental ability data was presented, resulting in a "mean-shift" model with non-invariant intercepts, invariant loadings, and invariant factor covariance matrix (factor means fixed at zero in all classes for identification given the class-varying intercepts). The generalization of factor mixture modeling to structural equation model mixtures has been studied in market research, for example by Jedidi, Jagpal, and DeSarbo (1997) with an application to market segmentation and customer satisfaction. Factor mixture work for continuous outcomes has also developed outside psychometrics. McLachlan and Peel (2000) discuss factor analyzers where the within-class item correlations are described by an exploratory factor analysis (EFA) model. All measurement parameters are allowed to differ across the latent classes. The EFA model fixed the factor covariance matrix to an identity matrix (orthogonal factors) and let the residual variances vary across classes. McLachlan, Do, and Ambroise (2004) apply this model to microarray expression data, arguing that allowing for within-class correlation creates scientifically more meaningful clusters. In the most general case for continuous outcomes, FMA provides a within-class model with unstructured mean vector and covariance matrix, a commonly used model in finite mixture analysis. A classic example is the analysis of Fisher's Iris data as discussed in, for example, Everitt and Hand (1981).

A separate strand of factor mixture applications can be found in the IRT literature with a focus on categorical outcomes and applications to achievement testing. Mislevy and Verhelst (1990) used a mixture version of the 1-parameter Rasch model to classify individuals according to their solution strategies. Here, the measurement intercept (the "difficulty") varies across the latent classes, resulting in measurement non-invariance. Spatial visualization tasks can be solved by both rotational and by non-spatial analytic strategies, with item difficulties being higher for some items and lower for others depending on the latent class (strategy) the person belongs to. The authors also gave an example where the Rasch model holds for one latent class of individuals whereas the other class consists of those who guess at random. Mislevy and Wilson (1996) give an overview of mixture IRT models, including the Saltus model of Wilson (1989) distinguishing individuals with respect to different patterns of difficulties in line with theory of developmental psychology. For more recent work along the Saltus lines, see de Boeck, Wilson, and Acton (2005). The HYBRID model of Yamamoto (see Yamamoto & Gitomer, 1993) is a mixture model where an IRT model holds for one of the latent classes, whereas an LCA model model holds for other classes. Yamamoto and Gitomer apply this model to a test battery where several types of misunderstandings create item response patterns corresponding to latent classes. In the mixture IRT setting, measurement non-invariance is not a problem because the factor dimension of the different classes are recognized as different ability dimensions. Several chapters in this book describe further mixture IRT work.

Factor mixture analysis developments for categorical outcomes have also been made outside the IRT literature. Muthén (2006) and Muthén and Asparouhov (2006) considered dichotomous diagnostic criteria for substance use disorders, comparing LCA, FA/IRT, and FMA. FMA was chosen as the best model in both cases. Muthén, Asparouhov, and Rebollo (2006) applied FMA to alcohol criteria to provide latent variable phenotype modeling in a twin study of heritability.

Model testing is a challenging topic with mixture models in general and in particular with hybrid models. There are two reasons. First, it is difficult to test the model against data because no simple sufficient statistics such as mean vectors and covariance matrices exist. Second, comparing nested models, it is difficult to decide on the number of latent classes given that the regular likelihood ratio testing (LRT) does not give a chi-square test variable. For the second problem, Nylund, Asparouhov, and Muthén (2006) carried out a Monte Carlo simulation study of common indices such as the Bayesian Information Criterion (BIC), as well as the two newer approaches to LRT using non-chi-square distributions: Lo-Mendel-Rubin (LMR) and bootstrapped LRT (BLRT). The naïve LRT approach that incorrectly assumes a chi-square distribution was also studied (NCS). Table 1.2 shows how these four indices are able to pick the correct four classes for an LCA with 10 binary items. For each row, percentages are given for how frequently certain numbers of classes are chosen. It is seen that BIC tends to underestimate the number of classes, NCS tends to overestimate the number of classes, LMR falls in between, and BLRT does best. Research is needed on approaches for comparing models that differ not only in the number of classes, but also in the number of random effects (factors with non-zero variance).

To illustrate the previous points, the following is an FMA application in the area of diagnosing Attention Deficit Hyperactivity Disorder (ADHD). The analysis considers a UCLA clinical sample of 425 males ages 5–18. Subjects were assessed by clinicians through direct interview with the child (> 7 years) and through interview with mother about child using the KSADS instrument which has 9 inattentiveness items and nine hyperactivity items as shown in Table 1.3. The items were dichotomously scored. The research question concerned what types of ADHD are found in a treatment population. Table 1.4 shows model fitting results for three types of models: LCA, FA/IRT, and FMA. It is seen that the preferred LCA model is a 3-class model when

TABLE 1.2 Monte Carlo Simulation Excerpt from Nylund, Asparouhov, and Muthén (in press)

		BIC classes			NCS classes			LMR classes			BLRT classes		
	n	3	4	5	3	4	5	3	4	5	3	4	5
10-item (complex structure) with 4 latent classes	200	92	8	0	2	48	41	34	43	9	16	78	6
	500	24	76	0	0	34	45	9	72	14	0	94	6
	1000	0	100	0	0	26	41	2	80	17	0	94	6

TABLE 1.3 The Latent Structure of ADHD

Inattentiveness items	Hyperactivity items
Difficulty sustaining attention on task/play	Difficulty remaining seated
Easily distracted	Fidgets
Makes a lot of careless mistakes	Runs or climbs excessively
Doesn't listen	Difficulty playing quietly
Difficulty following instructions	Blurts out answers
Difficulty organizing tasks	Difficulty waiting turn
Dislikes/avoids tasks	Interrupts or intrudes
Loses things	Talks excessively
Forgetful in daily activities	Driven by motor

TABLE 1.4 The Latent Structure of ADHD: Model Fit Results

Model	Log Likelihood	# parameters	BIC	BLRT *p* value for k–1
LCA—2c	–3650	37	7523	0.00
LCA—3c	**–3545**	**56**	**7430**	**0.00**
LCA—4c	–3499	75	7452	0.00
LCA—5c	–3464	94	7496	0.00
LCA—6c	**–3431**	**113**	**7547**	**0.00**
LCA—7c	–3413	132	7625	0.27
EFA—2f	**–3505**	**53**	**7331**	
FMA—2c, 2f	–3461	59	7280	
FMA—2c, 2f Class-varying factor loadings	**–3432**	**75**	**7318**	**χ^2-diff (16) = 58 $p < 0.01$**

judged by BIC, but is a 6-class model when judged by BLRT. The item profile plots corresponding to these two models are shown at the top of Figure 1.10. The items are arranged along the x axis with the nine inattentiveness items first, followed by the nine hyperactivity items. The 3-class model suggests a combined class, an inattentiveness only class, and a weakly defined hyperactivity only class. The 6-class model appears to show several variations on these three themes and is suggestive of a more dimensional representation. As seen in Table 1.4, an exploratory factor analysis (EFA) is a strong alternative to the LCA models. Here, EFA is the same as a two-dimensional IRT model, using 2-parameter logistic item characteristic curves. EFA has a better log-likelihood than the 3-class LCA for fewer parameters and a considerably better BIC than the 6-class model. Given these results, FMA is an

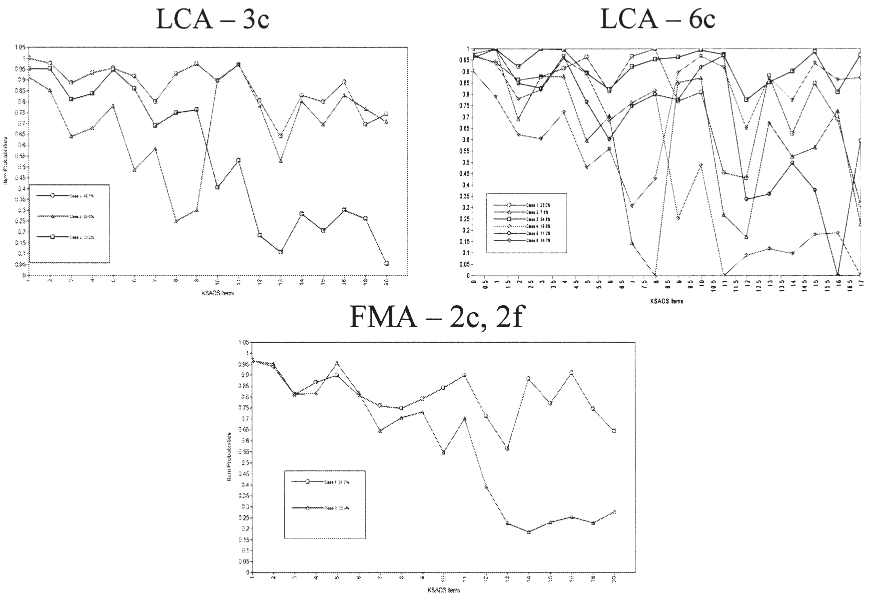

Figure 1.10 Item profiles for three-class LCA, six-class LCA, and two-class, two-factor FMA.

interesting alternative. As seen in Table 1.4, a 2-class FMA with 2 factors (one for inattentiveness and one for hyperactivity) has as good of a log-likelihood as the 6-class LCA but with far fewer parameters, and has a better BIC value than the EFA. Figure 1.11 shows that the hyperactivity only class disappears in the FMA model. The plot shows the mean probability of item endorsement, but it should be noted that variations in the item probabilities are produced within both classes as a function of the factor values.

Longitudinal Analysis

Longitudinal examples in branch 3 do not appear to have been published. Two different types of approaches can be considered. One model type is based on growth modeling where random effects influence an item measured at several time points. As seen when comparing the model diagrams in branches 1 and 3, this is different from the growth mixture modeling discussed in branch 1. The branch 3 model diagram shows that the latent class variable influences the items directly. Without the factor, this is an LCA model where the time structure is ignored and T repeated measures of the item is seen as T different items. This is a useful first analysis before turning to the branch 2 latent class growth analysis (LCGA) where growth factors govern the change over time in items means/probabilities. The LCA can be used to explore growth shapes in the data without impos-

Transition Probabilities

Mover Class (c=1)

		c2 = 1	c2 = 2
c1	1	0.6	0.4
	2	0.3	0.7

Stayer Class (c=2)

		c2 = 1	c2 = 2
c1	1	0.90	0.10
	2	0.05	0.95

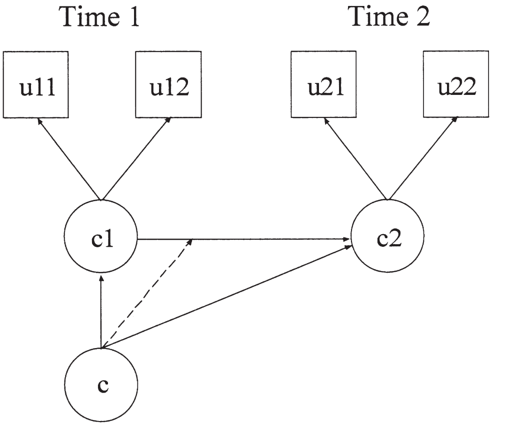

Figure 1.11 Latent transition analysis.

ing a particular growth function. As shown in the Figure 1.1 model diagram for branch 3, a factor may also be included to account for within-class correlations across time, allowing for a more flexible model in line with FMA. This factor does not have to be a growth factor where time determines the factor loadings, but a single-factor model with free factor loadings could for example be used.

Another longitudinal model type is based on auto-regressive modeling. As an example of the auto-regressive model type, latent transition analysis (LTA) considers several items measured at each of several time points to capture changes in a latent class variable. The latent class variable at one time point influences the latent class variable at the next time point in an auto-regressive fashion. Conventional LTA does not include a factor (factor variance is zero). A hypothetical example with two items measured at two time points is shown in Figure 1.11. LTA is an auto-regressive model in the sense that the time 2 status of the latent class variable c2 is dependent on the time 1 status of the latent class variable c1. Top left of Figure 1.11 is a hypothetical transition probability table (see "Mover Class"). For examples, individuals starting in the c1 = 1 class have the probability 0.4 to transition to the c2 = 2 class. The probabilities in each row of the table sum to one. The bottom table for the "Stayer Class" shows smaller probabilities for transitioning between classes. Conventional LTA does not include the latent class variable c at the bottom of the model diagram. Including this additional latent class variable makes it possible to distinguish between the latent classes of Movers and Stayers (see, e.g., Langeheine & van de Pol, 2002; Mooijaart, 1998).

Figure 1.12 shows a two-level extension of LTA by Asparouhov and Muthén (2007) in this volume. The application concerns aggressive-disruptive be-

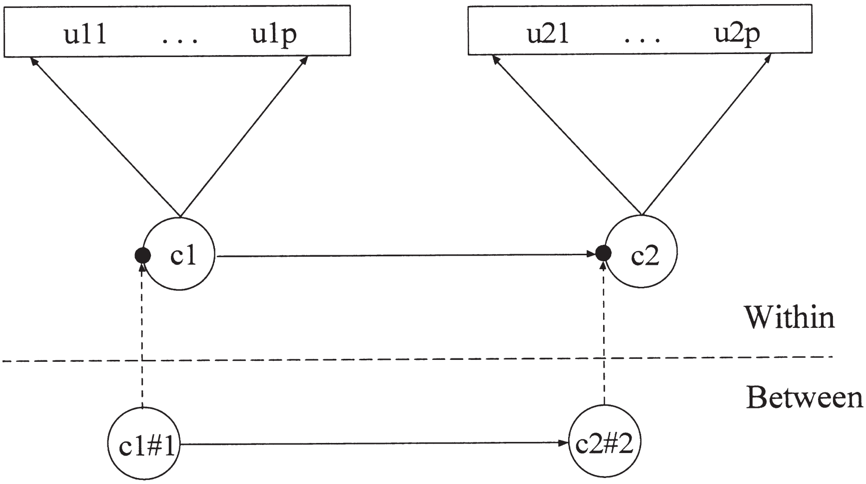

Figure 1.12 Two-level latent transition analysis.

havior in the classroom in Fall and Spring of first grade in Baltimore public schools (see also growth mixture modeling of these data in Muthén et al., 2002). The top part of the model diagram describes the within-level part of the model where the variables vary across students. The bottom part of the model diagram describes the between-level part of the model with variation across classrooms. The within-level shows filled circles next to the latent class variables c1 and c2, representing random intercepts. For example, the filled circle for c2 is the random intercept in the multinomial logistic regression of c2 on c1. On the between level, these random intercepts are continuous latent variables, shown as c1#1 and c2#1, representing the amount of classroom-level aggressive-disruptive behavior. On the between level, these two variables are connected via linear regression. Asparouhov and Muthén (2007) show that the classroom variation is large at both time points. The Fall between-classroom effect has a large impact on students' aggressive-disruptive behavior in the Fall. However, the effect also carries over into Spring, both through the individual level and through the classroom level.

A new model, which is a generalized, hybrid latent transition model will now be presented. This model includes f as shown in the Figure 1.1 model diagram of branch 3. A hypothetical example with p items at two time points is shown in Figure 1.13. At each time point, an FMA measurement model is specified with c having direct effects on the u's, and f describing continuous heterogeneity among individuals that reflects within-class correlation among the items. The latent class variable c2 is influenced by c1, but is also potentially influenced by f1. As shown by the vertical bars for the two latent classes in the bottom graph, individuals who at time 1 are low in the high

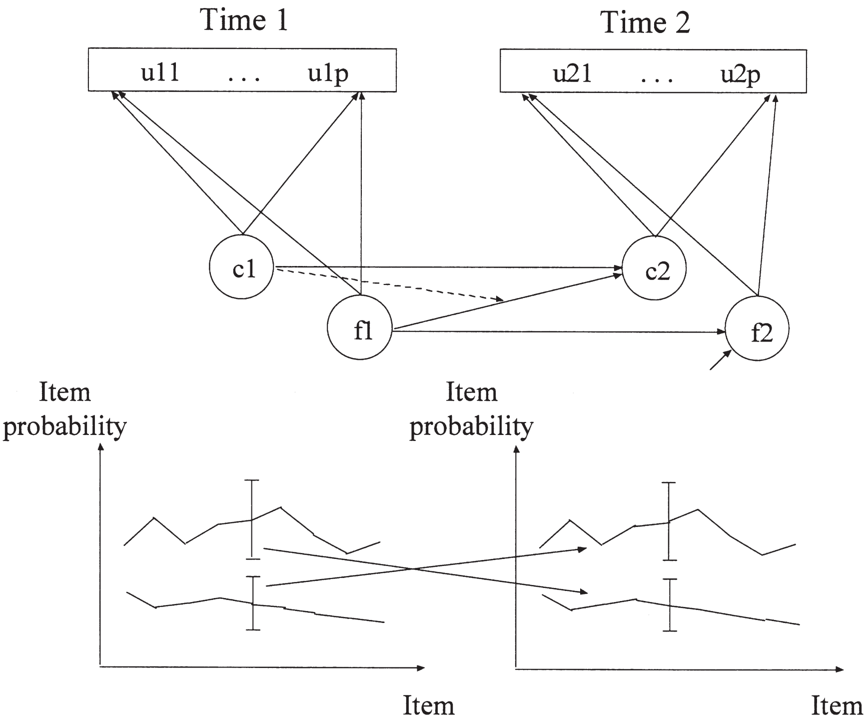

Figure 1.13 Factor mixture latent transition analysis.

class may be more likely to be in the low class at time 2 than individuals who are high in the high class. Similarly, individuals who are high in the low class may be more likely to be in the high class at time 2 than individuals who are low in the low class.

The FMA-LTA was applied to the data on aggressive-disruptive behavior in the classroom in Fall and Spring of first grade in Baltimore public schools referred to earlier. A model with two latent classes (high and low aggressive-disruptive behavior) and one factor dimension was found suitable for each of the two occasions. Table 1.5 shows the results of fitting the FMA-LTA versus the conventional LTA. It is seen that the log-likelihood is considerably better for FMA-LTA. Although this comes at the expense of 19 more parameters, the log-likelihood is so much better that this is more than compensated for. This advantage is reflected in the considerably better BIC value for the FMA-LTA model.

Table 1.6 shows the resulting estimates of the transition probability tables for the two model alternatives. The conventional LTA shows low probabili-

TABLE 1.5 Factor Mixture Latent Transition Analysis: Aggressive-Disruptive Behavior in the Classroom

Model	LogLikelihood	# parameters	BIC
Conventional LTA	–8,649	21	17,445
FMA LTA factors related across time	–8,012	40	16,306

TABLE 1.6 Estimated Latent Transition Probabilities, Fall to Spring

	Low	High
Conventional LTA		
Low	0.93	0.07
High	0.17	0.83
FMA-LTA		
Low	0.94	0.06
High	0.41	0.59

ties for transitioning from one class in Fall to a different class in Spring. In contrast, the FMA-LTA shows that there is a rather high probability (0.41) of transitioning from the high-aggressive class in Fall to the low-aggressive class in Spring.

CONCLUSION

This discussion has attempted to bring together seemingly disparate hybrid latent variable modeling efforts in many different application areas. The aim was to show that the various models are only slight variations on a few key themes. The critical aspects of the models are whether or not they specify measurement invariance and whether or not a parametric latent variable distribution is specified. By clearly showing the connections between different modeling branches and types of applications, researchers may be enabled to more easily learn from analysis experiences in neighboring fields. It is clear that much more methodological research is needed in this emerging research topic of mixture modeling and hopefully this chapter will stimulate such developments.

NOTE

1. This research was supported under grant K02 AA 00230 from NIAAA.

REFERENCES

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, *55*, 117–128.

Asparouhov, T., & Muthén, B. (2007). Multilevel mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models*. Charlotte, NC: Information Age Publishing, Inc.

Blafield, E. (1980). *Clustering of observations from finite mixtures with structural information*. Unpublished doctoral dissertation, Jyvaskyla studies in computer science, economics, and statistics, Jyvasykla University, Jyvaskyla, Finland.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459.

de Boeck, P., Wilson, M., & Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. *Psychological Review*, *112*, 129–158.

Dolan, C. V., Schmittmann, V. D., Lubke, G. H., & Neale, M. C. (2005). Regime switching in the latent growth curve mixture model. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*, 94–119.

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

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.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426.

Kreuter, F., & Muthén, B. (2006). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. Manuscript submitted for publication.

Langeheine, R., & van de Pol, F. (2002). Latent Markov chains. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 304–341). Cambridge, UK: Cambridge University Press.

Lin, H., Turnbull, B. W., McCulloch, C. E., & Slate, E. (2002). Latent class models for joint analysis of longitudinal biomarker and event process data: Application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association*, *97*, 53–65.

Lubke, G., & Muthén, B. (2003). Performance of factor mixture models. Manuscript submitted for publication.

McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometric Monograph No. 15*.

McDonald, R. P. (2003). A review of Multivariate taxometric procedures: Distinguishing types from continua. *Journal of Educational and Behavioral Statistics*, *28*, 77–81.

McLachlan, G. J., Do, K-A., & Ambroise, C. (2004). *Analyzing microarray gene expression data.* New York: Wiley.

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

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, *19*, 187–206.

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

Mislevy, R. J., & Verhulst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*, 195–215.

Mislevy, R. J., & Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika*, *61*, 41–71.

Mooijaart, A. (1998). Log-linear and Markov modeling of categorical longitudinal data. In C. C. J. H. Bijleveld & T. van der Kamp (Eds.), *Longitudinal data analysis: Designs, models, and methods.* Newbury Park, CA: Sage Publications.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*, 81–117.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage Publications.

Muthén, B. (2006). Should substance use disorders be considered as categorical or dimensional? *Addiction*, *101* (Suppl. 1), 6–16.

Muthén, B., & Asparouhov. T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors*, *31*, 1050–1066.

Muthén, B., Asparouhov, T., & Rebollo, I. (2006). Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. *Twin Research and Human Genetics*, *9*, 313–324.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., et al. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, *3*, 459–475.

Muthén, L. K., & Muthén, B. (1998–2006). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.

Nagin, D. S. (2005). *Group-based modeling of development.* Cambridge, MA: Harvard University Press.

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.

Nylund, K. L., Asparouhov, T., & Muthén, B. (in press). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal.*

Olsen, M. K., & Schafer, J. L. (2001). A two-part random effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association*, *96*, 730–745.

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

Roeder, K., Lynch, K. G., & Nagin, D. S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association*, *94*, 766–776.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229–239.

Verbeke, G., & Lesaffre, E. (1996) A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association, 91*, 217–221.

Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin, 105*, 276–289.

Yamamoto, K., & Gitomer, D. H. (1993). Application of a HYBRID model to a test of cognitive skill representation. In N. Fredriksen, R. Mislevy, & I. Beijar (Eds.), *Test theory for a new generation of tests.* Hillsdale, NJ: Erlbaum.

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