

Understanding Linkages Among Mixture Models

Sonya K. Sterba
Vanderbilt University

The methodological literature on mixture modeling has rapidly expanded in the past 15 years, and mixture models are increasingly applied in practice. Nonetheless, this literature has historically been diffuse, with different notations, motivations, and parameterizations making mixture models appear disconnected. This pedagogical review facilitates an integrative understanding of mixture models. First, 5 prototypic mixture models are presented in a unified format with incremental complexity while highlighting their mutual reliance on familiar probability laws, common assumptions, and shared aspects of interpretation. Second, 2 recent extensions—hybrid mixtures and parallel-process mixtures—are discussed. Both relax a key assumption of classic mixture models but do so in different ways. Similarities in construction and interpretation among hybrid mixtures and among parallel-process mixtures are emphasized. Third, the combination of both extensions is motivated and illustrated by means of an example on oppositional defiant and depressive symptoms. By clarifying how existing mixture models relate and can be combined, this article bridges past and current developments and provides a foundation for understanding new developments.

Over the past 15 years, the number of finite mixture modeling applications has dramatically increased in the social sciences (Collins & Lanza, 2010; Muthén & Muthén, 2000; Nagin & Odgers, 2010). Mixture models are now more commonly used for accommodating discrete population heterogeneity than are non-model-based classification methods such as cluster analysis and taxometrics (for reasons discussed in Lubke & Tueller, 2010; McLachlan, 2011; Vermunt,

Correspondence concerning this article should be addressed to Sonya K. Sterba, Quantitative Methods Program, Department of Psychology and Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203. E-mail: Sonya.Sterba@Vanderbilt.edu

2011). When mixture modeling is viewed as a unified methodological approach, many commonalities can be noted. Each mixture model can be broken down into a within-class model and a between-class model. The *within-class model* defines a data-generating mechanism for persons in class k . The *between-class model* defines how likely it is that a person will be a member of one class versus another. Combining these yields a *combined model*—a marginal (across-class) density function. Further, the construction of mixture models can be seen to rely on the same probability laws. These models also share assumptions, and they allow the same approach for obtaining predictions of class assignment given observed data. Mixture models also all allow “direct” and “indirect” interpretations of latent classes. Under the “direct” interpretation, extracted classes are hypothesized to correspond with unobserved population subgroups (e.g., Mazza, Fleming, Abbott, Haggerty, & Catalano, 2010). Under an “indirect” interpretation, classes can be used to discretely approximate unknown distributional forms even without the assumption that population subgroups exist (e.g., Bauer, 2005; Nagin, 2005).

An integrative understanding of shared aspects of mixture models is beneficial in highlighting similarities in how key assumptions can be relaxed and in suggesting how these models can be extended or combined in new ways. However, currently it remains difficult for psychology researchers to obtain an integrative understanding of mixture models. Historically, one reason has been the different parameterizations¹ and notational conventions used in mixture literatures for continuous versus categorical outcomes (Heinen, 1996; McCutcheon, 2002; McLachlan & Peel, 2000). Three other reasons for this difficulty, described here, jointly have not been thoroughly addressed.

First, reviews commonly focus on a pair of closely related mixture models rather than on a broader spectrum. For instance, reviews of such pairs of mixture models have included latent class versus latent profile models (e.g., McDonald, 1962; Wolfe, 1970), latent class versus latent transition models (e.g., Collins & Lanza, 2010), and groups-based trajectory versus growth mixture models (e.g., Kreuter & Muthén, 2008; Muthén, 2004; Pickles & Croudace, 2010). The first goal of this article is to provide more inclusive coverage of prototypic mixture models, employing a cumulative model-building approach incrementing from

¹For mixture models with discrete outcomes, a probability parameterization, most commonly, or loglinear parameterization, have historically been used (e.g., Biemer, 2011; Collins & Flaherty, 2002; Heinen, 1996; McCutcheon, 2002). For mixture models with continuous outcomes, a logistic parameterization is typically used for the between-class model, as employed here also. For consistency, we also employ a logistic parameterization with discrete outcomes (following, e.g., Humphreys & Janson, 2000; B. O. Muthén, 2001, 2004; Reboussin, Reboussin, Liang, & Anthony, 1998). This logistic parameterization can be used to compute probabilities, implicitly achieves the same constraints as in the popular probability parameterization, and readily expands to accommodate covariates (unlike the probability parameterization; Magidson & Vermunt, 2004).

the simplest univariate mixture model, to some of the most complex recent extensions, under unified notation.

Second, sources that do provide general coverage of mixture models have had somewhat different presentation formats in the literature aimed at statisticians versus psychologists. General presentations in the statistics literature have placed more emphasis on the combined model (marginal density), whereas within-class data models may not be presented for every mixture model (e.g., McLachlan & Peel, 2000, Chapter 1; Schlattmann, 2009, Chapter 1; Vermunt, 2008). This format is more useful for highlighting the “indirect” approximation function of mixtures and for highlighting how assumptions are imposed. General presentations aimed at psychologists place more emphasis on the within-class data models whereas a marginal density may not be presented for every mixture model (e.g., Muthén, 2002); alternately, path diagrams may be presented (e.g., Sterba & Bauer, 2010). This format is more helpful for understanding the “direct” interpretation of mixtures; additionally, interpreting the path diagrams requires an equations-based familiarity with the model architecture.² This article represents each mixture model by a within-class data model and density, between-class model, and marginal density in order to facilitate its second goal. The second goal is to highlight mixture models’ similar construction through their reliance on shared probability laws.

Third, relationships among two recent extensions of mixture models have not been fully emphasized. However, these two extensions have a common objective: relaxing a key “local independence” assumption (defined later). One extension, *hybrid mixtures*, accomplishes this objective by allowing for systematic individual differences in the within-class model. Hybrid mixtures include, for instance, factor mixtures, item response theory (IRT) mixtures, growth mixtures, and latent transition analysis with a factor measurement model. Another extension, *parallel-process mixtures*, accomplishes this objective via expansion of the between-class model. Parallel-process mixtures include, for instance, joint groups-based trajectory models and associative latent transition models. The third goal of this article is to pedagogically relate models implementing each extension under shared assumptions and shared construction based on probability laws. With this foundation, we consider how and why both extensions can be employed at once in an empirical illustration.

²Reasons are that (a) there is no formal set of rules (akin to Wright’s tracing rules in SEM) allowing mixture model equations to be directly reproduced from their diagrams and (b) diagrams do not fully represent all aspects of the mixture model. For instance, current path diagrams do not communicate how many classes there are, if a particular parameter differs across some but not all classes, if a parameter is fixed to 0 in some classes but estimated in others, or which is the reference class.

The remainder of this article proceeds as follows: First, five prototypical mixture models are presented: the univariate finite normal mixture, latent profile analysis, latent class analysis, groups-based trajectory, and latent transition analysis models. The presentation of each mixture model intentionally takes a similar form—a within-class, between-class, and combined model—followed by description of calculating posterior probabilities of class membership. The shared construction across mixture models is explicitly related to the following four probability laws, here stated for discrete variables³ (each law has a continuous variable analogue).

Law of total probability:

$$P(A) = P(B_1)P(A|B_1) + \dots + P(B_k)P(A|B_k) \quad (1)$$

Multiplication rule for conditionally independent events:

$$P(A, C|B) = P(A|B)P(C|B) \quad (2)$$

Multiplication rule for dependent events:⁴

$$P(A, B, C, D) = P(A)P(B|A)P(C|A, B)P(D|A, B, C) \quad (3)$$

Bayes' rule: $P(B|A) = P(B)P(A|B)/P(A)$. (4)

After the classic mixture models are presented, two kinds of recent extensions—hybrid and parallel-process mixtures—are related and examples of each kind are reviewed. Finally, a novel application illustrates how an integrative perspective facilitates flexibly combining these two extensions to meet particular modeling challenges. Syntax (in Mplus; L. K. Muthén & Muthén, 1998–2013) for model fitting is deferred to an online appendix,⁵ as are path diagrams for all but the final empirical example model.

UNIVARIATE FINITE NORMAL MIXTURE MODEL (UFNM)

The simplest mixture model we consider, a univariate finite normal mixture (UFNM; e.g., Pearson, 1894), requires one observed, continuous variable y_i measured cross-sectionally in a sample of $i = 1 \dots N$ persons. Suppose there is a latent classification variable, c_i , which can take one of $k = 1 \dots K$ values for person i . K is the finite, total number of classes in an analysis. K is determined

³Here we use a shorthand: $P(A)$ represents $P(A = a)$, where a is a realization of random variable A .

⁴This is also sometimes called the chain rule.

⁵Online appendix is available at <http://www.vanderbilt.edu/peabody/sterba/appxs.htm>

via comparing the fit of competing models with different numbers of classes using model selection indices, for instance, the Bayesian information criterion (BIC; see Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2007).

Within-Class Model

For person i in class k , that is, where $c_i = k$, we have the following data model:

$$y_i = \mu^{(k)} + \varepsilon_i \quad (5)$$

$$\varepsilon_i \sim N(0, \sigma^{2(k)}). \quad (6)$$

In class k , $\mu^{(k)}$ is the model-implied mean, and $\sigma^{2(k)}$ is the model-implied variance of the normally distributed person-specific residual ε_i . A parameter's k superscript denotes that its value is allowed to differ across classes. The UFNM assumes that, within each class, the outcome is univariate normally distributed with probability density function (PDF)

$$f(y_i | c_i = k) = \frac{1}{\sigma^{(k)} \sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu^{(k)})^2}{2\sigma^{2(k)}}\right). \quad (7)$$

Between-Class Model

The model-implied probability that person i will be a member of class k (also called the class probability) can be obtained via a multinomial logistic parameterization

$$p(c_i = k) = \exp(\omega^{(k)}) / \sum_{k=1}^K \exp(\omega^{(k)}). \quad (8)$$

$\omega^{(k)}$ is a multinomial intercept, and for identification purposes (to avoid redundancy), $\omega^{(K)} = 0$.

Combined Model

Applying the *law of total probability* in Equation (1), the marginal density of y_i in the population, denoted $f(y_i)$, can be obtained as a weighted sum of within-class densities:

$$f(y_i) = \sum_{k=1}^K p(c_i = k) f(y_i | c_i = k). \quad (9)$$

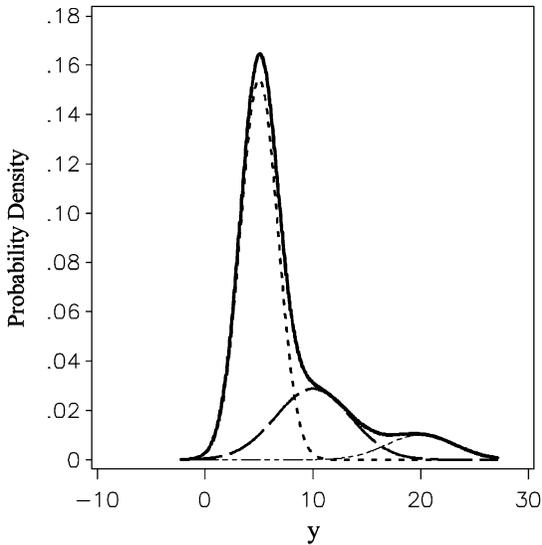


FIGURE 1 Illustrative univariate finite normal mixture density where $K = 3$. *Note.* Bold solid line is the marginal density. Dashed lines represent the within-class densities weighted by their respective class probabilities: Class 1 (left), Class 2 (middle), Class 3 (right). Generating parameters: $p(c_i = 1) = .67$, $p(c_i = 2) = .25$, $p(c_i = 3) = .08$, $\sigma^{2(1)} = 3$, $\sigma^{2(2)} = 12$, $\sigma^{2(3)} = 10$, $\mu^{(1)} = 5$, $\mu^{(2)} = 10$, $\mu^{(3)} = 20$.

Suppose a UFNМ is fit where y_i is a total score on a depression inventory such as the Hamilton Rating Scale, and $K = 3$. A “direct” interpretation may focus on estimates of data model parameters in Equations (5), (6), and (8) that define the weighted class-specific densities in Figure 1. Doing so may suggest that Classes 1, 2, and 3 correspond to normal, subclinical, and clinically depressed population subgroups, consistent with the fact that $\mu^{(3)}$ (for Class 3, constituting 8% of the sample) falls in the Hamilton’s clinical range, whereas $\mu^{(1)}$ (for Class 1, constituting 67% of the sample) falls in the normal range. In contrast, an “indirect” interpretation of the UFNМ could suggest that these three classes instead serve to flexibly approximate a nonnormal-but-continuous population density of depression when marginalizing across class (bold solid line in Figure 1). This interpretation emphasizes that different combinations of locations, $\mu^{(k)}$; spreads, $\sigma^{2(k)}$; and weights, $p(c_i = k)$, could account for nearly any marginal shape.

Posterior Probabilities

Estimates from the best-fitting UFNМ may be used to calculate posterior probabilities of class membership—each person’s model-implied probability of being

in each class given their observed data. Conveniently, these K posterior probabilities per person i are produced at the final step of the algorithm usually used for fitting mixture models—the Expectation-Maximization algorithm (see McLachlan & Peel, 2000; B. O. Muthén & Shedden, 1999). These posterior probabilities are computed using *Bayes' rule* from Equation (4). For UFNM this yields

$$p(c_i = k | y_i) = \frac{p(c_i = k) f(y_i | c_i = k)}{f(y_i)}, \quad (10)$$

where the numerator consists of Equations (7) and (8), and the denominator is Equation (9), all with estimates substituted for corresponding parameters. Posterior probabilities can be further collapsed into *modal class assignments*—assignments of each person to the class for which he or she has the highest posterior probability of membership. This constitutes an Empirical Bayes Modal approach to assigning each person a score on the discrete latent variable (Skrondal & Rabe-Hesketh, 2004). These modal class assignments can be used for descriptive purposes, diagnostics, assessments of classification accuracy, or subsequent predictive analyses (but see the Discussion).

LATENT PROFILE ANALYSIS (LPA)

For LPA (e.g., Gibson, 1959) *multivariate* cross-sectional data are instead required on $j = 1 \dots J$ continuous outcomes, for instance, severity ratings on $J = 4$ depression symptoms. These outcomes are theorized to measure the same underlying latent construct—for example, a categorical latent depression variable with K classes. If K discrete subpopulations with different profiles of means and variances on the J variables are mixed together, this can give rise to nonnormal marginal distributions of the variables (as in the UFNM) but also to associations among the variables in the population at large.

Within-Class Model

For outcome variable j and person i in class k (i.e., where $c_i = k$) we have

$$y_{ij} = \mu_j^{(k)} + \varepsilon_{ij} \quad (11)$$

$$\varepsilon_{ij} \sim N(0, \sigma_j^{2(k)}). \quad (12)$$

The means $\mu_j^{(k)}$ and variances $\sigma_j^{2(k)}$ may vary across $j = 1 \dots J$ and $k = 1 \dots K$. Within-class, each outcome is assumed univariate normally distributed

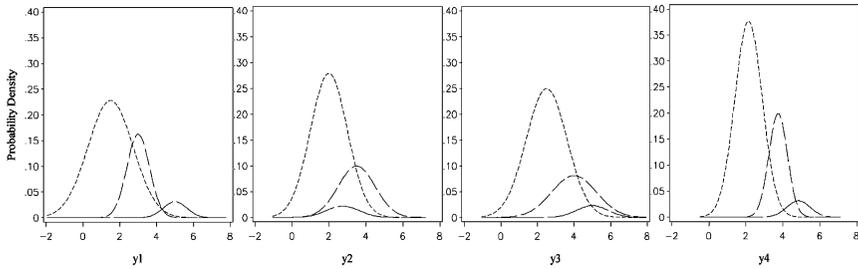


FIGURE 2 Illustrative LPA model-implied class-specific densities for each of $J = 4$ outcomes (shown weighted by corresponding class probabilities, where $K = 3$). *Note.* Dashed lines represent within-class densities for each outcome, weighted by their respective class probabilities. The weighting is used here only for visualization purposes. In LPA estimation, weighting is done for the joint outcome density, not individual outcome densities (see Equation (14)). LPA = latent profile analysis. Classes 1, 2, and 3 correspond with short, medium, and long dashes.

(Equation (7)) as in the UFGM; this is illustrated in Figure 2 for a simulated depression symptom severity LPA, where $K = 3$ and $J = 4$.⁶ For each of the outcome panels in Figure 2, the k th within-class density has a location centered at $\mu_j^{(k)}$, with spread $\sigma_j^{2(k)}$.

According to the important *axiom (assumption) of local independence*, the joint density of outcomes within class is the product of J univariate normal densities, using a continuous-variable analogue of the *multiplication rule for conditionally independent events* in Equation (2).

$$f(\mathbf{y}_i | c_i = k) = \prod_{j=1}^J f(y_{ij} | c_i = k). \quad (13)$$

This assumption implies that associations among y 's in the population at large are fully accounted for by between-class mean differences. As such, *within-class* residual covariances are assumed to be 0.

Between-Class Model

In LPA, the probability of membership in class k is as in Equation (8).

⁶For each outcome in the LPA, Figure 2 depicts class-specific densities weighted by their class probabilities; however, weighting is used here only for ease of visualization. In LPA estimation, weighting is done for the joint outcome density, not individual outcome densities (see Equation (14)).

Combined Model

Similarly to the UFM, using the *law of total probability* in Equation (1), the marginal PDF for person i is obtained as a sum (across classes) of the joint within-class density for the J variables weighted by the probability of membership in that class:

$$f(\mathbf{y}_i) = \sum_{k=1}^K p(c_i = k) f(\mathbf{y}_i | c_i = k). \quad (14)$$

Posterior Probabilities

As in the UFM, in LPA applying *Bayes' rule* in Equation (4) yields

$$p(c_i = k | \mathbf{y}_i) = \frac{p(c_i = k) f(\mathbf{y}_i | c_i = k)}{f(\mathbf{y}_i)}. \quad (15)$$

Suppose each of person i 's depression symptom severity outcomes (in \mathbf{y}_i) falls in a range of great overlap among the $K = 3$ class-specific densities in Figure 2. For instance, $\mathbf{y}_i = 4.25, 2.5, 5.0, 3.75$. Then, person i 's corresponding posterior probabilities of class membership (.01, .67, .32 for Classes 1, 2, and 3, respectively) will imply relatively large classification uncertainty. But if person i 's \mathbf{y}_i falls in a range of little overlap in Figure 2 (say, $\mathbf{y}_i = 1.5, 1.0, 2.0, 2.5$), his or her corresponding posterior probabilities (>.99, <.01, <.01 for Classes 1, 2, and 3, respectively) will imply relatively small classification uncertainty.

LATENT CLASS ANALYSIS (LCA)

As in the LPA, for LCA (e.g., Clogg, 1995; Lazarsfeld & Henry, 1968) data are multivariate and cross-sectional. But now we have J categorical (here, binary) outcomes, for instance, depression symptoms where $y_{ij} = 1$ if endorsed and $y_{ij} = 0$ if denied. In LCA applications, binary outcomes are most common (Collins & Lanza, 2010). These J outcomes are anticipated to be associated in the population because they are all indicators of the same categorical latent construct (e.g., depression), with K classes. LCA can account for the marginal distribution of each outcome (described by its proportion endorsed) as well as associations among outcomes (described by a contingency table). Recall that a contingency table for binary outcomes has 2^J cells. Each cell contains a count of participants providing a certain *response pattern* (vector of 0 or 1 responses to all J outcomes). Associations among outcomes give rise to patterns that occur

with disproportionate prevalence.⁷ Just as LPA accounted for covariances among metric variables in the population at large using between-class mean differences, LCA can account for marginal associations among categorical outcomes using between-class differences in patterns of outcome endorsement probabilities.

Within-Class Model

In LCA with binary outcomes, a binary logistic within-class model can be specified for the probability of endorsing outcome j within class k (denoted $\pi_j^{(k)}$):

$$\pi_j^{(k)} = 1/(1 + \exp(\tau_j^{(k)})). \quad (16)$$

$\tau_j^{(k)}$ is an estimated threshold for the j th outcome in class k ; it can be considered a cutpoint on an unobserved “underlying propensity” to endorse an outcome. If one’s underlying propensity exceeds the threshold, one responds 1. If not, one responds 0. Because this underlying propensity is unobserved, for identification purposes its residuals are here assumed to follow a standard logistic distribution (see Long, 1997, pp. 41–42). Consequently, in LCA, residual variances are not free parameters and cannot vary across class as in LPA. In LCA, thresholds $\tau_j^{(k)}$ are allowed to differ across class. The j th variable in class k is Bernoulli distributed; its within-class probability mass function (PMF) is

$$p(y_{ij}|c_i = k) = (\pi_j^{(k)})^{y_{ij}} (1 - \pi_j^{(k)})^{1-y_{ij}}. \quad (17)$$

According to the important assumption of local independence, responses are assumed independent within class. This means if we were to construct a contingency table just for persons in class k there should be no remaining association among their depression symptoms. This assumption would be violated in a binary depression symptom LCA if, within class k , persons still had a higher probability of endorsing y_{i1} (say, fatigue) if they had endorsed y_{i2} (say, insomnia). Assuming local independence, the joint PMF of all outcomes in class k is the product of the J outcome probabilities in class k , using the *multiplication rule for conditionally independent events* in Equation (2):

$$p(\mathbf{y}_i|c_i = k) = \prod_{j=1}^J p(y_{ij}|c_i = k). \quad (18)$$

⁷Even when discrete subpopulations exist, many more response patterns than classes can arise due to measurement error.

Between-Class Model

LCA's between-class model is Equation (8), as in the UFNM and LPA.

Combined Model

The marginal (across-class) PMF⁸ of the response vector \mathbf{y}_i in the overall population is constructed using the *law of total probability*, Equation (1), as in the UFNM and LPA:

$$p(\mathbf{y}_i) = \sum_{k=1}^K p(c_i = k)p(\mathbf{y}_i|c_i = k). \quad (19)$$

Posterior Probabilities

As in previous models, posterior probabilities are calculated in LCA using *Bayes' rule* (Equation (4)):

$$p(c_i = k|\mathbf{y}_i) = \frac{p(c_i = k)p(\mathbf{y}_i|c_i = k)}{p(\mathbf{y}_i)}. \quad (20)$$

GROUPS-BASED TRAJECTORY (GBT) MODEL

Prototypical mixtures we have considered thus far—UNFM, LPA, and LCA—all used cross-sectional data. For the GBT model (B. O. Muthén, 2001; Nagin, 1999; also called a latent class growth model) longitudinal panel data (J repeated measures on a single outcome) are required. Most GBT applications employ conditionally normal repeated measures (see Sterba, Baldasaro, & Bauer, 2012), as we do here; however, it has been applied with other within-class distributions (e.g., Poisson or Bernoulli). The GBT model accounts for the pattern of means and (co)variances of repeated measures in the population via a mixing of homogeneous classes of persons who, within class, follow the same trajectory of change apart from pure error. That is, the local independence assumption, familiar from LPA and LCA, is imposed within trajectory-class. Because classes can differ in mean trajectory parameters, they can differ in functional form. For instance, in Campbell, Matestic, von Stauffenberg, Mohan, and Kirchner's (2007) GBT application for $J = 7$ repeated measures of a depressive symptom scale, the best fitting $K = 6$ consisted of three nonlinear trajectories (cubic-increasing, cubic-decreasing, cubic-unstable) and three linear trajectories (high-chronic, moderate-decreasing, low-decreasing).

⁸When evaluated for a particular response pattern, it is a probability.

Within-Class Model

In each class, a model for change is specified, most commonly polynomial change. However, in theory any change process could be specified (for other examples see Dolan, Schmittmann, Lubke, & Neale, 2005; Grimm, Ram, & Estabrook, 2010). The GBT extends the LPA's within-class model in the manner of a regression mixture (e.g., Wedel & DeSarbo, 2002) by structuring the means $\mu_j^{(k)}$ from Equation (11) to make them dependent on time, such as

$$\mu_j^{(k)} = \eta_0^{(k)} + \eta_1^{(k)} \text{time}_{ij} + \eta_2^{(k)} \text{time}_{ij}^2. \quad (21)$$

This yields a quadratic GBT model for the j th repeated measure on the i th person in the k th class:

$$y_{ij} = \eta_0^{(k)} + \eta_1^{(k)} \text{time}_{ij} + \eta_2^{(k)} \text{time}_{ij}^2 + \varepsilon_{ij} \quad (22)$$

$$\varepsilon_{ij} \sim N(0, \sigma_j^{2(k)}). \quad (23)$$

time_{ij} is the measurement time for person i at time j , usually centered at a meaningful value such as initial status. $\eta_0^{(k)}$, $\eta_1^{(k)}$, and $\eta_2^{(k)}$ are a fixed class-specific intercept, linear slope, and quadratic slope, respectively. To facilitate interpreting these slopes as indicating quantitative differences in the construct over time, for class k , researchers typically assume the J outcomes are measurement invariant—that they measure the same construct at all timepoints. ε_{ij} is person i 's deviation from his or her class trajectory, at time j , and $\sigma_j^{2(k)}$ is the within-class residual variance at time j (often constrained equal across class and time: σ^2).

The within-class univariate normal PDF for repeated measure j is given in Equation (7) with model-implied mean in Equation (21). The imposition of the local independence assumption for GBT implies that, within class, residual covariances = 0, as in LPA. Under this assumption, the joint density of all repeated measures for person i is constructed as in Equation (13), using a continuous-variable analogue of the *multiplication rule for conditionally independent events* (Equation (2)).

Between-Class Model

The between-class model is Equation (8), as in UFNM, LPA, and LCA.

Combined Model

The marginal density of the repeated measures in GBT, given in Equation (14), is constructed using the *law of total probability* (Equation (1)), as for UFNM, LPA, and LCA.

Posterior Probabilities

For the GBT model, posterior probabilities are computed as in Equation (15) (Nagin, 2005), using *Bayes' rule* (Equation (4)), as in previous models.

LATENT TRANSITION ANALYSIS (LTA)

The final prototypic mixture we consider is a member of the Markov model family. Markov models in general allow (observed or latent) discrete states at timepoint t to be regressed on discrete states at previous timepoint(s), usually just timepoint $t - 1$. They can be fit with one outcome repeatedly measured at each of T timepoints (e.g., Langeheine, 1994; Vermunt, Tran, & Magidson, 2006) or J outcomes repeatedly measured at each of T timepoints—that is, *multivariate longitudinal* data. Here we consider a Markov model involving J outcomes at T timepoints that is particularly popular in psychology—termed Latent Transition Analysis (LTA; e.g., Collins & Wugalter, 1992; Reboussin, Reboussin, Liang, & Anthony, 1998). We index $j = 1 \dots J$ and $t = 1 \dots T$. LTAs are usually specified with categorical (here, binary) outcomes, although in principle Markov models can be fit with different outcome types (e.g., Schmittmann, Dolan, van der Maas, & Neale, 2005). Here, for outcome j at time t , suppose $y_{ijt} = 1$ if endorsed and $y_{ijt} = 0$ if denied.

LTA employs a categorical latent variable at *each* timepoint $c_{i1} \dots c_{iT}$. The c_{it} explains associations among the J outcomes at time t . Latent classes at a given timepoint are called *latent states* in LTA to underscore that their membership can shift across time—for instance, from State 2 at Time 1 to State 3 at Time 2. To explain associations among the J outcomes *across* time, LTA regresses consecutive categorical latent variables on each other. As such, LTA can capture qualitative, state-sequential change patterns rather than quantitative change patterns, as in the conventional GBT model. An LTA is also called a *multiple indicator latent Markov model* (Langeheine, 1994) because it can be conceptualized as an LCA measurement model at each time (relating *multiple* outcomes at time t to latent states at time t) connected by a *latent* Markov structural model (regressing latent state membership at time t on membership at time $t - 1$). LTA serves to account for the marginal distribution of each binary outcome (described by the proportion endorsed for each of the $J \times T$ outcomes) and associations among outcomes (described by a contingency table with 2^{JT} cells). Each contingency table cell contains the number of participants providing a particular complete response pattern—a vector of responses (0 or 1) to all $J \times T$ outcomes.

Within-State Model

In LTA we have a within-state model for *each* $c_{i1} \dots c_{iT}$. Here consider $T = 2$ timepoints where c_{i1} has $k = 1 \dots K$ states and c_{i2} has $m = 1 \dots M$ states. (The

online appendix gives the model for $T = 3$). Although potentially $K \neq M$, often they may be equal. In our empirical LTA example (details given later), we have $K = M = 2$ and have $J = 4$ binary depression symptoms measured at time t : sad/depressed, slow-moving, irritability, and concentration problems. At Time 1, these are denoted $y_{i11} - y_{iJ1}$ and are indicators of c_{i1} (latent depression states at Time 1). At Time 2, these are denoted $y_{i12} - y_{iJ2}$ and are indicators of c_{i2} (latent depression states at Time 2). Equation (24) gives the within-state model at $t = 1$ for persons in state k , as does Equation (25) at $t = 2$ for persons in state m :

$$\pi_{j1}^{(k)} = 1/(1 + \exp(\tau_{j1}^{(k)})) \quad (24)$$

$$\pi_{j2}^{(m)} = 1/(1 + \exp(\tau_{j2}^{(m)})). \quad (25)$$

$\pi_{j1}^{(k)}$ is the endorsement probability for the j th outcome at Time 1 given membership in state k . $\pi_{j2}^{(m)}$ is the endorsement probability for the j th outcome at Time 2 given membership in state m . For our LTA example, Figure 3 depicts these endorsement probabilities at Time 1 (Panel 1) and Time 2 (Panel 2). Figure 3 shows that at both times, persons in State 2 (solid line; “low risk”) have lower endorsement probabilities for all outcomes than persons in State 1 (dashed line; “clinical risk”). Also, at both times, persons in State 1 have much higher probabilities of irritability and concentration problems.

In state k at Time 1, y_{ij1} has a Bernoulli PMF, as does y_{ij2} in state m at Time 2:

$$p(y_{ij1}|c_{i1} = k) = (\pi_{j1}^{(k)})^{y_{ij1}} (1 - \pi_{j1}^{(k)})^{1-y_{ij1}} \quad (26)$$

$$p(y_{ij2}|c_{i2} = m) = (\pi_{j2}^{(m)})^{y_{ij2}} (1 - \pi_{j2}^{(m)})^{1-y_{ij2}}. \quad (27)$$

Outcomes at time t are assumed locally independent within state: hence, persons within state at time t should have interchangeable responses. Reflecting the imposition of this assumption, the *multiplication rule for conditionally independent events* is used to calculate the joint density of outcomes at Time 1 (Equation (28)) or Time 2 (Equation (29)):

$$p(\mathbf{y}_{i1}|c_{i1} = k) = \prod_{j=1}^J p(y_{ij1}|c_{i1} = k) \quad (28)$$

$$p(\mathbf{y}_{i2}|c_{i2} = m) = \prod_{j=1}^J p(y_{ij2}|c_{i2} = m). \quad (29)$$

\mathbf{y}_{i1} and \mathbf{y}_{i2} are the response patterns for the J outcomes at Times 1 and 2, respectively, for person i . Equations (24)–(29) look identical to the LCA Equations

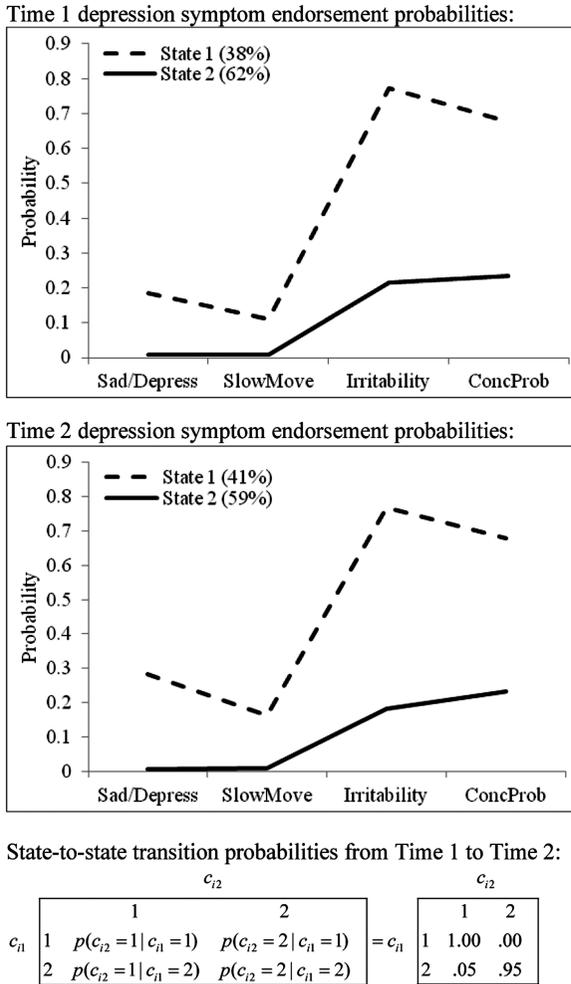


FIGURE 3 Empirical example results: latent transition analysis of depressive symptoms. Note. *Depress* = depressed; *SlowMove* = slow moving; *ConcProb* = concentration problems.

(16)–(18) except for the time-specific subscript. Further increasing the similarity, when $K = M^9$ the j th outcome threshold is commonly constrained equal across time within state—hence removing the t subscript. If this constraint were

⁹Having $K = M$ corresponds with configural invariance of the categorical latent variable across time.

imposed in the LTA depicted in Figure 3, the line plots in Panel 1 versus Panel 2 would look identical. If this constraint is justified, measurement invariance holds¹⁰ for the depression categorical latent variables across time, and transition probabilities from state to state are easier to interpret.

Between-State Model

The between-state model in LTA determines the Time 1 latent state probabilities. Additionally, by regressing c_{i2} on c_{i1} , it determines the probabilities of longitudinally transitioning from a given state to any other state.¹¹ The same multinomial logit model in Equation (8) is used to obtain the Time 1 latent state probabilities, denoted $p(c_{i1} = k)$. Subsequent latent state probabilities can be solved for from estimated parameters (Langeheine, 1994), as described in the Appendix. Because c_{i2} has ≥ 2 nominal categories, to regress c_{i2} on c_{i1} , a multinomial logit specification is used. Because c_{i1} also has ≥ 2 nominal categories, a set of $K - 1$ dummy variables can be used to represent it (e.g., Asparouhov & Muthén, 2008; Nylund, 2007; Reboussin et al., 1998). To illustrate, if $K = M = 3$, the multinomial regression of c_{i2} on c_{i1} is

$$p(c_{i2} = m | c_{i1} = k) = \frac{\exp(\alpha_m + \beta_{m1}d_{i1} + \beta_{m2}d_{i2})}{\sum_{g=1}^M \exp(\alpha_g + \beta_{g1}d_{i1} + \beta_{g2}d_{i2})}. \quad (30)$$

$p(c_{i2} = m | c_{i1} = k)$ is a *transition probability*—the probability of membership in state m at Time 2 given membership in state k at Time 1. Dummy variable d_{i1} is set to 1 to represent $k = 1$ and d_{i2} is set to 1 to represent $k = 2$. Different combinations of dummy variables are used to calculate different transition probabilities. For instance, calculating $p(c_{i2} = m | c_{i1} = 1)$ requires substituting $d_{i1} = 1$, $d_{i2} = 0$. Calculating $p(c_{i2} = m | c_{i1} = 2)$ needs $d_{i1} = 0$, $d_{i2} = 1$. Calculating $p(c_{i2} = m | c_{i1} = 3)$ needs $d_{i1} = d_{i2} = 0$. Terms involving the reference state at Time 2 (here, State 3) are set to 0: $\alpha_3 = \beta_{31} = \beta_{32} = 0$. Equation (30) implies that everyone in state k at Time 1 has the same probability of transitioning to state m at Time 2.

Even though the multinomial coefficients in Equation (30) are estimated, they are usually not interpreted directly; rather, they are simply used to calculate transition probabilities, which are then interpreted. These transition probabilities are often presented in a $K \times M$ transition probability matrix (Collins & Wugalter,

¹⁰The testability of measurement invariance in the conventional LTA contrasts to its untestability in the conventional GBT (unless extended to a second-order GBT, as in Grimm & Ram, 2009).

¹¹If $T > 2$, it could be possible to regress latent states at time t on prior states at both $t - 1$ and $t - 2$.

1992). Because our depression LTA empirical example only has $K = M = 2$, when using Equation (30) we can omit terms involving d_{i2} . The 2×2 transition probability matrix for this example (bottom of Figure 3) shows considerable stability in depression state membership. But when change occurs, it is more likely to be into a high-risk state than out of it.

Combined Model

For a $T = 2$ LTA, the marginal PMF for a response pattern for $J \times T$ outcomes for person i (denoted \mathbf{y}_i) is typically given as in Equation (31); it is discussed further in the next section.

$$p(\mathbf{y}_i) = \sum_{m=1}^M \sum_{k=1}^K p(c_{i1} = k) p(c_{i2} = m | c_{i1} = k) p(\mathbf{y}_{i1} | c_{i1} = k) p(\mathbf{y}_{i2} | c_{i2} = m). \quad (31)$$

Posterior Probabilities

If $T = 2$, we can compute person i 's posterior probability of membership in each of $K \times M$ patterns of state-to-state memberships (defined as a *sequence* in the next section).

$$\begin{aligned} p(c_{i1} = k, c_{i2} = m | \mathbf{y}_i) \\ = \frac{p(c_{i1} = k) p(c_{i2} = m | c_{i1} = k) p(\mathbf{y}_{i1} | c_{i1} = k) p(\mathbf{y}_{i2} | c_{i2} = m)}{p(\mathbf{y}_i)}. \end{aligned} \quad (32)$$

SUMMARY OF RELATIONSHIPS AMONG UFNM, LPA, LCA, GBT, AND LTA

Table 1 summarizes parameters discussed here for the five prototypical mixtures. All of these parameters may not be uniquely estimable in a given application. For instance, if the membership probability for one class approaches 0 and/or if the parameter estimates in one class nearly coincide with those of another class, empirical underidentification issues can arise, necessitating additional constraints on the Table 1 parameters. For other identification considerations, see, for example, Frühwirth-Schnatter (2009) and Abar and Loken (2012).

Many interconnections among the five prototypical mixtures reviewed here could be highlighted. For instance, removing the time covariate from a GBT yields an LPA, removing all but one outcome from an LPA yields a UFNM, changing the conditional distribution of the outcome from normal to Bernoulli

TABLE 1
Parameters in the Five Prototypical Mixture Models Reviewed

<i>Model</i>	<i>Maximum No. of Estimated Parameters^a</i>
UFNM	K outcome means K outcome variances $K - 1$ multinomial intercepts
LPA	$K \times J$ outcome means $K \times J$ outcome variances $K - 1$ multinomial intercepts
LCA (binary)	$K \times J$ outcome thresholds $K - 1$ multinomial intercepts
GBT	$K \times (1 + b)$ growth coefficients $K \times J$ repeated measure variances $K - 1$ multinomial intercepts
LTA (assuming #states, K , constant over time)	$K \times J \times T$ outcome thresholds $K - 1$ multinomial intercepts, for initial latent states $(T - 1)K(K - 1)$ multinomial coefficients, for transitions

Note. b = the order of the polynomial. UFNM = univariate finite normal mixture; LPA = latent profile analysis; LCA = latent class analysis; GBT = groups-based trajectory model; LTA = latent transition analysis. J , T , K , and M were defined in the text.

^aFurther constraints are common for parsimony and/or to prevent empirical underidentification (see text). Further constraints are also used to, for instance, impose threshold invariance within state across time in LTA.

converts an LPA to a LCA, and an LCA for an initial timepoint is embedded in the LTA. However, in this section we focus on three other relationships among these models that pertain to their reliance on probability laws introduced earlier. These relations are summarized in Table 2 and detailed here.

Combined Model and the Law of Total Probability

For UFNM, LPA, LCA, and GBT, the construction of the combined model (marginal density) was a straightforward application of the law of total probability (Equation (1)). However, for LTA, the conventional representation of the combined model needs to be rewritten to better see how this law is applied. Specifically, Equation (31) may be expressed as a sum of *within-sequence PMFs* weighted by respective *sequence probabilities*:

$$p(\mathbf{y}_i) = \sum_{m=1}^M \sum_{k=1}^K p(c_{i1} = k, c_{i2} = m) p(\mathbf{y}_i | c_{i1} = k, c_{i2} = m). \quad (33)$$

TABLE 2
Summary of Unifying Themes Among Mixture Models Illustrated in This Review

<i>Theme</i>	<i>Pertinent Equations</i>
The marginal density of each mixture model is constructed using the law of total probability, as a sum of within-class (-sequence) densities weighted by class/sequence probabilities.	Equation (9) Equation (14) Equation (19) Equation (33)
Posterior probabilities of class/sequence membership—widely used in assessing classification accuracy—are computed for all mixture models considered, by applying Bayes' rule.	Equation (10) Equation (15) Equation (20) Equation (34)
For LPA, LCA, GBT, and LTA, the assumption of local independence implies that the joint density of responses within class (-state/sequence) is constructed using the multiplication rule for conditionally independent events.	Equation (13) Equation (18) Equations (28)–(29)
Hybrid mixtures relax the local independence assumption—using an extension of the multiplication rule for conditionally independent events—by conditioning on continuous latent factor(s) in the within-class (-state) model.	Equations (35)–(42)
Parallel-process mixtures relax the local independence assumption for outcomes from two processes by expanding the between-class (-state) model using the multiplication rule for dependent events.	Equation (43) Equation (45)

Note. LPA = latent profile analysis; LCA = latent class analysis; GBT = groups-based trajectory model; LTA = latent transition analysis.

A *sequence probability* is the joint probability of being in a particular state at Time 1 and a particular state at Time 2. By the multiplication rule for dependent events (Equation (3)) the sequence probability, $p(c_{i1} = k, c_{i2} = m)$, is the product of the marginal probability $p(c_{i1} = k)$ and the conditional probability $p(c_{i2} = m | c_{i1} = k)$ from Equation (30). The *within-sequence PMF*, $p(\mathbf{y}_i | c_{i1} = k, c_{i2} = m)$, is the product of Equations (28) and (29) because of assumptions that outcomes in \mathbf{y}_i are independent after accounting for Time 1 state ($c_{i1} = k$), outcomes in \mathbf{y}_i are independent accounting for Time 2 state ($c_{i2} = m$), and \mathbf{y}_{i1} and \mathbf{y}_{i2} are independent after accounting for the sequence of $c_{i1} = k, c_{i2} = m$. Note that instead of simply summing over classes of a single categorical latent variable—as in UFNM, LPA, LCA, and GBT—we sum over all $K \times M$ sequences in applying the law of total probability in LTA.

Also, each mixture's combined model (marginal density) helps us understand its "indirect" interpretation. Crucial to the "indirect" interpretation is the fact that distributional assumptions are imposed only in the within-class (-state) and between-class (-state) submodels and that no additional assumptions are placed on the marginal density itself in Equations (9), (14), (19), or (31). Because of this fact, the pooled across-class marginal density is allowed to flexibly approximate a variety of shapes for the population as a whole. In contrast, for each mixture, the "direct" interpretation does not focus on the marginal density but rather focuses on interpreting parameters of the within-class (-state) and between-class (-state) submodels as descriptive of corresponding discrete latent subpopulations in the real world.

Posterior Probabilities of Class Membership and Bayes' Rule

Earlier, posterior probabilities from UFNM, LPA, LCA, and GBT were obtained using Bayes' rule in Equation (4) by multiplying the "prior" probability of class membership (between-class model) by the conditional density of the outcome(s) given class (within-class density) and dividing by the marginal density (combined model). For LTA, Bayes' rule can be similarly applied to obtain posterior probabilities of *sequence* membership; this is more clearly seen if Equation (32) is rewritten as

$$p(c_{i1} = k, c_{i2} = m | \mathbf{y}_i) = \frac{p(c_{i1} = k, c_{i2} = m) p(\mathbf{y}_i | c_{i1} = k, c_{i2} = m)}{p(\mathbf{y}_i)}. \quad (34)$$

In UFNM, LPA, LCA, and GBT, the K posterior probabilities of class membership *for person i* sum to 1. Likewise, in LTA the $K \times M$ posterior probabilities of sequence membership *for person i* sum to 1. The fact that each person has a nonzero posterior probability of being in each class (or each sequence) highlights the probabilistic/fuzzy nature of class (sequence) membership in mixture models.

In Equations (10), (15), (20), and (34), \mathbf{y}_i has more influence on the posterior probability for classes (sequences) with higher "prior" membership probabilities.¹² For example, consider a $K = 2$ LCA with $J = 4$ depression symptoms where $k = 1$ is characterized by high symptom probabilities ($\pi_1^{(1)} = .8$; $\pi_2^{(1)} = .65$; $\pi_3^{(1)} = .9$; $\pi_4^{(1)} = .75$) and $k = 2$ by relatively low symptom probabilities ($\pi_1^{(2)} = .3$; $\pi_2^{(2)} = .15$; $\pi_3^{(2)} = .25$; $\pi_4^{(2)} = .2$). The posterior probability of Class 1 membership for a person with a clinically extreme response pattern, $p(c_i = 1 | \mathbf{y}_i = 1, 1, 1, 1)$, is .89 if the prior for the high endorsement class is

¹²Analogously, in continuous latent variable models, the posterior density can be "shrunk" toward the mean of the prior density (Skronal & Rabe-Hesketh, 2004).

very low, $p(c_i = 1) = .05$, but rises to $p(c_i = 1 | \mathbf{y} = 1, 1, 1, 1) = .99$ if instead $p(c_i = 1) = .50$.

Posterior probabilities are used in assigning persons to classes (sequences), usually using the modal assignment rule described earlier but sometimes using a random allocation rule (e.g., Goodman, 2007; Vermunt, 2010a). Using the random allocation rule with UNFM, LPA, LCA, or GBT, the class assignment for person i is generated from a multinomial distribution with K category probabilities set equal to person i 's own K posterior probabilities. Class k 's expected assignment proportion approximates $p(c_i = k)$ under the model. For LTA, the same procedure would instead be done for the $K \times M$ sequences. Modal class (sequence) assignment is a discrete latent variable analogue to familiar continuous latent variable scoring procedures (Empirical Bayes Modal or Modal a Posteriori; Skrondal & Rabe-Hesketh, 2004). The latter involve assigning each person a score that is the mode of their continuous posterior density. Imprecision of such assignments/scores is a concern. For mixtures, it may be investigated within sample, conditional on the parameter estimates. This may involve calculating the average posterior probability of belonging to a class/sequence among persons assigned to that class/sequence (e.g., Nagin, 2005) or calculating entropy (e.g., Jedidi, Jagpal, & Desarbo, 1997).

Local Independence and the Multiplication Rule for Conditionally Independent Events

Earlier, the multivariate mixtures LPA, LCA, and GBT were seen to impose local independence within class, enabling the joint density of responses within class to be computed using Equation (2). For LTA, because all across-time dependencies among outcomes are exclusively accounted for in the between-state model, local independence among outcomes is assumed to hold within *sequence*. Additionally, for LTA, local independence for outcomes at time t is assumed within-state at time t , as previously noted. These two assumptions are used in calculating LTA's joint within-sequence PMF from Equation (33) as $p(\mathbf{y}_i | c_{i1} = k, c_{i2} = m) = p(\mathbf{y}_{i1} | c_{i1} = k, c_{i2} = m) p(\mathbf{y}_{i2} | c_{i1} = k, c_{i2} = m) = p(\mathbf{y}_{i1} | c_{i1} = k) p(\mathbf{y}_{i2} | c_{i2} = m)$, where the first assumption permits the first equality (using an expansion of Equation (2)) and the second assumption permits the second equality. These assumptions imply that, in LTA, temporal relations among \mathbf{y}_i are solely accounted for by the mixing together of sequences with different patterns of state-to-state change.

In all of these multivariate mixtures, the local independence assumption not only simplifies estimation but also plays a key role in simplifying interpretation of classes as homogeneous, with responses interchangeable within class. However, there may be circumstances necessitating further relaxation of this assumption, such as when systematic individual heterogeneity within class/sequence is

posited. Two approaches¹³ for further relaxing local independence are considered here (beyond simply increasing the number of classes or states, which effectively makes this assumption less strict).

The first approach involves expanding the within-class (-state) model to condition on additional variable(s) besides class/state membership. In terms of general probability laws, this approach involves expanding Equation (2) as in $P(A, C|B, D) = P(A|B, D)P(C|B, D)$. This first approach is employed in the *hybrid mixtures* section later, wherein the conditioning variable within-class (-state) is a continuous latent factor, η_i . Local independence is now imposed such that, when the latent factor is held fixed, outcomes need to be independent conditional on class membership *and* factor score, for example, $f(\mathbf{y}_i|c_i = k, \eta_i) = \prod_{j=1}^J f(y_{ij}|c_i = k, \eta_i)$ (Bartholomew & Knott, 1999). Hybrid mixtures marginalize over η_i to get back to a representation of $f(\mathbf{y}_i|c_i = k)$ (e.g., B. O. Muthén, Asparouhov, & Rebollo, 2005). This generally involves integration, except when outcomes are normal within class; then a closed-form solution is available.

The second approach for relaxing local independence of outcomes involves expanding the between-class (-state) model to allow class membership to itself depend on another classification variable. Outcomes are then assumed locally independent given a particular *combination* of memberships on the classification variables. Incidentally, LTA already employed this second approach to some degree when accommodating *across-time* dependency for \mathbf{y}_{i1} and \mathbf{y}_{i2} exclusively through state-to-state autoregressions in the between-state model. This second approach is further applied in the *parallel-process mixture* section later to accommodate *across-process* dependency among outcomes in two processes. Specifically, the between-state (-class) model is further expanded using the *multiplication rule for dependent events* in Equation (3) to make state/class membership in one process (say, depression) conditional now on state/class membership of another process (say, oppositional defiant behavior). The parallel-process GBT will, as one of its assumptions, require local independence of outcomes given the combination of class memberships in both processes. The parallel-process LTA will, as one of its assumptions, require local independence of outcomes given the combination of sequence memberships in both processes.

EXTENSION: HYBRID MIXTURE MODELS

The recent extension called hybrid mixtures (Muthén, 2008) employs the first approach discussed earlier to further relax local independence. Outcomes within

¹³A third approach, not discussed here, is unavailable for LCA and LTA but available for LPA and GBT. It involves estimating residual covariances within class; estimating many such covariances and/or allowing them to differ across class can incur estimation problems (Lubck & Neale, 2006).

class can mutually depend on continuously distributed latent factor(s) to accommodate associations among responses within class and represent systematic individual variability within class. This extension has appeared under separate names; for LPA it is called a factor mixture model (e.g., Lubke & Muthén, 2005; Yung, 1997); for GBT, a growth mixture model (e.g., Muthén & Shedden, 1999; Verbeke & Lesaffre, 1996); for LCA, a categorical-item factor mixture or IRT mixture model (Lubke & Neale, 2008; Mislevy & Verhelst, 1990; Muthén & Asparouhov, 2006; Rost, 1990); and for LTA, LTA with a categorical-item factor/IRT measurement model (Cho, Cohen, Kim, & Bottge, 2010; Nylund, 2007). Here we highlight the similarity of this extension across multivariate mixtures by reviewing the inclusion of one factor per class/state ($q = 1$) in LPA, LCA, GBT, and LTA. For each such hybrid version, the between-class model and combined model are unchanged; only the within-class model changes.

LPA \rightarrow Factor Mixture Model

To convert an LPA into a factor mixture model (Lubke & Muthén, 2005; Yung, 1997) within class k , outcome j is regressed on latent factor(s)—here a single factor η_i with factor loading (regression weight) $\lambda_j^{(k)}$. Within class k , the factor has mean $\alpha^{(k)}$ and normally distributed individual deviation from the mean ζ_i . Rather than estimating a mean and variance for the j th outcome, as in the unconditional LPA, we estimate an intercept $v_j^{(k)}$ and residual variance $\theta_{\varepsilon_j}^{(k)}$:

$$y_{ij} = v_j^{(k)} + \lambda_i^{(k)}\eta_i + \varepsilon_{ij} \quad (35)$$

$$\eta_i = \alpha^{(k)} + \zeta_i$$

where

$$\zeta_i \sim N(0, \psi^{(k)}) \quad (36)$$

$$\varepsilon_{ij} \sim N(0, \theta_{\varepsilon_j}^{(k)}). \quad (37)$$

For identification we may, for instance, fix $\alpha^{(k)} = 0$ and estimate all intercepts and fix $\psi^{(k)} = 1$ and estimate all factor loadings. Now that outcomes within class depend on a factor, local independence is relaxed and Equation (13) no longer holds. Because ε_{ij} and ζ_i are both normal, the within-class PDF for \mathbf{y}_i , marginalizing over the factor, has a closed-form expression as a multivariate normal PDF:

$$\begin{aligned} f(\mathbf{y}_i | c_i = k) &= \\ &= \frac{1}{(2\pi)^{J/2} |\boldsymbol{\Sigma}^{(k)}|^{1/2}} \exp \left\{ -\frac{1}{2} [(\mathbf{y}_i - \boldsymbol{\mu}^{(k)})' (\boldsymbol{\Sigma}^{(k)})^{-1} (\mathbf{y}_i - \boldsymbol{\mu}^{(k)})] \right\}. \end{aligned} \quad (38)$$

A prototypic element of the model-implied mean vector $\boldsymbol{\mu}^{(k)}$ is $\mu_j^{(k)} = v_j^{(k)} + \lambda_j^{(k)}\alpha^{(k)}$ and prototypic elements of the covariance matrix $\boldsymbol{\Sigma}^{(k)}$ are $\sigma_{jj}^{(k)} = \lambda_j^{2(k)}\psi^{(k)} + \theta_{\varepsilon_j}^{(k)}$ and $\sigma_{jj'}^{(k)} = \lambda_j^{(k)}\psi^{(k)}\lambda_{j'}^{(k)}$ where $j \neq j'$.

GBT → Growth Mixture Model

To convert the GBT model into a growth mixture model (e.g., Muthén & Shedden, 1999), one or all of the fixed growth coefficients within class are replaced with continuously distributed growth factors within class. Here, we do so for the intercept only.

$$\begin{aligned} y_{ij} &= \eta_{0i} + \eta_{1i}time_{ij} + \eta_{2i}time_{ij}^2 + \varepsilon_{ij} \\ \eta_{0i} &= \alpha_0^{(k)} + \zeta_{0i} \\ \eta_{1i} &= \alpha_1^{(k)} \\ \eta_{2i} &= \alpha_2^{(k)}. \end{aligned} \tag{39}$$

Class-specific growth factor means are $\alpha_0^{(k)}$, $\alpha_1^{(k)}$, and $\alpha_2^{(k)}$ for the intercept, linear slope, and quadratic slope. Within class, systematic individual deviations in intercepts, ζ_{0i} , and residuals, ε_{ij} , are distributed as in Equations (36) and (37). Given the relaxation of local independence, the within-class PDF, $f(\mathbf{y}_i | c_i = k)$, is Equation (38), as in the factor mixture—with the exception that outcome intercepts ($v_j^{(k)}$) are now fixed to 0 and fixed time scores take the place of estimated factor loadings.

LCA → Categorical-Item Factor Mixture Model

To relax the local independence assumption for the LCA (e.g., Lubke & Neale, 2008; Rost, 1990) we can also allow the probability of endorsing an outcome to depend on a latent factor η_i with loading $\lambda_j^{(k)}$.

$$\begin{aligned} \pi_j^{(k)} &= 1 / (1 + \exp(\tau_j^{(k)} + \lambda_j^{(k)}\eta_i)) \\ \eta_i &= \alpha^{(k)} + \zeta_i. \end{aligned} \tag{40}$$

The individual deviation ζ_i from the factor mean is distributed as in Equation (36). For identification we may, for instance, fix $\alpha^{(k)} = 0$ and estimate all J thresholds and fix $\psi^{(k)} = 1$ and estimate all J factor loadings. Because local

independence is relaxed, Equation (18) no longer holds. Further, the within-class PMF for y_i does not have a closed form solution; obtaining it requires integrating over the factor:

$$p(y_i | c_i = k) = \int \left(\prod_{j=1}^J p(y_{ij} | c_i = k, \eta) f(\eta) \right) d\eta. \quad (41)$$

This involves numerical integration (one dimension per factor) and is computationally intensive compared to fitting hybrid mixtures with normal outcomes within class.

LTA \rightarrow LTA With Categorical-Item Factor Measurement Models

The approach to relaxing local independence from the LCA is used *per timepoint* in the LTA (Cho et al., 2010; Nylund, 2007). Equation (24) can be expanded to depend on η_{1i} at Time 1 and Equation (25) can be expanded to depend on η_{2i} at Time 2. These factors may have means $\alpha_1^{(k)} = \alpha_2^{(m)} = 0$ and normally distributed individual deviations $\begin{bmatrix} \zeta_{1i} \\ \zeta_{2i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_{11}^{(k)} & \\ & \psi_{12} \psi_{22}^{(m)} \end{bmatrix} \right)$ with variances usually fixed to 1. Often covariances (here, ψ_{12}) are fixed to 0 so all across-time associations are conveyed through transition probabilities, from regressing c_{i2} on c_{i1} . The within-sequence PMF in Equation (33) becomes

$$p(\mathbf{y}_i | c_{i1} = k, c_{i2} = m) = \quad (42)$$

$$\iint \left(\prod_{j=1}^J p(y_{ij1} | c_{i1} = k, \eta_1) \prod_{j=1}^J p(y_{ij2} | c_{i2} = m, \eta_2) f(\eta_1) f(\eta_2) \right) d\eta_1 d\eta_2$$

thus involving integration over two dimensions (one for each factor per timepoint).

This section highlighted similarities in how prototypical mixtures LPA, LCA, GBT, and LTA are extended to hybrid versions to relax local independence. In these hybrid mixtures, factor(s) are sometimes added within class simply to parsimoniously accommodate atheoretic residual associations. Other times, “direct” interpretations of hybrid mixtures involve theoretically meaningful within-class factor structures, and measurement invariance of these factor structures can be tested across classes/states (see Lubke & Muthén, 2005; Lubke & Neale, 2008). Particular issues arise in hybrid mixtures concerning whether to impose constraints on the factor structure within class before or after choosing the number of classes (see Bauer & Curran, 2003, 2004). Finally, hybrid mixtures may

be further extended to allow structural relations among latent and/or observed variables within class (Arminger & Stein, 1997; Dolan, 2009; Dolan & van der Maas, 1998; Muthén, 2002).

EXTENSION: PARALLEL-PROCESS MIXTURE MODELS

Whereas the first approach we discussed for relaxing local independence, hybrid mixtures, involved an extension of the within-class model, the second approach, used in parallel-process mixtures, involves an extension of the between-class model. Parallel-process mixtures are used when interest lies in relating latent class membership on more than one construct over time. This extension has been presented differently for change processes that are continuous (joint GBT models; Nagin & Tremblay, 2001) versus stage sequential (associative LTA models; Bray, Lanza, & Collins, 2010; Flaherty, 2008; see relatedly Vermunt, 2010b). However, here these extensions are shown to be closely related.

Denote the process for one construct the y -process and for the other the z -process. Parallel-process GBT and LTA account for dependency of outcomes *across process* by regressing classes/states for one process on classes/states of the other process. The specification of parallel-process mixtures can be summarized in terms of a between-, within-, and combined-model, as done earlier. The between-model is a joint probability of membership in classes/states from both processes, calculated using the *multiplicative rule for dependent events* in Equation (3). Note that in the between-model, there are alternative ways to express the joint probability depending on the manner in which z -process classes/states are regressed on y -process classes/states, or vice versa. One such alternative is shown here, which can be used to solve for probabilities of interest from other alternatives (see Appendix), so respecification is not necessary. In the within-model, local independence of outcomes *across process* is relaxed in the following respect. For GBT, local independence of outcomes across process now requires that responses be independent within a particular *combination* of z -process class and y -process class membership. Likewise, for LTA, local independence of responses across process is now required within a particular *combination* of z -process sequence and y -process sequence membership. The combined-model (i.e., marginal density of y 's and z 's) can be written to highlight its construction using the *law of total probability* in Equation (1), as shown further here.

Parallel-Process GBT Model

For the y -process, denote the outcome vector \mathbf{y}_i and latent classification variable c_i^y with classes $k = 1 \dots K$; corresponding quantities for the z -process are \mathbf{z}_i

and c_i^z , with classes $q = 1 \dots Q$. The marginal PDF (combined model) is

$$f(\mathbf{y}_i, \mathbf{z}_i) = \sum_{q=1}^Q \sum_{k=1}^K p(c_i^z = q) p(c_i^y = k | c_i^z = q) f(\mathbf{y}_i | c_i^y = k) f(\mathbf{z}_i | c_i^z = q). \quad (43)$$

Terms in Equation (43) can be defined using previously presented equations as follows (and are written out in full in the online appendix). The between-model consists of the probability of z -process class membership, $p(c_i^z = q)$, computed as in Equation (8), and the conditional probability of y -process class membership given z -process class membership, $p(c_i^y = k | c_i^z = q)$. The latter is computed using the multinomial logit specification as in Equation (30), only now $Q - 1$ dummy predictors are used to represent latent states in the z -process. The within-model consists of the joint PDFs of \mathbf{y}_i given c_i^y membership, $f(\mathbf{y}_i | c_i^y = k)$, and \mathbf{z}_i given c_i^z membership, $f(\mathbf{z}_i | c_i^z = q)$. Each is computed as in Equation (13) with parameters defined in Equations (21) and (23). Finally, given that the parallel-process GBT imposes local independence of y -outcomes within class k , z -outcomes within class q , and all outcomes within each k, q class pair, we can rewrite the within-model as $f(\mathbf{y}_i, \mathbf{z}_i | c_i^y = k, c_i^z = q) = f(\mathbf{y}_i | c_i^y = k) f(\mathbf{z}_i | c_i^z = q)$. Using the *multiplication rule for dependent events*, we can also rewrite the between-model. Doing so, the combined model in Equation (43) becomes a sum of within k, q class pair densities weighted by the probability of each k, q pair, following the *law of total probability*:

$$f(\mathbf{y}_i, \mathbf{z}_i) = \sum_{q=1}^Q \sum_{k=1}^K p(c_i^y = k, c_i^z = q) f(\mathbf{y}_i, \mathbf{z}_i | c_i^y = k, c_i^z = q). \quad (44)$$

Parallel-Process LTA

For the y -process, denote response pattern for person i as \mathbf{y}_i , Time 1 latent classification variable as c_{i1}^y with states $k = 1 \dots K$, and Time 2 latent classification variable as c_{i2}^y with states $m = 1 \dots M$. Corresponding quantities for the z -process are \mathbf{z}_i ; c_{i1}^z , with states $q = 1 \dots Q$; and c_{i2}^z , with states $s = 1 \dots S$. The marginal PMF (combined model) for a $T = 2$ parallel-process LTA is

$$f(\mathbf{y}_i, \mathbf{z}_i) = \sum_{s=1}^S \sum_{m=1}^M \sum_{q=1}^Q \sum_{k=1}^K p(c_{i1}^z = q) p(c_{i1}^y = k | c_{i1}^z = q) \quad (45)$$

$$p(c_{i2}^z = s | c_{i1}^y = k, c_{i1}^z = q) p(c_{i2}^y = m | c_{i1}^z = q, c_{i1}^y = k, c_{i2}^z = s)$$

$$p(\mathbf{y}_{i1} | c_{i1}^y = k) p(\mathbf{y}_{i2} | c_{i2}^y = m) p(\mathbf{z}_{i1} | c_{i1}^z = q) p(\mathbf{z}_{i2} | c_{i2}^z = s).$$

Terms in Equation (45) can be defined using previously presented equations, as follows (and are written in full in the online appendix). The between-model consists of the following four terms. The marginal probability of Time 1 z -process state membership, $p(c_{i1}^z = q)$, is computed as in Equation (8). The conditional probability of Time 1 y -process state given Time 1 z -process state, $p(c_{i1}^y = k | c_{i1}^z = q)$, is specified using a multinomial logit as in Equation (30) with $Q - 1$ dummy variable predictors used to represent latent states in the z -process at Time 1. The conditional probability of Time 2 z -process state given Time 1 y and z state, $p(c_{i2}^z = s | c_{i1}^y = k, c_{i1}^z = q)$, also is specified as in Equation (30) but now with $K - 1$ dummy predictors representing latent states in the y -process at Time 1 and $Q - 1$ representing latent states in the z -process at Time 1. Finally, the conditional probability of Time 2 y -process state given Time 1 y - and z -process states and Time 2 z -process state, $p(c_{i2}^y = m | c_{i1}^y = k, c_{i1}^z = s)$, also uses Equation (30). But now there are $K - 1$ dummy predictors for y -process states at Time 1, $Q - 1$ for z -process states at Time 1, and $S - 1$ for z -process states at Time 2. The within-model consists of the Time 1 PMFs of \mathbf{y}_{i1} given c_{i1}^y membership, $p(\mathbf{y}_{i1} | c_{i1}^y = k)$, and \mathbf{z}_{i1} given c_{i1}^z membership, $p(\mathbf{z}_{i1} | c_{i1}^z = q)$, specified as in Equation (28). The within-model also includes Time 2 PMFs of \mathbf{y}_{i2} given c_{i2}^y membership, $p(\mathbf{y}_{i2} | c_{i2}^y = m)$, and \mathbf{z}_{i2} given c_{i2}^z , $p(\mathbf{z}_{i2} | c_{i2}^z = s)$, that are specified as in Equation (29).

Finally, similar to parallel-process GBT, because local independence is imposed here within each state k , m , q , and s ; within each sequence km and qs ; and within each sequence-pair km , qs , we can rewrite the within-model as $p(\mathbf{y}_i, \mathbf{z}_i | c_{i1}^y = k, c_{i2}^y = m, c_{i1}^z = q, c_{i2}^z = s) = p(\mathbf{y}_{i1} | c_{i1}^y = k) p(\mathbf{y}_{i2} | c_{i2}^y = m) p(\mathbf{z}_{i1} | c_{i1}^z = q) p(\mathbf{z}_{i2} | c_{i2}^z = s)$. We again rewrite the between-model using the *multiplication rule for dependent events*. Doing so, the combined model becomes a sum of within km , qs sequence densities weighted by the probability of each km , qs sequence pair, following the *law of total probability*:

$$f(\mathbf{y}_i, \mathbf{z}_i) = \sum_{s=1}^S \sum_{m=1}^M \sum_{q=1}^Q \sum_{k=1}^K p(c_{i1}^y = k, c_{i2}^y = m, c_{i1}^z = q, c_{i2}^z = s) \quad (46)$$

$$p(\mathbf{y}_i, \mathbf{z}_i | c_{i1}^y = k, c_{i2}^y = m, c_{i1}^z = q, c_{i2}^z = s).$$

In sum, modeling architecture and assumptions used to link together longitudinal processes specific to subgroups of persons are very similar in the parallel-process GBT versus LTA. The main substantive difference lies in whether interest is in linking subgroup-specific continuous trajectory or state-sequential patterns of change for two constructs. All across-process associations are accounted for by a mixing together of subgroups with particular k , q class combinations in the parallel-process GBT and are accounted for by a mixing together of subgroups with particular km , qs sequence combinations in the parallel-

process LTA. There can be different numbers of trajectories with different functional forms across processes in the GBT framework, just as there can be different numbers of sequences with different transition probabilities across processes in the LTA framework. In each parallel-process mixture, interest often focuses on particular probabilities in the between-model: marginal probabilities of class, state, or sequence membership in one process; conditional probabilities of class/state/sequence membership in one process given class/state/sequence membership in the other process; and joint probabilities of class/state/sequence membership in both processes. Also, just for the parallel-process LTA, interest may lie in conditional probabilities of membership in a current state on one process given previous states on both processes and current state on the other process. Although not all of these probabilities are represented in Equations (43)–(46), others are solvable from them, as shown in the Appendix.

SUMMARY OF EXTENSION MODELS AND THEIR RELATIONSHIP

As summarized in Table 2, the first extension section on hybrid mixtures covered relaxing local independence of outcomes by adding factor(s) within class (state). The second extension section, on parallel-process mixtures, covered relaxing local independence of outcomes from two different constructs by expanding the between-class (-state) model to create dependency among classification variables for each construct.

Hence, at this point we are positioned to consider the combination of both approaches for relaxing local independence—in the within-model and in the between-model. Here, we do so by considering a hybrid parallel-process LTA (for a hybrid parallel-process GBT, see, e.g., Muthén, 2001). To formulate a hybrid parallel-process LTA, one possibility is to include factor(s) within-state (akin to Equation (42) inside a process of a parallel-process LTA). However, an integrative perspective on mixture models can suggest combining features in ways not necessarily transparent when focusing on one model's literature. In particular, there is substantive motivation for accommodating a different kind of dependency in longitudinal studies of psychiatric syndromes, for instance. Many syndromes share symptoms (including oppositional defiant and conduct disorder syndromes, generalized anxiety and depression syndromes, and oppositional defiant and depression syndromes). When studying longitudinal associations among psychiatric syndromes, overlapping symptoms can be a methodological confound if unaccounted for (see Angold, Costello, & Erkanli, 1999; Sterba et al., 2010). Overlapping symptoms could induce *across-process* local dependence among outcomes in a conventional parallel-process LTA. This in turn can inflate or alter across-process conditional probabilities of

state (or sequence) membership—thus interfering with their interpretability. To accommodate across-process dependency induced by overlapping symptom(s), we can expand the parallel-process LTA to include factor(s) within-time-but-across-process. This will further relax local independence among outcomes (from either process) within a pair of km , qs sequences. The latter integrative model is illustrated in a longitudinal analysis of psychiatric symptoms in the next section.

EMPIRICAL EXAMPLE

This example concerns the longitudinal association between oppositional defiant and depressive syndromes in early childhood. Longitudinal comorbidity and temporal sequencing of these syndromes are of considerable interest in clinical psychology (e.g., Boylan, Vaillancourt, Boyle, & Szatmari, 2007; Lavigne et al., 2001). An overlapping symptom arises when considering these syndromes in children. Irritability is a symptom of childhood depression, whereas a considerably overlapping oppositional defiant symptom is temper tantrums. Researchers may intend to model longitudinal associations among depression and oppositional defiant syndromes using, for instance, parallel-process LTA. They would encounter the obstacle that the local independence assumption imposed within each pair of sequences, km , qs (one pair from each process), is likely violated due to the overlapping symptom. A parallel-process LTA that allows for local dependence by including an across-process factor in the within-state model at each timepoint is depicted in Figure 4. One outcome from each process (irritability and temper tantrums) loads on the factor. Here, each factor's loadings are constrained equal across state for parsimony, and factors do not correlate across time, although the model does not require this. Also, for identification, one loading per factor as well as the factor variances are fixed to 1.

Outcomes were binary maternal-reported symptoms (1 = *sometimes or always*; 0 = *never*) for $N = 1,113$ children at first and fourth grades (see National Institute of Child Health and Human Development Early Child Care Research Network [NICHD-ECCRN], 2004, for study details). The $J = 4$ depression symptoms were feeling sad/depressed, slow moving, concentration problems, and irritability. The $J = 4$ oppositional defiant symptoms were temper tantrums, disobedience, whining, and lack of guilt/responsibility for misbehavior. First, single-process LTAs were fit. Using BIC, the best fitting oppositional defiant LTA had $K = M = 2$, and depression LTA (discussed earlier) also had $Q = S = 2$. The state-specific symptom endorsement probabilities per timepoint are plotted in Figure 3 for depression and Figure 5 for oppositional defiant behavior; transition probability matrices are shown at the bottom of each figure. Imposing measurement invariance on each symptom threshold within state across time

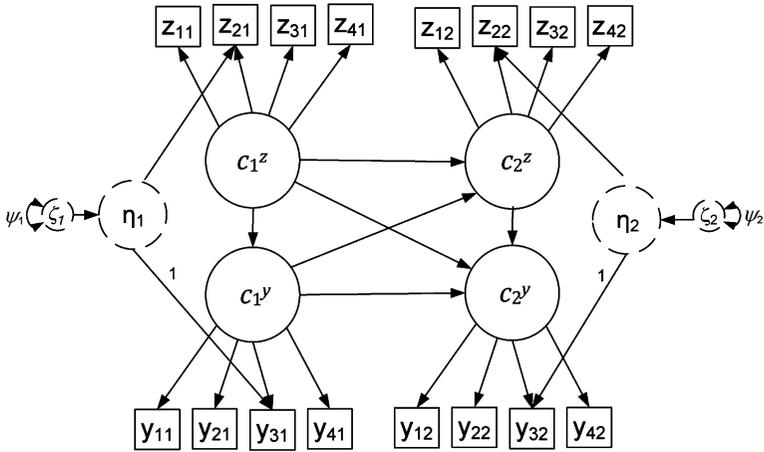
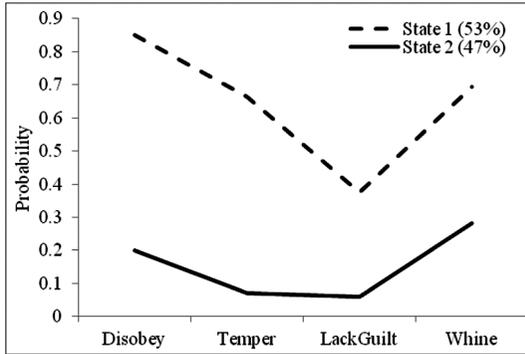


FIGURE 4 Path diagram for the hybrid parallel-process LTA empirical example model. *Note.* z -process is oppositional defiant behavior, with 4 measured symptoms (boxes) at 2 timepoints; y -process is depression, with 4 measured symptoms (boxes) at 2 timepoints. “Overlapping” symptoms across syndromes at each timepoint are z_2 (temper tantrums) and y_3 (irritability). Solid circles denoted with c correspond with a categorical latent variable. Dashed circles η correspond with a continuous latent factor or deviation. Other notation was described in the text. LTA = latent transition analysis.

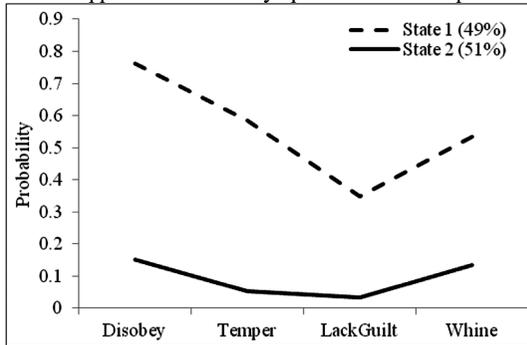
did not decrease fit according to BIC for the depression or oppositional defiant behavior LTAs. This indicates that states retain their interpretation over time. The measurement invariant LTAs for each process were then combined into a parallel-process LTA. Parallel-process LTAs were fit either allowing versus not allowing for local dependence between overlapping symptoms. The model allowing for symptom local dependence across process (BIC = 16,123.80, $df = 28$) fit better than the model requiring symptom local independence across process (BIC = 16,138.41, $df = 26$). In the locally dependent model, the estimated factor loading was significant at each timepoint ($\lambda_1 = 1.42$, $p < .01$; $\lambda_2 = 1.54$, $p < .01$).

Whereas many kinds of within- and across-process conditional and joint probabilities (calculable using formulas in the Appendix) may be of substantive interest to report in this example, we focus on one kind in Table 3: the conditional probability of sequence membership in one process given sequence membership in the other process. Estimated multinomial coefficients that can be used to solve for other probabilities of interest are provided in the online appendix. Note that in Table 3, for both processes at both times, State 1 corresponds with elevated problems and State 2 corresponds with low risk. Conditional probabilities in Table 3 are for the locally dependent parallel-process LTA from Figure 4. For the local independence counterpart model, these probabilities differed often by

Time 1 oppositional defiant symptom endorsement probabilities:



Time 2 oppositional defiant symptom endorsement probabilities:



State-to-state transition probabilities from Time 1 to Time 2:

	c_{i2}	
	1	2
c_{i1}	1	.90 .10
	2	.02 .98

FIGURE 5 Empirical example results: latent transition analysis of oppositional defiant symptoms.

.01–.02 but up to .07. The predominant pattern in Table 3 is that persons have the highest probability of sequence membership in a given process for the same sequence they followed in the other process. As an example, persons in an elevated-problems state at both timepoints for oppositional defiant behavior (oppositional defiant sequence 1,1) have the highest probability (.83) of being in an elevated-problems state at both timepoints for depression (depression sequence 1,1). One clinically interesting exception to this pattern is that persons transitioning from low to elevated depression over time (depression sequence 2,1) are most likely

TABLE 3
Empirical Example Parallel-Process LTA With Across-Process Local Dependence for Irritability/Temper Symptom: Conditional Probability of Sequence Membership on One Process Given Sequence Membership on the Other Process

		c_1^z, c_2^z				c_1^y, c_2^y				
		1,1	2,1	1,2	2,2					
c_1^y, c_2^y	1,1	.87	.00	.10	.03	1,1	.83	.07	.00	.10
	2,1	.54	.22	.00	.24	2,1	.06	.94	.00	.00
	1,2	.00	.00	1.00	.00	1,2	.28	.00	.37	.34
	2,2	.09	.00	.10	.82	2,2	.03	.03	.00	.94

Note. y -process is oppositional defiant behavior and z -process is depression. Conditional probabilities within the left table describe the probability of being in a z -process sequence given y -process sequence membership. Conditional probabilities within the right table describe the probability of being in a y -process sequence given z -process sequence membership. For both processes, State 1 is elevated problems and State 2 is low risk. For a given process, sequences are denoted as, for example, (1,1), meaning membership in State 1 at Time 1 and State 1 at Time 2. LTA = Latent Transition Analysis.

(.54 probability) to *stably* be in elevated-problems oppositional defiant states over time. That is, perhaps persistence of oppositional defiant problems is a risk factor for onset of depressive symptoms. The reverse phenomenon is not prominent; those transitioning from low to elevated oppositional defiant behavior are unlikely to be stably in an elevated depressive state (.06 probability).

To our knowledge, previous studies have not relaxed local independence *across process* in mixture models. Yet the integrated model is clearly relevant to studying longitudinal comorbidity.

DISCUSSION

Interest in mixture models continues to increase. However, even for psychologists acquainted with one or a pair of mixture models, there is currently a limited foundation for understanding shared features uniting mixture models generally. This article focused on concretely bridging understanding among five often-applied mixture models by presenting the specification of each in a unified format with incremental complexity while highlighting their shared reliance on familiar probability laws (summarized in Table 2). For instance, because the traditional presentation format for LTA did not strongly emphasize its relation to mixture models in general, it was reexpressed in order to make this connection. The integrative presentation of the five mixtures clarified common assumptions and aspects of interpretation. Additionally, these prototypic mixtures were extended

in two alternative ways—to form hybrid mixtures and parallel-process mixtures. Both extensions accomplished the same goal—relaxing the key assumption of local independence—but did so using different approaches. Popular implementations of each approach were seen to have similar construction and interpretation. Finally, an example illustrated how an integrative perspective on mixtures affords insight into new ways to combine both approaches. In the remainder of this section, other applications of these two approaches for relaxing local independence in mixtures are described. We also briefly mention other connections among mixture models that are deserving of increased emphasis in the future.

Additional Mixture Models Using the Two Approaches for Relaxing Local Independence

Our review of two common approaches for relaxing local independence in mixtures provides a foundation for understanding more advanced models, such as multivariate multilevel mixtures (e.g., Henry & Muthén, 2010; Vermunt, 2008). In such mixture models, there are similarly two options for relaxing local independence within Level 1 latent classes to account for the fact that outcome scores from persons sharing a cluster (e.g., a school) may be more similar to each other, even after accounting for Level 1 class membership. For instance, per the first approach discussed earlier, random effect(s) (i.e., factor(s)) can be added within Level 1 class. Or, per the second approach discussed earlier, another latent classification variable can be added—here, a Level 2 classification variable—and Level 1 latent class membership can be made dependent on Level 2 latent class membership. This implies local independence of outcomes per person given a *combination* of Level 1 and Level 2 class membership (Lukociene, Varriale, & Vermunt, 2010). It is also possible to use these two approaches in one model.

Addition of Observed Covariates

Modal class assignments—calculated from posterior probabilities of class membership, as described earlier—are commonly used in subsequent analyses predicting class membership. However, this procedure ignores measurement error involved in modal assignments. A second option is to directly add observed predictors of class membership to the between-class (-state) model to explain who is more likely to be in which class (also see the measurement error correction proposed by Vermunt, 2010a). For UFNM, LPA, LCA, and GBT this second option entails adding $K - 1$ multinomial slopes for each predictor to Equation (8) (e.g., Muthén, 2002). For LTA, the researcher may choose to allow the predictor to predict initial latent state, and/or predict subsequent state controlling for initial

state, in Equation (30) (see Muthén & Asparouhov, 2011). Mixture models also allow the possibility of adding observed predictors within class (state) in the manner that $time_{ij}$ was incorporated as a covariate in Equation (22). The slope of such predictors could be allowed to differ across class (state).

Addition of Within-Class and/or Between-Class Constraints

For the prototypical mixtures considered here, constraints on parameters within class (sequence) can be useful in evaluating more specific or parsimonious patterns. For example, in LTA it is common to compare unrestricted versus forward-only (or backward-only) versus no state-to-state change by imposing constraints¹⁴ on multinomial coefficients in Equation (30) (see Kaplan, 2008). Likewise, in GBT it is common to compare flat (no-change) trajectories to linear or higher order change among some or all classes (by testing zero-constraint(s) on slope(s) of time within class in Equation (22); Nagin, 2005). For each parallel-process mixture, researchers may be interested in testing whether certain conditional probabilities of class/state membership in one process given class/state membership in the other process are 0 (e.g., Bray et al., 2010). This may be done by testing corresponding constraints on the multinomial coefficients used to calculate these probabilities. Also, constraints placed within class in modified¹⁵ growth mixture models can test for regime switching—wherein persons are allowed to toggle between different continuous curves (regimes) at each timepoint (e.g., Dolan et al., 2005). The number of classes is fixed to the number of regime *sequences*, and local independence is imposed within sequence.

Limitations

Two common within-class outcome distributions were illustrated here, although mixtures have been fit with a variety of other within-class distributions (e.g., McLachlan & Peel, 2000). Further, space limitations prevented considering other important mixture models (e.g., SEM mixtures or survival analysis mixtures). In principle, virtually any existing whole-population model could be converted to a mixture by specifying it as a within-class model and allowing certain parameters

¹⁴For instance, in a $K = M = 2$ LTA, testing forward change (below-diagonal elements of the transition probability matrix = 0) requires fixing α_1 to a large negative number. Testing no-change (off-diagonal elements of transition probability matrix = 0) requires *also* fixing β_{11} to a very large positive number.

¹⁵Within class, the number of growth coefficients is $(1 + b) \times r$ (where b is the polynomial curve degree and r is the number of regimes). At timepoints where a particular regime goes off-line, its growth coefficients are fixed to 0 (for details see Dolan, Schmittmann, Lubke, & Neale, 2005).

to differ across class. Unifying principles in Table 2 would generally apply to other examples.

Conclusions

By supplying pedagogical detail on linkages among classic mixture models and recent extensions, this article bridges historically diffuse literatures on mixture modeling. This article clarified how features of existing models build on each other and can be combined in new ways, which can serve as a foundation for conceptualizing new mixture modeling developments.

ACKNOWLEDGMENTS

I thank the Editors, the Reviewers, Sun-Joo Cho, and Andrew Greer for helpful comments.

REFERENCES

- Abar, B., & Loken, E. (2012). Consequences of fitting nonidentified latent class models. *Structural Equation Modeling, 19*, 1–15. doi: 10.1080/10705511.2012.634701
- Angold, A., Costello, J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40*, 57–87. doi: 10.1111/1469-7610.00424
- Arminger, G., & Stein, P. (1997). Finite mixtures of covariance structure models with regressors. *Sociological Methods and Research, 26*, 148–182. doi: 10.1177/0049124197026002002
- Asparouhov, T., & Muthén, B. (2008). Multilevel mixture models. In G. Hancock & K. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 27–52). Charlotte, NC: Information Age.
- Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). London, UK: Arnold.
- Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling, 12*, 513–535. doi: 10.1207/s15328007sem1204_1
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8*, 338–363. doi: 10.1037/1082-989X.8.3.338
- Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*, 3–29. doi: 10.1037/1082-989X.9.1.3
- Biemer, P. (2011). *Latent class analysis of survey error*. Hoboken, NJ: Wiley.
- Boylan, K., Vaillancourt, T., Boyle, M., & Szatmari, P. (2007). Comorbidity of internalizing disorders in children with oppositional defiant disorder. *European Child & Adolescent Psychiatry, 16*, 484–494. doi: 10.1007/s00787-007-0624-1
- Bray, B., Lanza, S., & Collins, L. (2010). Modeling relations among discrete developmental processes: A general approach to associative latent transition analysis. *Structural Equation Modeling, 17*, 541–569. doi: 10.1080/10705511.2010.510043

- Campbell, S., Matestic, P., von Stauffenberg, C., Mohan, R., & Kirchner, T. (2007). Maternal depressive symptoms, maternal sensitivity, and children's functioning at school entry. *Developmental Psychology, 43*, 1202–1215. doi: 10.1037/0012-1649.43.5.1202
- Cho, S.-J., Cohen, A. S., Kim, S.-H., & Bottge, B. (2010). Latent transition analysis with a mixture IRT measurement model. *Applied Psychological Measurement, 34*, 583–604. doi: 10.1177/0146621610362978
- Clogg, C. (1995). Latent class models. In G. Arminger, C. Clogg, & M. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311–359). New York, NY: Plenum. doi: 10.1007/978-1-4899-1292-3_6
- Collins, L., & Flaherty, B. (2002). Latent class models for longitudinal data. In J. Hagenaars & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 287–303). West Nyack, NY: Cambridge. doi: 10.1017/CBO9780511499531.011
- Collins, L., & Lanza, S. (2010). *Latent class and latent transition analysis*. Hoboken, NJ: Wiley.
- Collins, L., & Wugalter, S. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research, 27*, 131–157. doi: 10.1207/s15327906mbr2701_8
- Dolan, C. (2009). Structural equation mixture modeling. In R. Millsap & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 568–591). London, UK: Sage. doi: 10.4135/9780857020994.n23
- Dolan, C., Schmittmann, V., Lubke, G., & Neale, M. (2005). Regime switching in the latent growth curve mixture model. *Structural Equation Modeling, 12*, 94–119. doi: 10.1207/s15328007sem1201_5
- Dolan, C., & van der Maas, H. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika, 63*, 227–253.
- Flaherty, B. P. (2008). Testing the degree of cross-sectional and longitudinal dependence between two discrete dynamic processes. *Developmental Psychology, 44*, 468–480. doi: 10.1037/0012-1649.44.2.468
- Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. New York, NY: Springer.
- Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika, 24*, 229–252. doi: 10.1007/BF02289845
- Goodman, L. A. (2007). On the assignment of individuals to latent classes. *Sociological Methodology, 37*, 1–22. doi: 10.1111/j.1467-9531.2007.00184.x doi: 10.1080/00273171.2010.531230
- Grimm, K., & Ram, N. (2009). A second-order growth mixture model for developmental research. *Research in Human Development, 6*, 121–143. doi: 10.1080/15427600902911221
- Grimm, K., Ram, N., & Estabrook, R. (2010). Nonlinear structured growth mixture models in Mplus and OpenMx. *Multivariate Behavioral Research, 45*, 887–909.
- Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage. doi: 10.1080/15427600902911221
- Henry, K., & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling, 17*, 193–215. doi: 10.1080/10705511003659342
- Humphreys, K., & Janson, H. (2000). Latent transition analysis with covariates, nonresponse, summary statistics and diagnostics. *Multivariate Behavioral Research, 35*, 89–118. doi: 10.1207/S15327906MBR3501_4
- Jedidi, K., Jagpal, H., & Desarbo, W. (1997). Finite mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16*, 39–59. doi: 10.1287/mksc.16.1.39
- Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology, 44*, 457–467. doi: 10.1037/0012-1649.44.2.457

- Kreuter, F., & Muthén, B. (2008). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. *Journal of Quantitative Criminology*, *24*, 1–31. doi: 10.1007/s10940-007-9036-0
- Langeheine, R. (1994). Latent variable Markov models. In A. von Eye & C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 373–395). Beverly Hills, CA: Sage.
- Lavigne, J., Cicchetti, C., Gibbons, R., Binns, H., Larsen, L., & DeVito, C. (2001). Oppositional defiant disorder with onset in the preschool years: Longitudinal stability and pathways to other disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 1393–1400. doi: 10.1097/00004583-200112000-00009
- Lazarsfeld, P., & Henry, N. (1968). *Latent structure analysis*. New York, NY: Houghton Mifflin. doi: 10.2307/2785875
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.
- Lubke, G., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21–39. doi: 10.1037/1082-989X.10.1.21
- Lubke, G., & Neale, M. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, *41*, 499–532. doi: 10.1207/s15327906mbr4104_4
- Lubke, G., & Neale, M. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. *Multivariate Behavioral Research*, *43*, 592–620. doi: 10.1080/00273170802490673
- Lubke, G., & Tueller, S. (2010). Latent class detection and class assignment: A comparison of the MAXEIG taxometric procedure and factor mixture modeling approaches. *Structural Equation Modeling*, *17*, 605–628. doi: 10.1080/10705511.2010.510050
- Lukociene, O., Varriale, R., & Vermunt, J. (2010). The simultaneous decision about the number of lower- and higher-level classes in multilevel latent class analysis. *Sociological Methodology*, *40*, 247–283. doi: 10.1111/j.1467-9531.2010.01231.x
- Magidson, J., & Vermunt, J. (2004). Latent class models. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 175–198). Thousand Oaks, CA: Sage. doi: 10.4135/9781412986311.n10
- Mazza, J. J., Fleming, C. B., Abbott, R. D., Haggerty, K. P., & Catalano, R. F. (2010). Identifying trajectories of adolescents' depressive phenomena: An examination of early risk factors. *Journal of Youth and Adolescence*, *39*, 579–593. doi: 10.1007/s10964-009-9406-z
- McCutcheon, A. (2002). Basic concepts and procedures in single- and multiple-group latent class analysis. In A. McCutcheon (Ed.), *Advances in latent class models* (pp. 56–85). West Nyack, NY: Cambridge. doi: 10.1017/CBO9780511499531.003
- McDonald, R. P. (1962). A note on the derivation of the general latent class model. *Psychometrika*, *27*, 203–206. doi: 10.1007/BF02289639
- McLachlan, G. (2011). Commentary on Steinley and Brusco (2011): Recommendations and cautions. *Psychological Methods*, *16*, 80–81. doi: 10.1037/a0021141
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.
- Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*, 195–215. doi: 10.1007/BF02295283
- Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Erlbaum.
- Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *20*, 81–117. doi: 10.2333/bhmk.29.81

- Muthén, B. O. (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. doi: 10.4135/9781412986311.n19
- Muthén, B. O. (2008). Latent variable hybrids: Overview of new and old models. In G. Hancock & K. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–26). Charlotte, NC: Information Age.
- Muthén, B. O., & Asparouhov, T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors, 31*, 1050–1066. doi: 10.1016/j.addbeh.2006.03.026
- Muthén, B. O., & Asparouhov, T. (2011). LTA in Mplus: Transition probabilities influenced by covariates. Webnote #13. Retrieved from www.statmodel.com
- Muthén, B. O., Asparouhov, T., & Rebollo, I. (2005). Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. *Twin Research and Human Genetics, 9*, 313–324. doi: 10.1375/183242706777591317
- Muthén, B. O., & Muthén, L. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882–891. doi: 10.1111/j.1530-0277.2000.tb02070.x
- Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463–469. doi: 10.1111/j.0006-341X.1999.00463.x
- Muthén, L. K., & Muthén, B. O. (1998–2013). *Mplus users guide* (7th ed.). Los Angeles, CA: Author.
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*, 139–157. doi: 10.1037//1082-989X.4.2.139
- Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.
- Nagin, D. S., & Odgers, C. (2010). Group-based trajectory modeling in clinical research. *Annual Reviews of Clinical Psychology, 6*, 109–138. doi: 10.1146/annurev.clinpsy.121208.131413
- Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods, 6*, 18–34. doi: 10.1037/1082-989X.6.1.18
- National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood: Predictors, correlates, and outcomes. *Monographs of the Society for Research in Child Development, 69* (pp. 26–41). doi: 10.1111/j.1540-5834.2004.00313.x
- Nylund, K. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization* (Doctoral dissertation, University of California, Los Angeles).
- Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte-Carlo simulation study. *Structural Equation Modeling, 14*, 535–569. doi: 10.1080/10705510701575396
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London, A, 185*, 71–110. doi: 10.1098/rsta.1894.0003
- Pickles, P., & Croudace, T. (2010). Latent mixture models for multivariate and longitudinal outcomes. *Statistical Methods in Medical Research, 19*, 271–289. doi: 10.1177/0962280209105016
- Reboussin, B. A., Reboussin, D. M., Liang, K., & Anthony, J. C. (1998). Latent transition modeling of progression of health risk behaviors. *Multivariate Behavioral Research, 33*, 457–478. doi: 10.1207/s15327906mbr3304_2
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271–282. doi: 10.1177/014662169001400305
- Schlattmann, P. (2009). *Medical applications of finite mixture models*. Berlin, Germany: Springer-Verlag. doi: 10.1007/978-3-540-68651-4

- Schmittmann, V., Dolan, C., van der Maas, H., & Neale, M. (2005). Discrete latent Markov models for normally distributed response data. *Multivariate Behavioral Research, 40*, 461–488. doi: 10.1207/s15327906mbr4004_4
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.
- Sterba, S. K., Baldasaro, R. E., & Bauer, D. J. (2012). Factors affecting the adequacy and preferability of semiparametric groups-based approximations of continuous growth trajectories. *Multivariate Behavioral Research, 47*, 590–634. doi: 10.1080/00273171.2012.692639
- Sterba, S. K., & Bauer, D. J. (2010). Matching method with theory in person-oriented developmental psychopathology research. *Development & Psychopathology, 22*, 239–254. doi: 10.1017/S0954579410000015
- Sterba, S. K., Copeland, W., Egger, H., Costello, J., Erkanli, A., & Angold, A. (2010). Longitudinal dimensionality of adolescent psychopathology: Testing the differentiation hypothesis. *Journal of Child Psychology and Psychiatry, 51*, 871–884. doi: 10.1111/j.1469-7610.2010.02234.x
- Tofighi, D., & Enders, C. (2007). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Greenwich, CT: Information Age.
- 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.
- Vermunt, J. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research, 17*, 33–51. doi: 10.1177/0962280207081238
- Vermunt, J. (2010a). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18*, 450–469. doi: 10.1093/pan/mpq025
- Vermunt, J. (2010b). Longitudinal research using mixture models. In K. Van Monfort, J. Oud, & A. Sattora (Eds.), *Longitudinal research with latent variables* (pp. 119–152). Heidelberg, Germany: Springer. doi: 10.1007/978-3-642-11760-2_4
- Vermunt, J. (2011). K-means may perform as well as mixture model clustering but may be much worse: Comment on Steinley and Brusco (2011). *Psychological Methods, 16*, 82–88. doi: 10.1037/a0020144
- Vermunt, J., Tran, B., & Magidson, J. (2006). Latent class models in longitudinal research. In S. Menard (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis* (pp. 373–385). Burlington, MA: Elsevier.
- Wedel, M., & DeSarbo, W. (2002). Mixture regression models. In A. McCutcheon (Ed.), *Advances in latent class models*. (pp. 366–382). West Nyack, NY: Cambridge. doi: 10.1017/CBO9780511499531.014
- Wolfe, J. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research, 5*, 329–350. doi: 10.1207/s15327906mbr0503_6
- Yung, Y.-F. (1997). Finite mixtures in confirmatory factor analysis models. *Psychometrika, 62*, 297–330. doi: 10.1007/BF02294554

APPENDIX

The Appendix reviews how probabilities appearing in the marginal density functions for latent transition analysis (LTA), parallel-process groups-based trajectory model (GBT), and parallel-process LTA—themselves a function of multinomial intercepts and slopes as described in the text—can be used to solve for other probabilities of interpretational interest. For each model, Equations (1)–(4) and

the definition of conditional probability are used in computing additional probabilities. Some mixture modeling software computes these automatically.

LTA. We adopt the following shorthand for this appendix section: $K = (c_{i1} = k)$. $M = (c_{i2} = m)$.

Already given in marginal model: $p(K)$ and $p(M|K)$. Also potentially of interest:

$$p(M) = \sum^K p(M|K)p(K) \quad p(K, M) = p(K)p(M|K).$$

Parallel-process GBT. Shorthand adopted for this appendix section: $K = (c_i^y = k)$. $Q = (c_i^z = q)$. Given in marginal model for 2 timepoints: $p(Q)$ and $p(K|Q)$. Also may be of interest:

$$p(K) = \sum^Q p(K|Q)p(Q) \quad p(K, Q) = p(Q)p(K|Q)$$

$$p(Q|K) = p(K, Q)/p(K)$$

Parallel-process LTA. Shorthand adopted for this appendix section: $Q = (c_{i1}^z = q)$. $K = (c_{i1}^y = k)$. $M = (c_{i2}^y = m)$. $S = (c_{i2}^z = s)$. Already given in marginal model for 2 timepoints: $p(Q)$, $p(K|Q)$, $p(S|K, Q)$ and $p(M|K, Q, S)$. Also potentially of interest:

Marginal:

$$p(K) = \sum^Q p(K|Q)p(Q)$$

$$p(M) = \sum^S \sum^Q \sum^K p(M|K, Q, S)p(K, Q, S)$$

$$p(S) = \sum^Q \sum^K p(S|K, Q)p(K, Q)$$

Conditional:

$$p(Q|K) = p(K, Q)/p(K)$$

$$p(S|M) = p(S, M)/p(M)$$

$$p(M|S) = \sum^Q \sum^K p(M|K, Q, S)p(K, Q|S)$$

$$p(M|K) = \sum^Q p(M|K, Q)p(Q|K)$$

Joint:

$$p(K, M) = p(M|K)p(K)$$

$$p(Q, S) = \sum^M \sum^K p(Q, S|K, M)p(K, M)$$

$$p(K, Q) = p(K|Q)p(Q)$$

$$p(S, M) = p(M|S)p(S)$$

$$p(K, Q, M, S) = p(Q)p(K|Q)p(S|K, Q)p(M|K, Q, S)$$

$$p(S|Q) = p(Q, S)/p(Q)$$

$$p(M|K, Q) = \sum^S p(M|K, Q, S)p(S|K, Q)$$

$$p(S|K, Q, M) = p(K, Q, M, S)/p(K, Q, M)$$

$$p(K, M|Q, S) = p(K, Q, M, S)/p(Q, S)$$

$$p(Q, S|K, M) = p(K, Q, M, S)/p(K, M)$$

$$p(S, M|K, Q) = p(K, Q, M, S)/p(K, Q)$$

Not usually of substantive interest but used earlier:

$$p(K, Q, M) = p(M|K, Q)p(K, Q) \quad p(K, Q|S) = p(S|K, Q)p(K, Q)/p(S)$$

$$p(K, Q, S) = p(S|K, Q)p(K, Q)$$