Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

Building Latent Class Growth Trees

Mattis van den Bergh & Jeroen K. Vermunt

To cite this article: Mattis van den Bergh & Jeroen K. Vermunt (2018) Building Latent Class Growth Trees, Structural Equation Modeling: A Multidisciplinary Journal, 25:3, 331-342, DOI: 10.1080/10705511.2017.1389610

To link to this article: https://doi.org/10.1080/10705511.2017.1389610

Structural Equation Modeling: A Multidisciplinary Journal, 25: 331–342, 2018
ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2017.1389610

Building Latent Class Growth Trees

Mattis van den Bergh and Jeroen K. Vermunt
Tilburg University

Researchers use latent class growth (LCG) analysis to detect meaningful subpopulations that display different growth curves. However, especially when the number of classes required to obtain a good fit is large, interpretation of the encountered class-specific curves might not be straightforward. To overcome this problem, we propose an alternative way of performing LCG analysis, which we call LCG tree (LCGT) modeling. For this purpose, a recursive partitioning procedure similar to divisive hierarchical cluster analysis is used: Classes are split until a certain criterion indicates that the fit does not improve. The advantage of the LCGT approach compared to the standard LCG approach is that it gives a clear insight into how the latent classes are formed and how solutions with different numbers of classes relate. The practical use of the approach is illustrated using applications on drug use during adolescence and mood regulation during the day.

Keywords: hierarchical clustering, latent class growth analysis, latent class growth trees, longitudinal data, mixture models

Longitudinal data are used by social scientists to study development of behaviors or other phenomena. The analysis will often be done with latent growth curve models (MacCallum & Austin, 2000), with the aim to assess interindividual differences in intraindividual change over time (Nesselroade, 1991). The typical growth model can be described as a multilevel model (Raudenbush & Bryk, 2002), in which the intercept and slopes of the time variables are allowed to vary across individuals. This heterogeneity is captured using random effects, which are continuous latent variables (Jung & Wickrama, 2008). This approach assumes that the growth trajectories of all individuals can be appropriately described by a single set of the growth parameters, and thus all individuals come from a single population. Growth mixture modeling relaxes this assumption by allowing for differences in growth parameters across unobserved subpopulations; that is, each latent class has a separate growth model. However, fully unrestricted growth mixture models are seldom used in practice, in part due to frequent estimation problems, as well as the preference for simpler, restricted models. The most widely used form of growth mixture modeling is latent class growth (LCG) analysis, whereby the variances and covariances of the growth factors within classes are fixed to zero (Jones, Nagin, & Roeder, 2001; Nagin & Land, 1993). This assumes that all individuals within a class follow the same trajectory and thus there is no residual heterogeneity within classes.

When an LCG model is applied, two key modeling decisions need to be made: the number of classes and the shape of the class-specific trajectories. In general, the decision on the number of classes is of more importance than the decision on the shape of the trajectory of each class as long as the shape is flexible enough (Nagin, 2005). Different fit statistics are available to handle the problem of model selection in LCG models, such as the Akaike's information criterion (AIC; Akaike, 1974), the Bayesian information criterion (BIC; Schwarz, 1978), the sample-size-adjusted BIC (Sclove, 1987), the Lo–Mendell–Rubin likelihood ratio test (Lo, Mendell, & Rubin, 2001), and the bootstrap likelihood ratio test (McLachlan & Peel, 2004). The benefits and limitations of these measures have been studied several times (e.g., Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2008). However, these indexes rarely point to the same best model. Grimm, Ram, and Estabrook (2017) therefore recommended considering all available fit information when selecting a model and supplementing this information

Correspondence should be addressed to Mattis van den Bergh, Methodology and Statistics, Tilburg University, Warandelaan 2, 5037 AB, Tilburg, The Netherlands. E-mail: m.vdnbergh@uvt.nl

with substantive knowledge of the phenomena being studied (see also Muthén, 2003). Although there is nothing wrong with such a procedure, in practice it is often perceived as being problematic, especially when the model is applied with a large data set; that is, when the number of time points or the number of subjects is large. One problem occurring in such situations is that the selected number of classes could be rather large (Francis, Elliott, & Weldon, 2016). This causes the class trajectories to pick up very specific aspects of the data, which might not be interesting for the research question at hand. Moreover, these specific trajectories are hard to interpret substantively and compare to each other. This, combined with the fact that usually one would select a different number of classes depending on the model selection criterion used (e.g., AIC or BIC), means that one might wish to inspect multiple solutions, as each of them could reveal specific relevant features in the data. However, it is fully unclear how solutions with different numbers of classes are connected, making it impossible to see what a model with more classes adds to a model with fewer classes.

To circumvent the issues just mentioned, it is most convenient if models with differing numbers of classes are substantively related; in other words, a model with $K + 1$ classes is a refined version of a model with K classes, where one of the classes is split in two parts. Such an approach would result in a hierarchical structure, comparable to hierarchical cluster analysis (Everitt, Landau, Leese, & Stahl, 2011) or regression trees (Friedman, Hastie, & Tibshirani, 2001). Van Der Palm, Van Der Ark, and Vermunt (2016) developed an algorithm for hierarchical latent class analysis that can be used for this purpose. Although they focused on density estimation, with some adaptations their algorithm has also been used to build so-called latent class trees for substantive interpretation (van Den Bergh, Schmittmann, & Vermunt, 2017). In this article, this procedure is extended to the longitudinal framework to construct latent class growth trees (LCGT).

With LCGT analysis, a hierarchical structure is imposed on the latent classes by estimating one- and two-class models on a "parent" node, which initially comprised the full data. If the two-class model is preferred according to a certain information criterion, the data are split into child nodes and separate data sets are constructed for each of the child nodes. The split is based on the posterior class membership probabilities; hence, the data patterns in each new data set will be the same as the original data set, but with weights equal to the posterior class membership probabilities for the child class concerned. Subsequently, each new child node is treated as a parent and it is checked again whether a two-class model provides a better fit than a one-class model on the corresponding weighted data set. This procedure continues until no node is split up anymore. Because of this sequential algorithm, the classes at different levels of the tree can be substantively related, as child classes are subclasses of a parent class. Therefore, LCGT modeling allows for direct interpretation of the relationship between solutions with different numbers of classes, while still retaining the same statistical basis.

The remainder of the article is set up as follows. In the next section, we discuss the basic LCG model and show how it can be used to build an LCGT. Also split criteria and guidelines for deviating from a binary split at the root of the tree are discussed, together with an entropy measure for the post-hoc evaluation of the quality of splits. Two empirical data sets are used to illustrate LCGT analysis. The article concludes with final remarks by the authors.

METHOD

Latent Class Growth Models

Let y_{it} denote the response of individual i at time point t, T_i the number of measurements of person i, and $\mathbf{y}_i$ the full response vector of person i. Moreover, let X be the discrete latent class variable, k a particular latent class, and K the number of latent classes. An LCG model is, in fact, a regression model for the responses y_{it}, where time variables are used as predictors and where intercept and slope parameters differ across latent classes. We define the LCG model within the framework on the generalized linear model, which allows dealing with different scale types of the response variable (Muthén, 2004; Vermunt, 2007).

Let $E(y_{it}|X = k)$ denote the expected value of the response at time point t for latent class k. After an appropriate transformation $g(\cdot)$, which mainly depends on the measurement level of the response variable, $E(y_{it}|X = k)$ is modeled as a linear function of time variables. The most common approach is to use polynomial growth curves, which yields the following regression model for latent class k:

$$g[E(y_{it}|X = k)] = \beta_{0k} + \beta_{1k} \cdot t + \beta_{2k} \cdot t^2 + \ldots + \beta_{sk} \cdot t^s \tag{1}$$

The choice of the degree of the polynomial (the value of s) is usually an empirical matter, although polynomials of a degree larger than three are seldom used. Recently, Francis et al. (2016) proposed an alternative approach involving the use of baseline splines in LCG models.

To complete the model formulation for the response vector $\mathbf{y}_i$, we have to define the form of the class-specific densities $f(y_{it}|X = k)$, which could be univariate normal for a continuous response, binomial for a binary response, and so on. The response density for class k is a function of the expected value $E(y_{it}|X = k)$ and for continuous variables also of the residual variance. The LCG model for $\mathbf{y}_i$ can now be defined as follows:

$$f(\mathbf{y}_i) = \sum_{k=1}^{K} P(X = k) \prod_{t=1}^{T_i} f(y_{it}|X = k), \tag{2}$$

where the size of class k is represented by $P(X = k)$. A graphical representation of an LCG model with $K = 3$ can be seen in Figure 1.

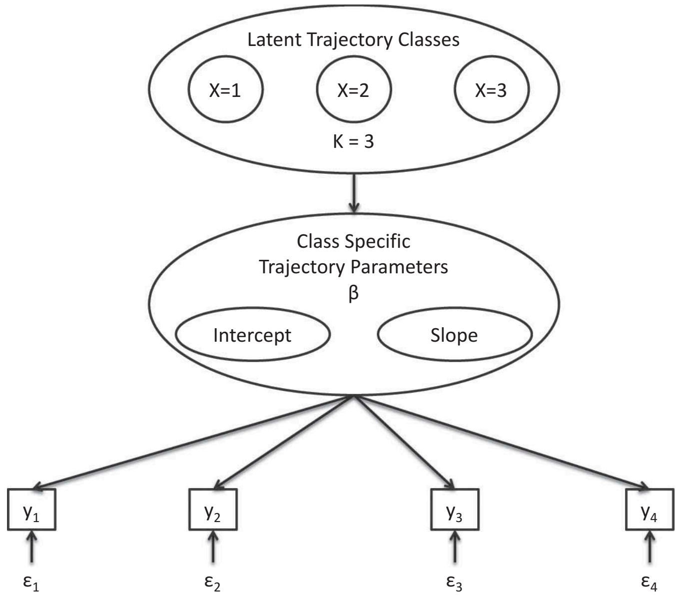

FIGURE 1 Graphical representation of a latent class growth model with three trajectory classes.

The model estimates (the β parameters and class sizes) can be obtained by maximizing the following log-likelihood function:

$$logL(\theta; \mathbf{y}) = \sum_{i=1}^{N} \log f(\mathbf{y}_i), \quad (3)$$

where $f(\mathbf{y}_i)$ takes the form defined in Equation 2 and N denotes the total sample size. Maximization is usually achieved through an expectation maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977), possibly combined with a Newton-type algorithm (Vermunt & Magidson, 2013).

After selecting a particular model, individuals could be assigned to latent classes based on their the posterior class membership probabilities. Using the Bayes theorem, these probabilities are obtained as follows:

$$P(X = k|\mathbf{y}_i) = \frac{P(X = k)\prod_{t=1}^{T_i} f(y_{it}|X = k)}{f(\mathbf{y}_i)}. \quad (4)$$

Latent Class Growth Tree Models

Using an algorithm similar to the algorithm developed by Van der Palm et al. (2016) for divisive latent class analysis, an LCG model can also be constructed in a tree form. Such an LCGT has the advantages that increasing K classes to $K + 1$ classes results in directly related classes. This is because newly formed classes are obtained by splitting one of the K classes. Due to this direct relation, models with different numbers of classes can be substantively related, while still retaining the same statistical basis. In what follows, we first describe the algorithm for constructing an LCGT in more detail, and subsequently discuss various statistics that can be used during this process.

An LCGT consists of parent and child nodes. Every set of child nodes is based on one parent node and the first parent node consists of the root node containing the complete data set. At each parent node, standard LCG models are used and its child nodes are the classes assessed with the selected parent model. At the next level of the tree, these child nodes, in their turn, become parent nodes, and conditional on each new parent node a new set of LCG models is defined. This process continues until a stopping criterion is reached, for example, when the BIC no longer decreases when splitting.

The basic equations of the growth curves of an LCGT model do not differ from those of a standard LCG model (e.g., Equation 1). The fact is that the LCGT model is based on LCG models at parent nodes, which can be formulated as follows:

$$Pf(\mathbf{y}_i|X_{parent}) = \sum_{k=1}^{K} P(X_{child} = k|X_{parent}) \prod_{t=1}^{T} f(y_{it}|X_{child} = k, X_{parent}), \quad (5)$$

where X_{parent} represents the parent class at level l and X_{child} represents one of the K possible newly formed classes at level $l + 1$, with in general K being 2. Furthermore, $P(X_{child} = k|X_{parent})$ represents the size of a class, given the parent node, and $f(y_{it}|X_{child} = k, X_{parent})$ represents the class-specific response density at time point t, given the parent class. In other words, as in a standard LCG analysis, a model for $\mathbf{y}_i$ is defined, but now conditioned on belonging to the parent class concerned.

Estimation of the LCG model at the parent node X_{parent} involves maximizing the following weighted log-likelihood function:

$$\log L(\theta; \mathbf{y}, X_{parent}) = \sum_{i=1}^{N} w_{i,X_{parent}} P(\mathbf{y}_i | X_{parent}), \tag{6}$$

where $w_{i,X_{parent}}$ is the weight for person i at the parent class, which equals this person's posterior probability of belonging to the parent class concerned. So, building an LCGT involves estimating a series of LCG models using weighted data sets.

To see how the weights $w_{i,X_{parent}}$ are constructed, let us first look at the posterior class membership probabilities for the child nodes, conditional on the corresponding parent node. Assuming a split is accepted, the posteriors are obtained as follows:

$$P(X_{child} = k | \mathbf{y}_i; X_{parent}) = \frac{P(X_{child} = k | X_{parent}) \prod_{t=1}^{T_i} f(y_{it} | X_{child} = k, X_{parent})}{P(\mathbf{y}_i | X_{parent})}. \tag{7}$$

As proposed by Van Der Palm et al. (2016), we use a proportional split based on these posterior class membership probabilities for the K child nodes conditional on the parent node, denoted by $k = 1, 2, ..., K$. If a split in two classes is performed, the weights for the two newly formed classes at the next level are obtained as follows:

$$w_{i,X_{child}=1} = w_{i,X_{parent}} P(X_{child} = 1 | \mathbf{y}_i; X_{parent}) \tag{8}$$

$$w_{i,X_{child}=2} = w_{i,X_{parent}} P(X_{child} = 2 | \mathbf{y}_i; X_{parent}). \tag{9}$$

In other words, a weight for individual i at a particular node equals the weight at the parent node times the posterior probability of belonging to the child node concerned conditional on belonging to the parent node. As an example, the weights $w_{i,X_1=2}$ used for investigating a possible split of class $X_1 = 2$ are constructed as follows:

$$w_{i,X_{12}} = w_{i,\, X=1} P(X_1 = 2 | \mathbf{y}_i, X = 1), \tag{10}$$

where in turn $w_{i,X=1} = P(X = 1 | \mathbf{y}_i)$. This implies:

$$w_{i,X_{12}} = P(X = 1 | \mathbf{y}_i) P(X_1 = 2 | \mathbf{y}_i, X = 1), \tag{11}$$

which shows that a weight at Level 2 is in fact a product of two posterior class membership probabilities.

Construction of an LCGT can be performed using standard software for latent class analysis, namely by running a series of latent class models with the appropriate weights. After each accepted split a new data set is constructed and the procedure repeats itself. We developed an R routine in which this process is fully automated. It calls the Latent GOLD program (Vermunt & Magidson, 2013) in batch mode to estimate one- and two-class models, evaluates whether a split should be made, and keeps track of the weights when a split is accepted. In addition, it creates various graphical displays, which facilitates the interpretation of the LCGT (see among others Figure 2). A novel graphical display is a tree depicting the class-specific growth curves for each of the classes in the tree and the newly formed child classes. In the trees, the name of a child class equals the name of the parent class plus an additional digit, a 1 or a 2. To prevent the structure of the tree from be affected by label switching resulting from the fact that the order of the newly formed classes depends on the random starting values, when building the LCGT we locate the larger class at the left branch with number 1 and the smaller class at the right branch with number 2.

Statistics for Building and Evaluating the LCGT

Different types of statistics can be used to determine whether a split should be accepted or rejected. Here, we will use the BIC (Schwarz, 1978), which is defined as follows:

$$BIC = -2\log L(\theta; \mathbf{y}, X_{parent}) + \log(N)P, \tag{12}$$

where $logL(.)$ represents the log-likelihood at the parent node concerned, N is the total sample size, and P is the number of parameters of the model at hand. Thus, a split is performed if at a parent node concerned the BIC for the two-class model is lower than the one of the one-class model. Note that using a less strict criterion (e.g., AIC) will yield the same splits as the BIC, but possible also additional splits, and thus a larger tree.

Special attention needs to be dedicated to the first split at the root node of the tree, in which one picks up the most dominant features in the data. In many situations, a binary split at the root might be too much of a simplification, and one would prefer allowing for more than two classes in the first split. For this purpose, we cannot use the usual criteria like AIC or BIC, as this would boil down to again using a standard LCG model. Instead, for the decision to use more than two classes at the root node, we propose looking at the relative improvement of fit compared to the improvement between the one- and two-class models. When using the log-likelihood value as the fit measure, this implies assessing the increase in log-likelihood between, say, the two- and three-class models and comparing it to the increase between the one- and two-class models. More explicitly, the relative improvement between models with K and K + 1 classes ($RI_{K,K+1}$) can be computed as:

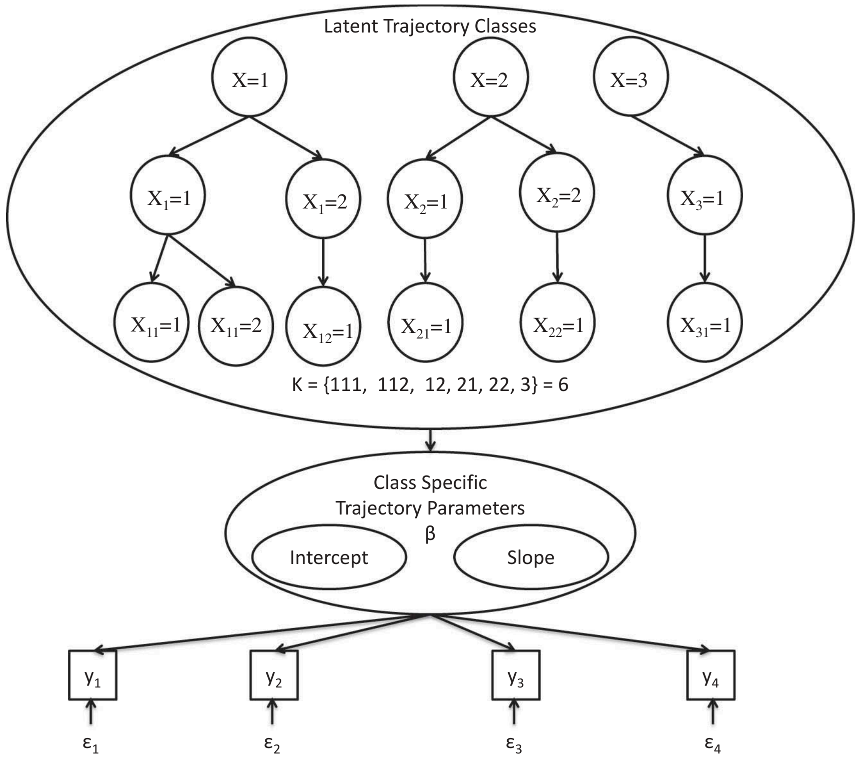

FIGURE 2 Graphical example of a latent class growth tree model with a root of three classes.

$$RI_{K,K+1} = \frac{\log L_{K+1} - \log L_K}{\log L_2 - \log L_1}, \tag{13}$$

which yields a number between 0 and 1, where a small value indicates that the K-class model can be used as the first split, and a larger value indicates that the tree might improve with an additional class at the root of the tree. Note that instead of an increase in log-likelihood, in Equation 13 one could use other measures of improvement of fit, such as the decrease of the BIC or the AIC. Scree plots depicting the difference in log-likelihood (or BIC or AIC) for models with one class difference can also be used to judge whether the relative improvement is large, as illustrated in the empirical examples presented later.

The *BIC* and $RI_{K,K+1}$ statistics are used to determine whether and how splits should be performed. However, often we are also interested in evaluating the quality of splits in terms of the amount of separation between the newly formed classes; that is, to determine how different the classes are. In other words, is a split substantively important?. This is also relevant if one would like to assign individuals to the classes resulting from an LCGT. Note that the assignment of individuals to the two child classes is more certain when the larger of the posterior probabilities $P(X_{child} = k|\mathbf{y}_i; X_{parent})$ is closer to 1. A measure to express this is the entropy; that is,

$$Entropy(X_{child}|\mathbf{y}) = \sum_{i=1}^{N} w_{i|X_{parent}} \sum_{k=1}^{2} -P(X_{child} = k|\mathbf{y}_i; X_{parent}) \\ logP(X_{child} = k|\mathbf{y}_i; X_{parent}). \tag{14}$$

Typically $Entropy(X_{child}|\mathbf{y})$ is rescaled to lie between 0 and 1 by expressing it in terms of the reduction compared to $Entropy(X_{child})$, which is the entropy computed using the unconditional class membership probabilities $P(X_{child} = k|X_{parent})$. This so-called $R^2_{Entropy}$ is obtained as follows:

$$R^2_{Entropy} = \frac{Entropy(X_{child}) - Entropy(X_{child}|\mathbf{y})}{Entropy(X_{child})} \tag{15}$$

The closer $R^2_{Entropy}$ is to one, the better the separation between the child classes in the split concerned.

EMPIRICAL EXAMPLES

The proposed LCGT methodology will be illustrated by the analyses of two longitudinal data sets. The data set in the first example contains a yearly dichotomous response on drug use collected using a panel design. The second data set contains an ordinal mood measure, recorded using an experience sampling design with eight measures per day during 1 week. The two

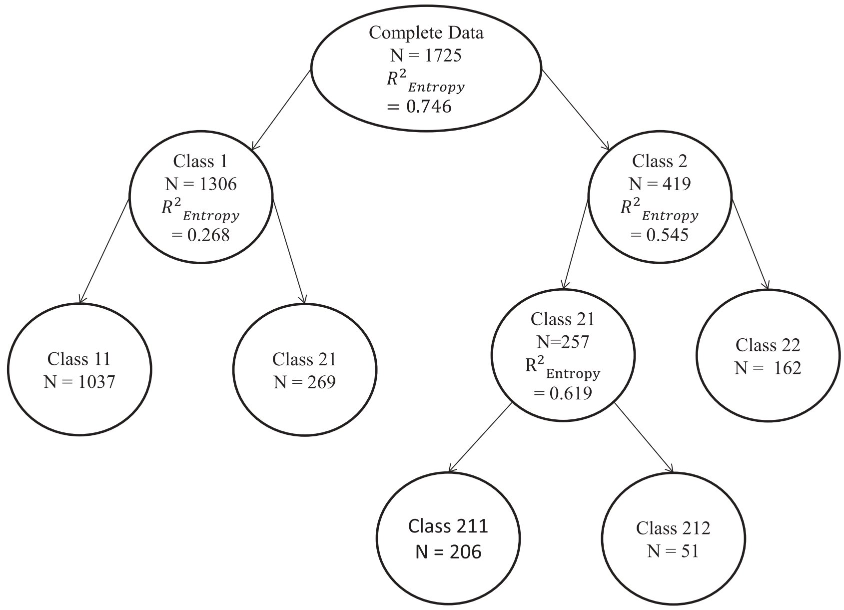

FIGURE 3 Layout, class sizes, and $R^2_{Entropy}$ of every split of a latent class growth tree with a root of two classes on drug use.

data sets illustrate LCGT analyses, differing in the number of classes at their root node. For both examples, the quality of the splits will also be evaluated using the entropy-based R^2.

Example 1: Drug Use

The first data set stems from the National Youth Survey (Elliot, Huizinga, & Menard, 1989). It contains nine waves, from 1976 to 1980 yearly and from 1980 to 1992 with 3-year intervals. The age at the first wave of the 1,725 respondents (53% men, 47% women) varied between 11 and 17 years. We use age at the panel wave concerned as the time variable, which takes on values ranging from age 11 to 33. Each respondent has been observed at most nine times (on average 7.93 times). Although the presence of (systematic) missing values could influence the results, as is typically done in growth modeling, in this illustrative example we assume the missing data to be missing at random. The dichotomous dependent variable of interest in our example will be whether the respondent used drugs or not during the past year.

The LCGT solution obtained with a third-order polynomial was very similar to the one with a second-order polynomial, both in terms of the shapes of the growth curves and the class assignments. In the LCGT starting with three classes, the class allocations differed for only 12 out of the 1,725 respondents. The tree structure and the class sizes at the splits[1] are presented in Figure 3. As can be seen, there are four binary splits, which result in a total of five latent classes at the end nodes.

To determine whether it would be better to increase the number of classes at the root of the tree, we can look at the relative improvement in fit of models with more than two classes according to the likelihood, BIC, and AIC as reported in Table 1. As can be seen, the relative improvement with a third class is around 10%. As this is quite low, we retain the tree with a binary split at the root.

To interpret the encountered classes, the growth curves can be plotted for the two newly formed classes at each node of the tree. This is displayed in Figure 4. As can be seen, the first split results in a class with a low probability to use drugs (Class 1) and a class with a high probability to use drugs (Class 2). Subsequently both of these classes are split further. Class 1 is split into Class 11 with a very low probability of using drugs (on average 0.01%) and Class 12 with a low probability during the first few years, but with

TABLE 1
Fit Statistics of a Traditional Latent Class Growth Model With One to Six Classes

	Log L	*P*	*BIC*	*AIC*	RI_{logL}	RI_{BIC}	RI_{AIC}
1	−5,089	3	10,200	10,183			
2	−4,246	7	8,543	8,505			
3	−4,156	11	8,394	8,334	0.106	0.090	0.102
4	−4,086	15	8,284	8,202	0.083	0.067	0.079
5	−4,046	19	8,233	8,129	0.048	0.031	0.043
6	−4,028	23	8,228	8,102	0.021	0.003	0.016

Note. BIC = Bayesian information criterion; AIC = Akiake's information criterion.

[1] Every split should sum up to the class size of its parent node. However, because the allocation is carried out on the basis of the posterior probabilities, the class sizes are not integers. For convenience, these numbers have been rounded, which causes slight deviations where the sum of two child nodes does not exactly add up to the parent node.

a slight increase from age 20 to 33. Class 2 is split into Class 21 and Class 22, which mainly differ in the moment at which the probability of drug use is the highest: Respondents of Class 21 start using drugs a few years earlier than respondents of Class 22. Finally, Class 21 is split further, where Class 211 has a moderate probability (around 0.6) of using drugs at an early age, but this probability also quickly declines. Class 212 has a very high probability (around 0.95) to start using drugs at an early age and this probability stays quite constant up to age 25.

The $R^2_{Entropy}$ values confirm what could also be seen from the depicted growth curves: The first split on the complete data set shows a large difference between the two classes with $R^2_{Entropy}$ of 0.746. Furthermore, Classes 11 and 12 are quite similar with $R^2_{Entropy}$ of 0.268, whereas the differences between Classes 21 and 22 and between Classes 211 and 212 are substantial (the $R^2_{Entropy}$ values are 0.545 and 0.619, respectively). Hence, after the first split, the branch of Class 2 contains more important additional differences than the one of Class 1. As an additional check we ran an LCGT using only the data from respondents with at least eight waves observed and obtained very similar results.

Example 2: Mood Regulation

The second data set stems from a momentary assessment study by Crayen, Eid, Lischetzke, Courvoisier, and Vermunt (2012). It contains eight mood assessments per day during a period of 1 week among 164 respondents (88 women and 76 men, with a mean age of 23.7, *SD* = 3.31). Respondents answered a small number of questions on a handheld device at pseudorandom signals during their waking hours. The delay between adjacent signals could vary between 60 and 180 min (*M* = 100.24 min, *SD* = 20.36, min = 62, max = 173). Responses had to be made within a 30-min time window after the signal, and were otherwise counted as missing. On average, the 164 participants responded to 51 (of 56) signals (*M* = 51.07 signals, *SD* = 6.05, min = 19, max = 56). In total, there were 8,374 nonmissing measurements. The missing data are assumed to be missing at random.

At each measurement occasion, participants rated their momentary mood on an adapted short version of the Multidimensional Mood Questionnaire (MMQ). Instead of the original monopolar mood items, a shorter bipolar version was used to fit the need for brief scales. Four items assessed pleasant or unpleasant mood (happy–unhappy, content–discontent, good–bad, and well–unwell). Participants rated how they momentarily feel on a 4-point bipolar intensity scale (e.g., *very unhappy, rather unhappy, rather happy, very happy*). For this analysis, we focus on the item well–unwell. Preliminary analysis of the response category frequencies showed that the lowest category (i.e., very unwell) was only chosen in approximately 1% of all occasions. Therefore the two lower categories were collapsed together into one unwell category. The following analysis is based on the recoded item with three categories (Crayen et al., 2012).

For the analysis, we used an LCG model based on an ordinal logit model. The time variable was the time during the day, meaning that we modeled the mood change during the day. Application of a traditional LCG model, using the BIC for model selection, resulted in seven classes with cubic growth curves (see Table 2). The cubic curves seemed to be flexible enough, as adding a quartic term did not improve model fit. The class-specific growth curves are displayed in Figure 5. As

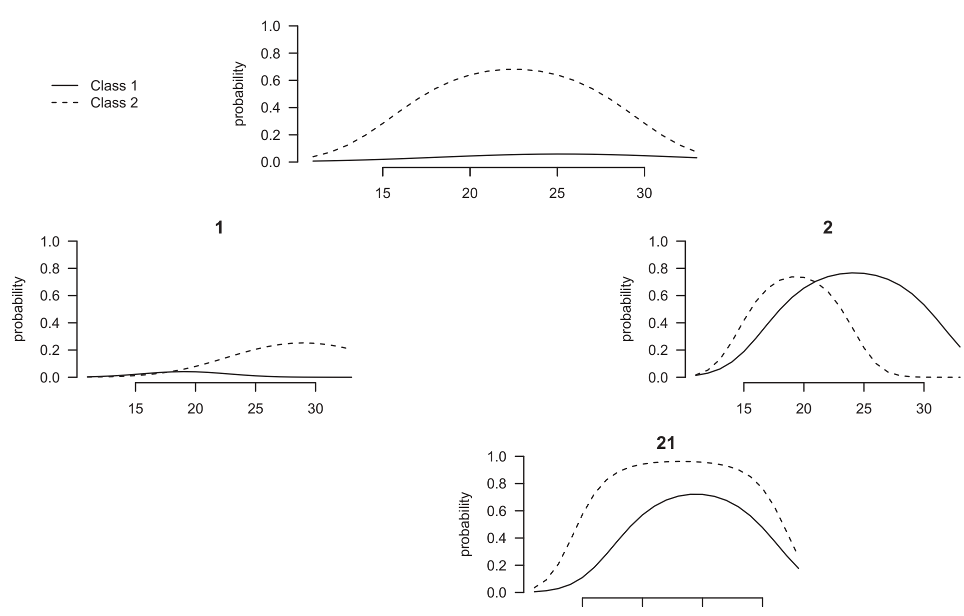

FIGURE 4 Latent class growth three with a root of two classes on drug use over age.

TABLE 2
Likelihood, Number of Parameters, Bayesian Information Criterion (BIC) and Relative Improvement of the Likelihood and BIC of a Traditional Latent Class Growth Model With One to Six Classes

	Log L	*P*	*BIC*	*AIC*	RI_{logL}	RI_{BIC}	RI_{AIC}
1	−7,199	4	14,424	14,408			
2	−6,741	9	13,538	13,504			
3	−6,578	14	13,244	13,191	0.355	0.333	0.347
4	−6,516	19	13,149	13,077	0.137	0.107	0.126
5	−6,471	24	13,091	13,001	0.097	0.065	0.085
6	−6,443	29	13,064	12,956	0.062	0.030	0.050
7	−6,424	34	13,058	12,931	0.040	0.007	0.028
8	−6,414	39	13,068	12,922	0.022	−0.012	0.009

Note. AIC = Akaike's information criterion.

can be seen, there is a clear high and a clear low class, whereas the remaining five "average" classes are rather similar to one another. This indicates that an LCGT might yield a simpler interpretation of the classes detected for this data set.

The LCGT model obtained with a root of two classes is quite large, with in total seven binary splits, resulting in a total of eight latent classes. A large tree already indicates that a larger number of classes at the root of the tree might be appropriate. Moreover, based on the relative improvement of the log-likelihood, BIC, and AIC (Table 2), it seems sensible to increase the number of classes at the root of the tree. A scree plot of the relative change in log-likelihood, BIC, and AIC also shows that after three classes the relative gain is quite small for both measures (Figure 6).

The layout and size of the LCGT with three root classes can be seen in Figure 7 and its growth curve plots in Figure 8. The growth plots show that at the root of the tree, the three different classes all improve their mood during the day. They differ in their overall mood level, with Class 3 having the lowest and Class 2 the highest overall score. Moreover, Class 1 seems to be more consistently increasing than the other two classes.

These three classes can be split further. Class 1 splits into two classes with both an average score around one, Class 11 just above and Class 12 just below. Moreover, the increase in Class 11 is larger than in Class 12. The split of Class 2 results in Class 21 consisting of respondents with a very good mood in the morning, a quick decrease until midday, and a subsequent increase. In general the mean score of Class 21 is high relative to the other classes. Class 22 starts with an average mean score and subsequently only increases. The splitting of Class 3 results in two classes with a below average mood. Both classes increase, Class 31 mainly in the beginning and Class 32 mainly at the end of the day.

The $R^2_{Entropy}$ of the different splits is quite high. The root of the tree has $R^2_{Entropy}$ of 0.889, and $R^2_{Entropy}$ of the subsequent splits is 0.734, 0.932, and 0.897, respectively. This indicates that the differences between Subclasses 21 and 22 are larger

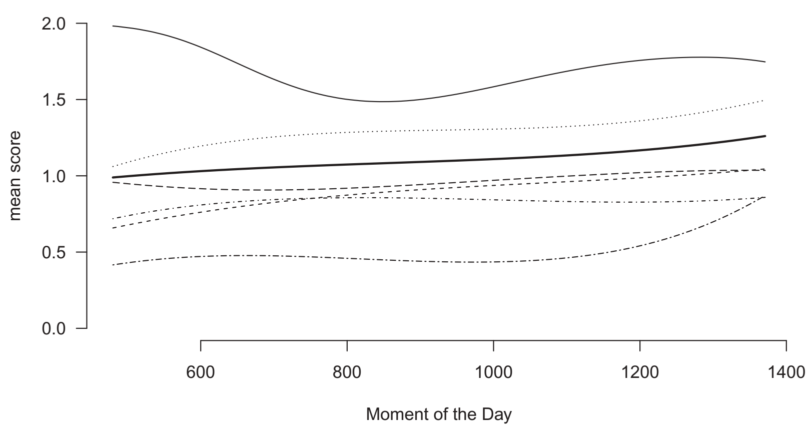

FIGURE 5 Profile plot of a latent class growth model on mood regulation with seven classes.

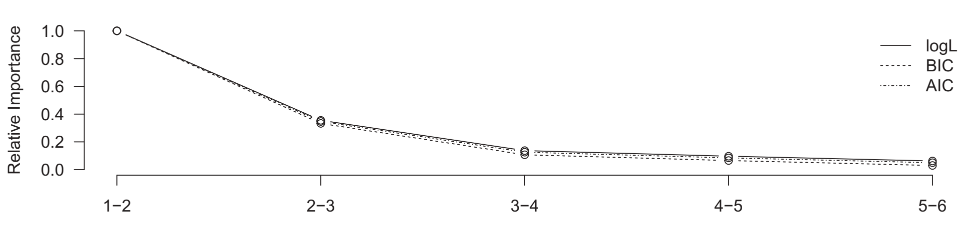

FIGURE 6 Scree plot of the difference in likelihood and Bayesian information criterion (BIC) of successive latent class growth models for the data on mood regulation. *Note*: logL = log-likelihood; AIC = Akaike's information criterion.

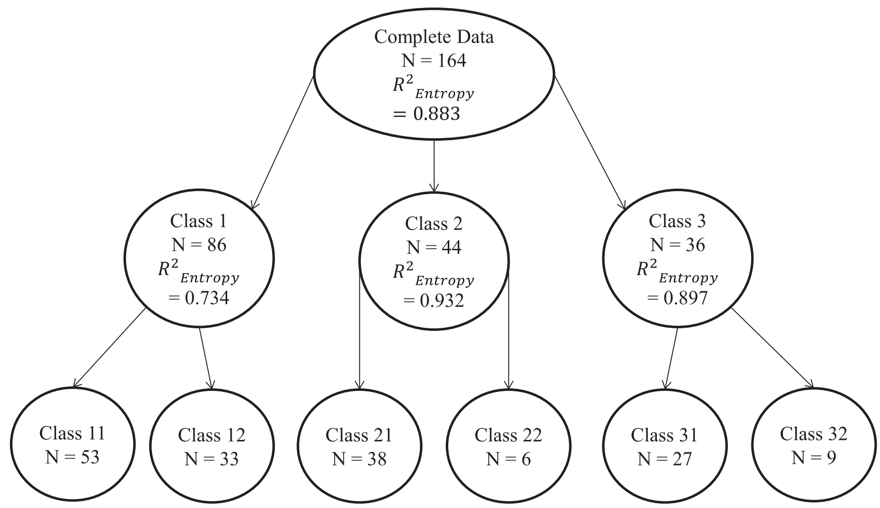

FIGURE 7 Layout, class sizes, and $R^2_{Entropy}$ of every split of a latent class growth three with a root of three classes on mood regulation during the day.

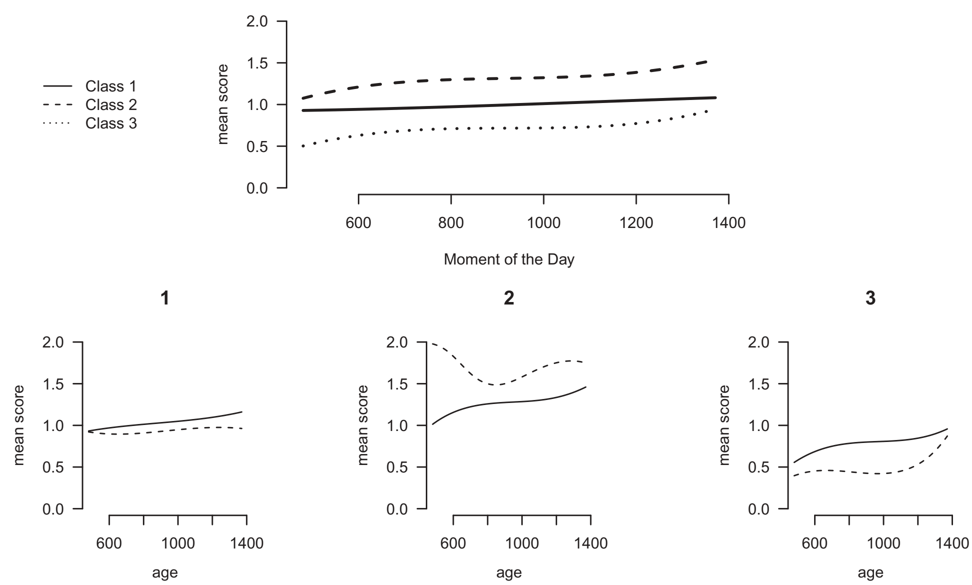

FIGURE 8 Latent class growth tree with a root of three classes of mood regulation during the day.

than those between Subclasses 31 and 32, whereas Classes 11 and 12 differ the least.

To compare the results obtained with the traditional LCG model with those of the LCGT we looked at the modal class assignments. Table 3 cross-tabulates the class allocations of the two methods. It can be seen that some classes are quite some similar; for instance, traditional Classes 3, 6, and 7 contain fairly the same respondents as the tree-based Classes 3, 4, and 6, respectively. This is not the case when using a tree with a root of two classes, as can be seen in Table 4. The differences between the remaining classes of an LCGT with a root of three classes and the seven-class LCG model are due to the first split of the tree. This can be seen in Table 5, which cross-tabulates class allocation from a seven-class and a three-class LCG model (the first level of the preferred LCGT). This also illustrates the problem of traditional LCG models, as the seven classes cannot be seen as resulting from a further splitting of these three classes. In contrast, this is exactly what is achieved with the LCGT approach, as can be seen in Table 6: Pairs of classes of the six-class model form a class of the three-class model.[2]

[2] This is true for all respondents, except for one that is in Class 1 at the first level of the tree and Class 5 at the second level of the tree. Respondents can still switch between branches of an LCGT, but this is much more restricted than in traditional LCG models.

TABLE 3
A Cross Table Showing the Differences and Similarities in Modal Assignment of a Traditional Latent Class Growth Model With Seven Classes (Rows) and a Latent Class Growth Tree With a Root of Three Classes (Columns)

	1	2	3	4	5	6
1	37	0	3	0	0	0
2	17	0	2	0	14	0
3	0	0	32	0	0	0
4	1	11	0	0	13	0
5	0	21	0	0	0	0
6	0	0	0	0	0	9
7	0	0	0	5	0	0

TABLE 4
A Cross Table Showing the Differences and Similarities in Modal Assignment of a Traditional Latent Class Growth Model With Seven Classes (Rows) and a Latent Class Growth Tree With a Root of Two Classes (Columns)

	1	2	3	4	5	6	7	8
1	13	13	0	0	0	0	14	0
2	14	0	0	16	0	3	0	0
3	0	0	0	0	0	25	3	4
4	8	0	8	9	0	0	0	0
5	1	0	20	0	0	0	0	0
6	0	0	0	2	7	0	0	0
7	0	0	0	0	0	0	0	5

TABLE 5
A Cross Table Showing the Differences and Similarities in Modal Assignment of a Latent Class Growth Tree With Six Classes (Rows) and a Traditional Latent Class Growth Model With Three Classes (Columns)

	1	2	3
1	37	3	0
2	17	2	14
3	0	32	0
4	13	0	12
5	21	0	0
6	0	0	9
7	0	5	0

TABLE 6
A Cross Table Showing the Differences and Similarities in Modal Assignment of a Traditional Latent Class Growth Model With Three Classes (Rows) and a Latent Class Growth Tree With a Root of Three Classes (Columns)

	1	2	3
1	55	0	0
2	32	0	0
3	0	37	0
4	0	5	0
5	1	0	26
6	0	0	9

DISCUSSION

LCG models are used by researchers who wish to identify (unobserved) subpopulations with different growth trajectories using longitudinal data. However, often the number of latent classes encountered is rather large, making interpretation of the results difficult. Moreover, because solutions with different numbers of classes are unrelated, a substantive comparison of models with different numbers of classes is not possible, which is especially problematic when different model selection criteria point at a different optimal number of classes. To resolve these issues, we proposed using LCGT models in which the identification of the latent classes is done in a sequential manner. The constructed hierarchical tree will show the most important distinctions in growth trajectories in the first splits, and more detailed distinctions in latter splits. Although we primarily used binary splits, we also showed how to decide about larger splits using relative improvement of fit measures. The latter is mainly of interest at the root of the tree. The proposed LCGT algorithm and graphical displays that are available as R code were illustrated with two empirical examples. The two illustrative examples showed that easily interpretable solutions are obtained using our new procedure. The fact that we impose a tree structure is both a strong point (it simplifies interpretation of the classes) and also a limitation in the sense the method will be less useful if there are no classes that are similar and that can thus be seen as children of the same parent. This is partially dealt with by allowing for a larger number of classes at the initial split, where the classes are not assumed to be hierarchically linked. In future research we want to investigate via simulation studies how well the LCGT modeling approach can pick up known class structures. As with other clustering methods, the LCG model is typically used as an exploratory clustering tool. In the context of exploratory clustering, it has been shown that even when the true class structure is not hierarchical, hierarchical clustering methods could perform very well (Ghattas, Michel, & Boyer, 2017).

The fact that we impose a tree structure is both a strong point (it simplifies interpretation of the classes) and a limitation in the sense the method will be less useful if there are no classes that are similar and that can thus be seen as children of the same parent. This is partially dealt with by allowing for a larger number of classes at the initial split, where the classes are not assumed to be hierarchically linked. In future research we want to investigate via simulation studies how well the LCGT modeling approach can pick up known class structures. As other clustering methods, the LCG model is typically used as an exploratory clustering tool. In the context of exploratory clustering, it has been shown that even when the true class structure is not hierarchical, hierarchical clustering methods might perform very well (Ghattas et al., 2017).

Various extensions and variants of the proposed procedure are possible and worth studying in more detail. Whereas in this article we restricted ourselves to LCGTs with only binary splits after the split at the root of the tree, and also at the second and next levels, it might be of interest to use larger split sizes, which could result in a tree with different split sizes within branches. Because the size of the splits could strongly affect the structure of the constructed LCGT, we recommend deciding this separately per split rather than using a fully automated procedure. Note that at lower branches of a tree there is also more substantive information available to guide the decision regarding the number of child classes.

The BIC was used to decide whether or not to split a class, as it is the most commonly used criterion and has been shown to perform well for standard latent class and LCG analysis (Nylund et al., 2007). However, other measures could be used as well, where their strictness will influence the likelihood of starting a new branch within a tree. For instance, the AIC is more lenient than the BIC and would therefore result in a larger tree with more splits, but also containing the splits of the BIC-based tree. It should be noted that the relative improvement in fit is rather similar for different fit measures, as the difference in penalty terms are canceled out. Hence, the decision criterion used mostly affects the bottom part of the tree and much less the decision regarding the number of initial classes. Therefore, the exact choice of a criterion depends on the required specificity of the encountered growth trajectories, where a less strict criterion could be used if one wishes to see more specific classes at the bottom of the tree. Note furthermore that other alterations are possible, such as a BIC with a sample size adjustment for every split (Sclove, 1987). Other criteria, such as the minimum class size, can be incorporated in the decision as to whether to perform a split. Note that sometimes classes with a very small size might point to the presence of outliers, and could thus be useful to detect.

Although LCG models are becoming very popular among applied researchers, the use of these models is not easy at all (Van De Schoot, Sijbrandij, Winter, Depaoli, & Vermunt, 2017). We hope that the proposed LCGT methodology will simplify the detection and interpretation of underlying growth trajectories. This, of course, does not mean that the standard LCG model is not useful anymore. In practice, a researcher might start with a standard LCG analysis, and switch to our LCGT approach when encountering difficulties in deciding about the number of classes or interpreting the differences between a possibly large number of classes.

FUNDING

This work was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek [453-10-002].

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. doi:10.1109/TAC.1974.1100705

Crayen, C., Eid, M., Lischetzke, T., Courvoisier, D. S., & Vermunt, J. K. (2012). Exploring dynamics in mood regulation: Mixture latent Markov modeling of ambulatory assessment data. *Psychosomatic Medicine*, *74*, 366–376. doi:10.1097/PSY.0b013e31825474cb

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological), 39*, 1–38.

Elliott, D.S., Huizinga, D., and Menard, S.(1989). *Multiple problem youth: Delinquency, substance use, and mental health problems*. New York: Springer-Verlag

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Hierarchical clustering. In *Cluster analysis* (5th ed., pp. 71–110). Chichester: John Wiley & Sons.

Francis, B., Elliott, A., & Weldon, M. (2016). Smoothing group-based trajectory models through b-splines. *Journal of Developmental and Life-Course Criminology*, *2*, 113–133. doi:10.1007/s40865-016-0025-6

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Berlin, Germany: Springer.

Ghattas, B., Michel, P., & Boyer, L. (2017). Clustering nominal data using unsupervised binary decision trees: Comparisons with the state of the art methods. *Pattern Recognition*, *67*, 177–185. doi:10.1016/j.patcog.2017.01.031

Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: Structural equation and multilevel modeling approaches*. New York, NY: Guilford.

Jones, B. L., Nagin, D. S., & Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods & Research*, *29*, 374–393. doi:10.1177/0049124101029003005

Jung, T., & Wickrama, K. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*, 302–317. doi:10.1111/j.1751-9004.2007.00054.x

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767–778. doi:10.1093/biomet/88.3.767

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–226. doi:10.1146/annurev.psych.51.1.201

McLachlan, G., & Peel, D. (2004). *Finite mixture models*. New York, NY: Wiley.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods*, *8*, 384–393.

Muthén, B. (2004). Latent variable analysis. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 345–368). Thousand Oaks, CA: Sage Publication.

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

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, *31*, 327–362. doi:10.1111/crim.1993.31.issue-3

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 92–105). Washington, DC: American Psychological Association.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (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

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: Sage.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464. doi:10.1214/aos/1176344136

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333–343. doi:10.1007/BF02294360

Tofighi, D., & Enders, C. K. (2008). 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.

Van De Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2017). The GRoLTS-checklist: Guidelines for reporting on latent trajectory studies. *Structural Equation Modeling*, *24*, 451–467.

van Den Bergh, M., Schmittmann, V. D., & Vermunt, J. K. (2017). Building latent class trees, with an application to a study of social capital. *Methodology,13*, 13–22.

van Der Palm, D. W., Van der Ark, L. A., & Vermunt, J. K. (2016). Divisive latent class modeling as a density estimation method for categorical data. *Journal of Classification, 33*(1), 52–72.

Vermunt, J. K. (2007). Growth models for categorial response variables: Standard, latent-class, and hybrid approaches. In K. Van Montfort, H. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 139–1580). Mahwah, NJ: Erlbaum.

Vermunt, J. K., & Magidson, J. (2013). *Technical guide for Latent GOLD 5.0: Basic, advanced, and syntax*. Belmont, MA: Statistical Innovations Inc.