

# UC Davis

## UC Davis Previously Published Works

### Title

Longitudinal Models for Ordinal Data With Many Zeros and Varying Numbers of Response Categories

### Permalink

<https://escholarship.org/uc/item/4f52b1mt>

### Journal

Structural Equation Modeling: A Multidisciplinary Journal, 22(2)

### ISSN

1070-5511 1532-8007

### Authors

McTernan, Melissa  
Blozis, Shelley A

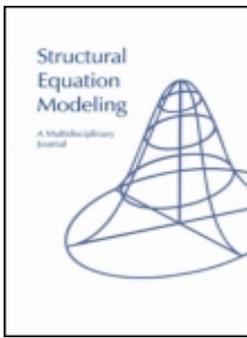
### Publication Date

2014-11-06

### DOI

10.1080/10705511.2014.935915

Peer reviewed



## Longitudinal Models for Ordinal Data With Many Zeros and Varying Numbers of Response Categories

Melissa McTernan & Shelley A. Blozis

**To cite this article:** Melissa McTernan & Shelley A. Blozis (2015) Longitudinal Models for Ordinal Data With Many Zeros and Varying Numbers of Response Categories, *Structural Equation Modeling: A Multidisciplinary Journal*, 22:2, 216-226, DOI: [10.1080/10705511.2014.935915](https://doi.org/10.1080/10705511.2014.935915)

**To link to this article:** <http://dx.doi.org/10.1080/10705511.2014.935915>



Published online: 06 Nov 2014.



Submit your article to this journal [↗](#)



Article views: 171



View related articles [↗](#)



View Crossmark data [↗](#)

# Longitudinal Models for Ordinal Data With Many Zeros and Varying Numbers of Response Categories

Melissa McTernan and Shelley A. Blozis

*University of California–Davis*

Ordinal response scales are often used to survey behaviors, including data collected in longitudinal studies. Advanced analytic methods are now widely available for longitudinal data. This study evaluates the performance of 4 methods as applied to ordinal measures that differ by the number of response categories and that include many zeros. The methods considered are hierarchical linear models (HLMs), growth mixture mixed models (GMMMs), latent class growth analysis (LCGA), and 2-part latent growth models (2PLGMs). The methods are evaluated by applying each to empirical response data in which the number of response categories is varied. The methods are applied to each outcome variable, first treating the outcome as continuous and then as ordinal, to compare the performance of the methods given both a different number of response categories and treatment of the variables as continuous versus ordinal. We conclude that although the 2PLGM might be preferred, no method might be ideal.

**Keywords:** GMMM, HLM, ordinal data, two-part latent curve models, zero-inflated

Surveys are valuable tools that allow researchers to collect longitudinal data in large samples. Commonly, survey items rely on a response scale that is categorical in nature, with response categories often following an incremental order. Often the resulting response distribution is nonsymmetric across response categories. A common instance of this is the case for which a large proportion of the responses fall into a single category, in particular, the “zero” category, often referring to a “never” or “none” response for an outcome.

Response variables that include a large proportion of zeros are common in behavioral research. Examples include measures of problem behaviors, including alcohol and substance use, and time use. Consider a survey item, for instance, that prompts participants to report how often they engaged in binge drinking in the past month. Response categories might then include “never” (coded 0), “1–2 days” (coded 1), “3–5 days” (coded 2), and so on. This survey item, if used to measure the behavior in a general adolescent sample, for instance, might contain many zero responses due to a low incidence of binge drinking for adolescents in general. Further, the positive values will be positively skewed if most

adolescents who do report drinking behaviors, for example, tend to report relatively low frequencies of the behavior.

Several methods that were developed for analysis of normally distributed longitudinal data have been applied to ordinal outcomes that also include a high proportion of zeros. Feldman, Masyn, and Conger (2009) examined four methods in particular: hierarchical linear models (HLMs), growth mixture mixed models (GMMMs), latent class growth analysis (LCGA), and longitudinal latent class analysis (LLCA). Although these advanced methods might be applicable to ordinal data, there are often issues concerning the estimation of these models in practice. As a result, researchers might find a need to modify the approach to a data analysis and use methods that might not provide the best theoretical match to the given problem. In other words, researchers could choose to sacrifice the use of more theoretically sound approaches in exchange for an approach that is more likely to provide results.

Bentler and Chou (1987) considered the use of structural equation models that assume normally distributed outcomes to ordinal response measures. In their evaluation, they recommended that if an ordered response is measured using four or more response categories, then the response could be treated as continuous to avoid the estimation problems often associated with categorical data analysis methods

---

Correspondence should be addressed to Shelley A. Blozis, Department of Psychology, University of California–Davis, 135 Young Hall, 1 Shields Avenue, Davis, CA 95616. E-mail: [sablozis@ucdavis.edu](mailto:sablozis@ucdavis.edu)

without much worry about a loss of information about the measured response (Bentler & Chou, 1987, p. 88). Other researchers have made similar observations (see Dolan, 1994). Rhemtulla, Brosseau-Liard, and Savalei (2012) performed a simulation study to assess the performance of continuous and categorical estimators on ordinal data under varying conditions, including number of response categories, asymmetric versus symmetric thresholds, and asymmetric versus symmetric underlying distributions. They recommended categorical methods if the number of response categories is fewer than five. Otherwise, they recommended the direct application of maximum likelihood methods. Notably, their simulation demonstrated that if data thresholds are extremely asymmetric or if the underlying distribution is asymmetric, parameter estimates are likely to be biased using either continuous or categorical estimators.

This study applies these ideas in the context of modern methods for longitudinal data that include many zeros by comparing the performance of different models applied to a set of ordinal measures that differ in terms of the number of response categories, with one category that includes only zeros. The outcome is a substance use variable that is likely to be asymmetrically distributed in the population. The model is applied to each outcome variable as if it were a continuous response and once again treating the response as ordinal. Comparisons between the sets of results within each method (e.g., HLM) and data treatment (as continuous vs. ordinal) are made based on the relatively best fitting model, resulting parameter estimates and standard errors, and whether interpretable parameter estimates could be obtained.

Many of the methods prescribed for categorical response data assume that underlying the observed response is a normally distributed variable. This is especially problematic for data that include a high proportion of zeros and that are otherwise positive. Thus, one of the goals of this study is to evaluate the performance of several modern methods as applied to such data but that vary with regard to the number of response categories. This study seeks to further understand whether continuous methods could serve as alternatives to categorical methods for ordinal data that have many zeros and a relatively large number of response categories.

## MEASUREMENT OF ORDINAL RESPONSES

Ordinal scales require respondents to categorize behaviors that might or might not be inherently categorical. Consider again the example of measuring frequency of adolescent binge drinking behavior. An individual who drinks once a week drinks too often to include himself or herself in the “never” category, but not often enough to be included in the third category that is reserved for people who drink three to four times a week. Thus, this participant falls in the second response category of 1–2 days. For response scales such as this, the structure of the scale forces responses into one of a

set of ordered categories. Importantly, the behavior itself is quantifiable, although the measure is recorded as categorical or, specifically, as ordinal.

In describing an ordered response that is assumed to have an underlying continuous distribution, it is useful to provide notation that could be used to define such variables. First, let  $Y_{ij}$  denote an observed response and  $Y^*_{ij}$  denote the true underlying continuous response. Let  $m$  refer to a response category with  $m = 1, \dots, c$ , where  $c$  represents the category that includes the maximum possible score. Further, let  $\nu_q$  represent a latent threshold parameter that separates the ordered response categories with  $Q$  thresholds. Given this,

$$Y_{ij} = m_q \text{ if } \nu_q \leq Y^*_{ij} < \nu_{q+1}.$$

With regard to an actual analysis of an ordinal variable, a technique could be applied directly to the observed outcome  $Y$  in which  $Y$  is treated as continuous, such as by applying a model that assumes a normal response. Conversely, a method could be applied in which the observed response is treated as ordinal, and it is the underlying continuous response  $Y^*$ , with  $Y^*$  often assumed to be normal, that is modeled. Based on the suggestions in Bentler and Chou (1987), this article evaluates both approaches to understand if the former approach might be reasonable for responses based on four or more ordered response categories with the goal of obtaining parameter estimates that perform well in describing the data without a great computational burden that might otherwise result from treating the response as ordinal.

## METHODS FOR LONGITUDINAL DATA

### Hierarchical Linear Models

HLMs can be used to model change in a normal response variable over time. This approach allows for individual differences in response trajectories, such as by including a subject-specific intercepts and slopes (Bryk & Raudenbush, 1987; Raudenbush & Bryk, 2002). Take, for example, the aforementioned adolescent binge drinking survey question. It might be reasonable to assume that at the start of a study when participants are young, all individuals will have an intercept of zero (with time centered at the start of the study) but that the rate of change in the frequency of binge drinking behaviors for some individuals might be higher than that of others. Thus, an appropriate model might include a fixed response level at the start of the study (fixed intercept) and a random time effect (random slope).

Let  $Y_{ij}$  denote a normal response measured for person  $i$  at time  $j$ , where  $i = 1, \dots, N$  with  $N$  equal to the total number individuals, and where  $j = 1, \dots, n_i$ , with  $n_i$  denoting the total number of observations for person  $i$ . Let  $t_{ij}$  denote the particular time at which  $Y_{ij}$  is observed. Assuming that  $Y_{ij}$  follows a two-level hierarchical linear model,  $Y_{ij}$  could be expressed as

$$\text{Level 1 : } Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij}$$

$$\text{Level 2 : } \beta_{0i} = \Upsilon_{00} + u_{0i}$$

$$\beta_{1i} = \Upsilon_{10} + u_{1i}$$

where  $\beta_{0i}$  is the expected response for person  $i$  at  $t_{ij} = 0$ . The coefficient  $\beta_{1i}$  is the expected rate of change in  $Y_{ij}$  for person  $i$  per unit of time. At Level 2, the random intercept and slope are each sums of a fixed effect ( $\Upsilon_{00}$  and  $\Upsilon_{10}$ , respectively) and a random effect ( $u_{0i}$  and  $u_{1i}$ , respectively). The model here includes only a linear slope, but higher order slopes might be included to account for nonlinear forms of change in  $Y_{ij}$ . Additionally, covariates could be included at the first level to account for occasion-to-occasion variation, controlling for change in  $Y_{ij}$  due to time, as well as at the second level to account for individual differences in the individual-level intercept and slope.

At the first level of the model, the set of individual and time-specific errors,  $\mathbf{e}_i = (e_{1i}, \dots, e_{n_i i})'$ , is assumed to be normal with mean equal to zero and covariance matrix  $\Theta$ . The errors could be assumed to be independent with constant variance across time, such that  $\Theta = \mathbf{I}\theta$ , where  $\mathbf{I}$  is an identity matrix of order  $n_i$ , and  $\theta$  is the common variance; other patterns are possible. At the second level, the random effects  $u_{0i}$  and  $u_{1i}$  are assumed to be normal with means equal to zero and covariance matrix  $\Phi$ , where  $\Phi$  is usually unstructured so as not to impose constraints on the variances and covariances of the random coefficients. The Level 1 error and Level 2 random effects are assumed to be independent. The parameters of the model, expressed generally, are  $(\Upsilon_{00}, \Upsilon_{01}, \dots)'$ ,  $(\phi_1, \phi_2, \dots)'$ , and  $(\theta_1, \theta_2, \dots)'$ .

A hierarchical generalized linear model (HGLM; a.k.a. generalized linear mixed model) is a generalization of HLM to handle a range of response variables including normal, ordinal, and nominal data. For an ordinal response and assuming a logit link, the individual's underlying response,  $Y^*$ , is modeled by a linear mixed model:

$$\text{Level 1 : } Y_{ij}^* = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij}$$

$$\text{Level 2 : } \beta_{0i} = \Upsilon_{00} + u_{0i}$$

$$\beta_{1i} = \Upsilon_{01} + u_{1i}$$

where  $e_{ij}$  is assumed to be distributed according to a logistic distribution, with expected value of zero and variance  $\left(\frac{\pi^2}{3}\right)$  and to be independent across individuals and conditionally independent across time points. The fixed intercept  $\Upsilon_{00}$  is fixed at zero given that the scale for the latent response is arbitrary and therefore must be assigned a value for model identification. The observed ordinal response  $Y_{ij}$  is related to the latent continuous response  $Y_{ij}^*$  by way of the thresholds described earlier:  $Y_{ij} = m_q$  if  $v_q \leq Y_{ij}^* < v_{q+1}$ , where  $v_q$  is

a threshold parameter for  $q = 1, \dots, Q$ , and  $Q$  is the total number of thresholds. Under the model for a longitudinal response, the thresholds are typically assumed to be constant across time. Similar to HLM, covariates could be included at either level of the model. The parameters of the model to be estimated are  $(\Upsilon_{01}, \dots)'$ ,  $(\phi_1, \phi_2, \dots)'$ , and  $(v_1, \dots)'$ . See Figure 1 for a structural representation of these models.

### Growth Mixture Mixed Model

The random effects at the second level of HLM are assumed to be normally distributed. An extension of HLM is a model in which the random effects are assumed to be related to two or more latent classes in what is called a GMMM (Verbeke & Lesaffre, 1996). Models with latent classes are commonly used to model substance use and risk-behaviors, for example (Henry & Muthen, 2010; Laska, Pasch, Lust, Story, & Ehlinger, 2009; Lubke & Muthén, 2005; Nylund, Bellmore, Nishina, & Graham, 2007). Mixture models were developed for continuously distributed data in which a response is assumed to be due to two or more latent classes. The data are assumed to be normally distributed within classes but not necessarily across classes. For longitudinal data, GMMM allows for variation in the random effects that characterize the individual-level responses to be due to a finite number of classes (the number of which are specified by the researcher).

GMMM is similar to HLM but includes class-level information. Assuming linear growth, the first level of the model can be written as

$$\text{Level 1 : } Y_{ijk} = \beta_{0ik} + \beta_{1ik}t_{ijk} + e_{ijk}$$

where  $k$  denotes a particular class, with  $k = 1, \dots, K$  and  $K$  being the total number of classes. Assuming random variation in both coefficients of the Level 1 model, the second level allows for variation across classes:

$$\text{Level 2 : } \beta_{0ik} = \Upsilon_{00k} + u_{0ijk}$$

$$\beta_{1ik} = \Upsilon_{01k} + u_{1ijk}$$

For an ordinal response and assuming linear growth, Level 1 of the model can be written as

$$Y_{ijk}^* = \beta_{0ik} + \beta_{1ik}t_{ijk} + e_{ijk}$$

where  $Y_{ijk}^*$  is the latent continuous response for individual  $i$  at time point  $j$  and in class  $k$ . The second level of the model would be identical to Level 2 of the GMMM that treats the response as continuous. GMMM for ordinal response data also include threshold parameters that are typically assumed to be equal across classes and time. The distributional assumptions of GMMM are identical to HLM conditional on the class-level information. Similar to

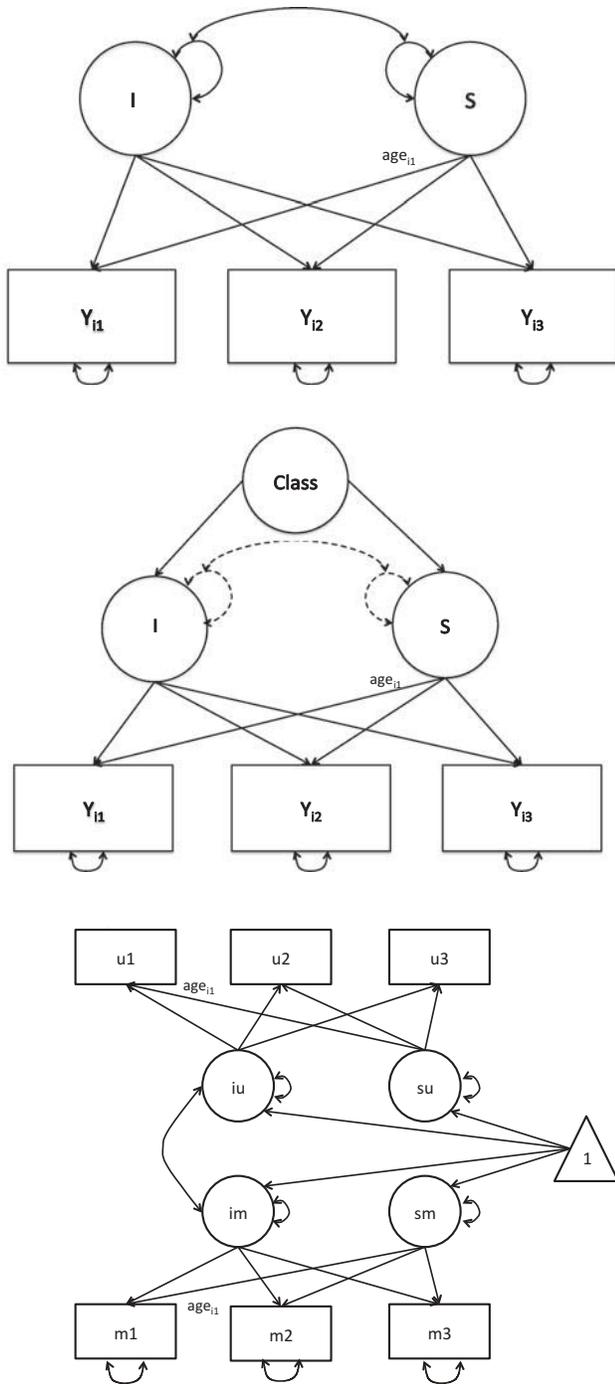


FIGURE 1 Hierarchical linear model/hierarchical generalized linear model (HLM/HGLM; top), growth mixture mixed model/latent class growth analysis (GMMM/LCGA; middle), and two-part latent growth (bottom) conceptual models.

the model for continuous data, classes might have unique mean and covariance structures. See Figure 1 for a structural representation of GMMM.

Although GMMM has been applied to ordinal data (Li, Duncan, Duncan, & Acock, 2001; Muthén & Muthén, 2000),

a study by Hipp and Bauer (2006) showed that if GMMM is applied to ordinal data, estimates of the model are often difficult to obtain. Although a solution to this has been to apply constraints to the model or to provide starting values, either approach could yield biased results given that the data are not normally distributed (Hipp & Bauer, 2006).

### Latent Class Growth Analysis

In cases where the estimation of a growth mixture model is difficult due to low within-class variability, an option might be to constrain one or more of these variances to zero. If all of the variances of the latent growth coefficients are equal, then the model reduces to what is often called a latent class growth model. Thus, unlike GMMM, in which individual differences can be represented by a series of normal distributions with each distribution representing a class, under LCGA all individuals within a class are assumed to have identical growth trajectories. Level 1 of the LCGA model is identical to that of a growth mixture model. Level 2 of the LCGA model can be expressed as

$$\beta_{0ik} = \gamma_{00k}$$

$$\beta_{1ik} = \gamma_{01k}$$

Similar to GMMM, LCGA may be applied to ordinal data to handle the nonnormality of an outcome (Li et al., 2001; Muthén & Muthén, 2000). LCGA, however, is a restricted form of a GMMM, and consequently, the arguments in Hipp and Bauer (2006) cautioning against this treatment might apply. Figure 1 is a structural display of LCGA.

### Two-Part Models for Semicontinuous Data

The problem of data with many zeros has been addressed by a two-part model proposed by Duan, Manning, Morris, and Newhouse (1983) for continuous cross-sectional data that also include many zeros. In a two-part model, the original response serves as the basis for the creation of two new variables for analysis. The first variable is dichotomous and coded 0 if the original response is equal to zero and coded 1 if the original response is positive. Thus, this variable is an indicator of whether an individual engaged in the measured outcome. The second variable is equal to the original response if greater than zero and is missing if the original response is equal to zero. This second variable is a measure of the magnitude or frequency of an outcome given that an individual engaged in the behavior. The two variables are analyzed separately using a model for the dichotomous response (e.g., logistic regression) and a model for the conditional continuous response (e.g., regression).

Olsen and Schafer (2001) extended a two-part model to handle longitudinal data by allowing for random effects in

both model parts at the second level with possible covariances between them. Indeed, ignoring nonzero covariances between the random effects of the two model parts has been shown to result in biased parameter estimates (Su, Tom, & Farewell, 2009). Figure 1 includes a path model for a two-part latent growth model that includes two growth functions with random coefficients that covary. Thus, a two-part latent growth model is one more approach to the analysis of longitudinal data with many zeros. Given an original response variable that is measured on an ordinal scale, the second part of a two-part latent growth model could be specified to treat the response as continuous, or alternatively, to treat the response by using a categorical model, such as an ordered logistic model.

### EMPIRICAL EXAMPLE

To test the performance of the four statistical methods, ordinal response data were drawn from the National Longitudinal Study of Adolescent Health (Add Health) conducted by the Inter-University Consortium for Political and Social Research (ICPSR) at the University of North Carolina, Chapel Hill (Harris & Udry, 1994–2008). Data from this study include survey responses on physical and psychological health and general well-being. There were four waves of data collection, with the last occurring in 2008. The first wave of the survey was conducted during the 1994–1995 school year and included 27,000 adolescents who were in Grades 7 through 12. Schools were randomly selected from the full national sample of high schools that met two conditions: student enrollment equal to at least 30 students and the school included an 11th grade. Feeder schools that included a seventh grade were then selected from each community.

A random selection of schools from the initial sample received follow-up in-home surveys. One year later, a second wave of data collection included as many students as could be located from the Wave 1 in-home sample, with the exclusion of 471 disabled students and the majority of 12th graders who had been interviewed at Wave 2 but then exceeded the grade eligibility requirement (Harris & Udry, 1994–2008). A third in-home survey occurred 6 years after the initial data collection and included as many participants from Wave 1 as could be located.

From this data set, we chose a substance use variable that was publicly available for the first three waves of data collection. This variable was chosen because it was considered representative of the kind of data researchers might encounter in many kinds of behavioral studies, namely an ordinal response that includes many zeros. The prompt for this variable was, “In the past 12 months, on how many occasions did you drink alcohol?” Respondents selected a response from seven categories: 1 (*every day/almost every day*), 2 (*3–5 days per week*), 3 (*1 or 2 days per week*), 4

(*2 or 3 days per month*), 5 (*once a month or less; 3–12 times in past 12 months*), 6 (*1 or 2 days in past 12 months*), and 7 (*never*). There were also options for *refused*, *don’t know*, and *legitimate skip*. The latter option was automatically selected if the participant had responded to a filter question that they had never had a drink (as asked in Wave 1) or had not had a single drink since the previous data collection (as asked in Waves 2 and 3).

Some preliminary steps were taken to simplify the data analysis for the central purpose of this study. First, data were limited to only White respondents. The data were also filtered to include only those who were in the seventh or eighth grade in Wave 1 to help limit the sample to exclude those who would reach a legal drinking age during the study. For those who by Wave 3 had reached the legal drinking age (21), their Wave 3 data were excluded from analysis, although their Waves 1 and 2 data were retained. Individuals identified as teetotalers were also excluded from analysis. These individuals were selected according to responses given to a filter question about drinking. At Wave 1, participants were asked if they had ever tried alcohol and at Waves 2 and 3 they were asked if they had drunk any alcohol since the previous interview. Participants who answered “No” at all three waves or had any combination of “No” responses and missing data for a filter question were considered teetotalers and were excluded from analyses. All other response patterns were included, for a final sample size of 1,038. The mean age of the sample at Wave 1 was 13.5 years old ( $SD = 0.78$ ). The average age at Wave 2 was 14.4 years old ( $SD = 0.82$ ) and 19.5 years old ( $SD = 0.55$ ) at Wave 3. At Waves 1, 2, and 3, the number of responses in the *zero* category were 641 (61.8%), 510 (49.1%), and 120 (11.6%), respectively. Figure 2 displays the proportions of responses across time for each of the scale versions.

### Data Analysis

The Add Health study relied on a random sample of schools and then a sampling of students within schools. Prior to conducting the primary data analysis, variation in the drinking variable was studied according to variation due to the nesting of measures within children and children within schools. Assuming a normal response, a three-level unconditional means model that assumed constant variance across schools and children was used to estimate the proportion of variance in the responses due to nesting. Neither of the intraclass correlations was statistically different from zero based on a deviance test that compared the model to a model that treated the scores as independent between children and schools. Assuming that these values were close to the true values (given that the responses were measured using ordinal scales and thus were not normal), the nesting of data within schools was considered trivial and not studied further. To study the dependencies of scores within children further, an unconditional means model was fit in which the variances could vary

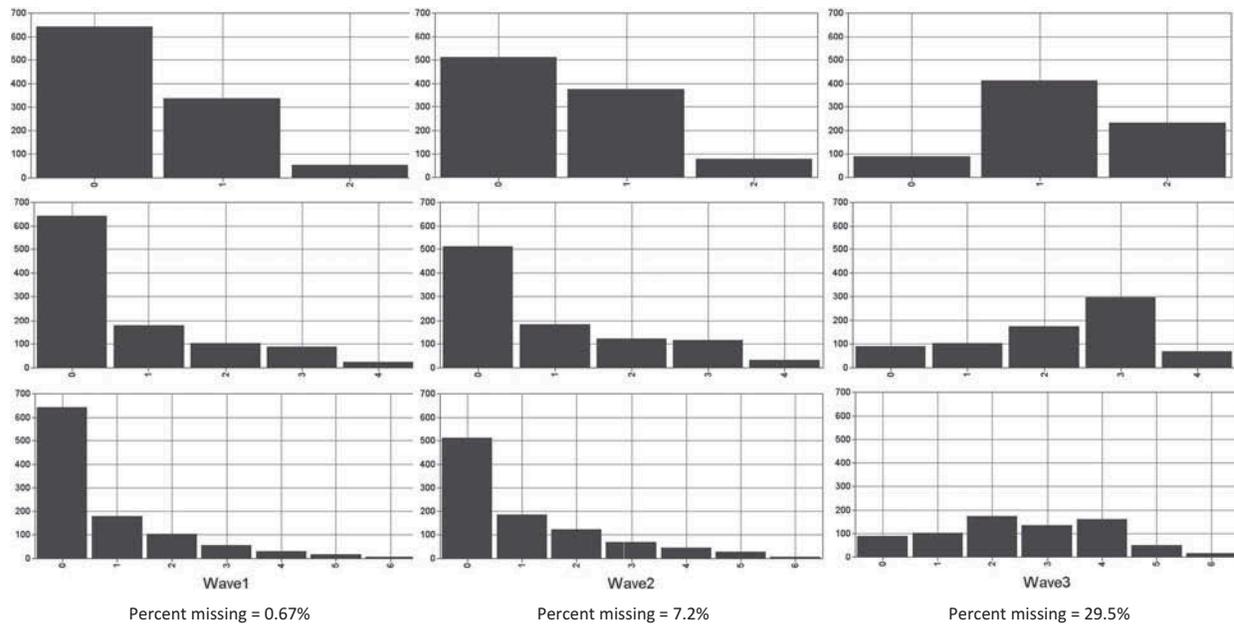


FIGURE 2 Histograms of the original response variables at Waves 1, 2, and 3.

between waves. The random intercept of this model was statistically significant, suggesting some degree of dependence of scores within children. Thus, the need to account for the nesting of data within children was considered important.

For HLM and GMMM, five growth models were tested: (a) no growth with a random intercept, (b) linear growth with a random intercept, (c) linear growth with a random intercept and slope, (d) quadratic growth with a random intercept, and (e) quadratic growth with a random intercept and linear slope. A quadratic growth model with a random linear and quadratic slope (assuming nonzero covariances between the random intercept, linear, and quadratic slopes) is not identified given that data were available for at most three measurement waves. For the growth models tested, the variances of the time-specific errors were allowed to differ by measurement wave. For LCGA, three growth functions (no growth, linear growth, and quadratic growth) were tested. No growth, linear growth, and quadratic growth forms (allowing for random intercepts and linear slopes as considered for other methods) were tested under the two-part latent growth model. Age was centered at 15 years for all analyses. Methods involving two or more latent classes (GMMM and LCGA) were fit assuming up to three classes. An increase in the number of classes was discontinued if after fitting a model with  $k$  classes the adjusted Lo–Mendell–Rubin (LMR) statistic favored  $k - 1$  classes. In such cases, the  $k - 1$  class model was selected as the final model.

Before applying the two-part latent growth model, a dichotomous variable  $u_{ij}$  was coded 0 for an individual who reported “never” to drinking at time  $j$  and was coded 1 if the original response was in a higher response category. A second variable  $m_{ij}$  was coded as missing at time  $j$  for any person

$i$  who reported “never” drinking at that time point and was coded as  $m_{ij} = Y_{ij}$  if an individual reported any amount of drinking at time  $j$ . For both variables, the values were coded as missing if the original response was missing.

Models that treated the response as categorical were specified first with a logit link. Under this specification, the response is assumed to have an underlying latent continuous distribution that follows a logistic distribution, and the parameter estimates are in log-odds units. The models were also evaluated by using a probit link function that assumes that the underlying response follows a normal distribution. Results from both treatments could be compared directly, such as by comparing Akaike’s information criterion (AIC) values.

To compare methods and data treatments across the different numbers of response categories, the original 7-point response variable was transformed to create two new variables. Specifically, the original seven response categories were collapsed into five categories (0 = 0; 1 = 1; 2 = 2; 3 = 3, 4; 4 = 5, 6) to create a second response variable and then into three categories (0 = 0; 1 = 1, 2, 3; 2 = 4, 5, 6) to create a third response variable. Scores were collapsed so that responses at the upper end of the distribution that tended to be sparse were collapsed into fewer categories.

Full information maximum likelihood estimates with robust standard errors (DiStefano, 2002) were obtained for both response treatments using *Mplus* version 6.12 (Muthén & Muthén, 1998–2011). Numerical integration was used for analyses based on HLM, GMMM, and LCGA by using ALGORITHM = INTEGRATION. Analyses were done using a Core 2 Duo E8400 at 3 GHz (one processor, two cores) desktop computer with 4 GB of RAM. The computer

operated as a shared server computer, so it was not possible to document meaningful computation times given that access by multiple users could not be controlled.

## RESULTS

For comparable models using the logit or probit link function, the AIC consistently favored the logit link, suggesting a better approximation to the data under this model. Given the consistency of this result, further discussion of categorical results is limited to models using a logit link. For each method, data treatment, and number of response categories, model fit was evaluated using combined results from model fit indices and the entropy value and the adjusted LMR statistic if applicable. Model fit indices according to method and data can be provided on request. For those models that did not result in a converged solution given the default settings in *Mplus*, we attempted to provide useful starting values, increased the number of initial and final stage starts, or increased the number of initial stage iterations or iterations for the expectation maximization (EM) algorithm, as applicable. Models provisionally selected as the final model according to method, variable treatment, and number of response categories are summarized in Table 1. Plots of the fitted fixed values are shown in Figure 3 through 5. Fit statistics are not reported for the final models because it would be inappropriate to statistically compare models across number of response categories, across variable treatment, or across models with different numbers of classes.

### HLM

**Continuous treatment.** Models based on HLM converged quickly without difficulty across the numbers of response categories relative to other methods. A linear growth model with a random intercept and slope provided the best fit to the data regardless of the number of response categories. Overall, drinking increased with age according to these models. The estimated fixed intercept and slope increased in value as the number of response categories increased, as would be expected. Similarly, the estimated covariance between the intercept and slope increased as the number of response categories increased.

**Categorical treatment.** All but 3 of the 15 HGLMs resulted in a converged solution. The quadratic growth models with a random intercept only across the three versions of the response scale failed to converge. In contrast to the continuous data treatment, the categorical treatment suggested quadratic growth in drinking levels with individual differences in the response level and rate of change at age 15. This was consistent across the numbers of response categories. The intercept and slope estimates of these models and their covariances were also comparable across the varying numbers of response categories.

TABLE 1  
Final Models by Method, Variable Treatment, and Number of Response Categories Considering HLM/HGLM, GMMM, and LCGA

Method	Treatment	Response Categories	Final Model		
			Growth	Random Coefficients	Classes
HLM	Continuous	3	Linear	$\beta_{0i}, \beta_{1i}$	N/A
		5	Linear	$\beta_{0i}, \beta_{1i}$	N/A
		7	Linear	$\beta_{0i}, \beta_{1i}$	N/A
	Categorical	3	Quadratic <sup>a</sup>	$\beta_{0i}, \beta_{1i}$	N/A
		5	Quadratic <sup>a</sup>	$\beta_{0i}, \beta_{1i}$	N/A
		7	Quadratic <sup>a</sup>	$\beta_{0i}, \beta_{1i}$	N/A
GMMM	Continuous	3	Quadratic <sup>a</sup>	$\beta_{0i}$	2
		5	Quadratic	$\beta_{0i}$	3
		7	Quadratic	$\beta_{0i}, \beta_{1i}$	2
	Categorical	3	Linear <sup>a</sup>	$\beta_{0i}, \beta_{1i}$	2
		5	Linear <sup>a</sup>	$\beta_{0i}, \beta_{1i}$	2
		7	Linear <sup>a</sup>	$\beta_{0i}, \beta_{1i}$	2
LCGA	Continuous	3	Quadratic <sup>a</sup>	N/A	2
		5	Quadratic	N/A	3
		7	Quadratic	N/A	3
	Categorical	3	Linear <sup>a</sup>	N/A	3
		5	Linear <sup>a</sup>	N/A	3
		7	Linear <sup>a</sup>	N/A	2

Note. HLM = hierarchical linear model; HGLM = hierarchical generalized linear model; GMMM = growth mixture mixed model; LCGA = latent class growth analysis.

<sup>a</sup>These models were selected because one or more competing models (with more classes, different growth trajectory, or different random coefficient combination) did not converge with trustworthy estimates. Fit statistics for all models are available on request.

### GMMM

**Continuous treatment.** GMMM under continuous treatment of the response was notably more burdensome than HLM and HGLM. Although estimates converged quickly for those models in which solutions were obtained, there were many estimation problems using this method. The best fitting model under this approach was one based on quadratic growth, across the number of response categories. A random slope did not add any explanatory power in the 3- or 5-point category models; only the 7-point scale preferred the added random slope term. Two classes were adequate for 3- and 7-point scales, but three classes were preferred for the 5-point scale. As with the continuous data treatment under HLM, the estimated fixed intercept and slope under this method increased with the number of response categories as expected.

**Categorical treatment.** Estimation of GMMM under the categorical data treatment was most burdensome of all the tested models. Ten of the 15 models resulted in estimation difficulties, with results not obtained for 6 models. Importantly, estimates could not be obtained for what was the best fitting model when treating the data as continuous under GMMM (quadratic growth with a random intercept and fixed slope). The selected model across all numbers

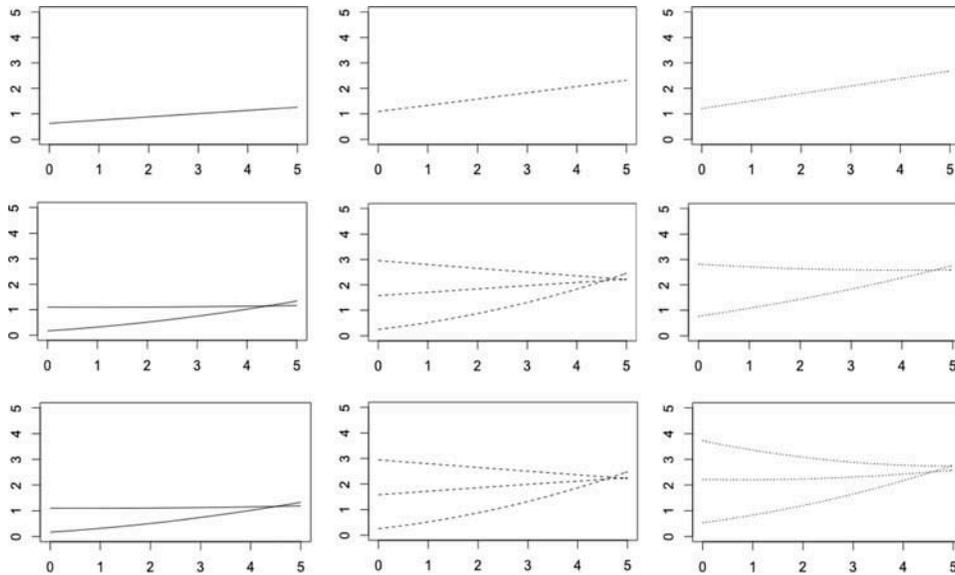


FIGURE 3 Fitted values under continuous data treatment. Columns 1 through 3 represent three-, five-, and seven-response category scales, respectively. Rows 1 through 3 represent hierarchical linear models (HLMs), growth mixture mixed models (GMMMs), and latent class growth analysis (LCGA), respectively.

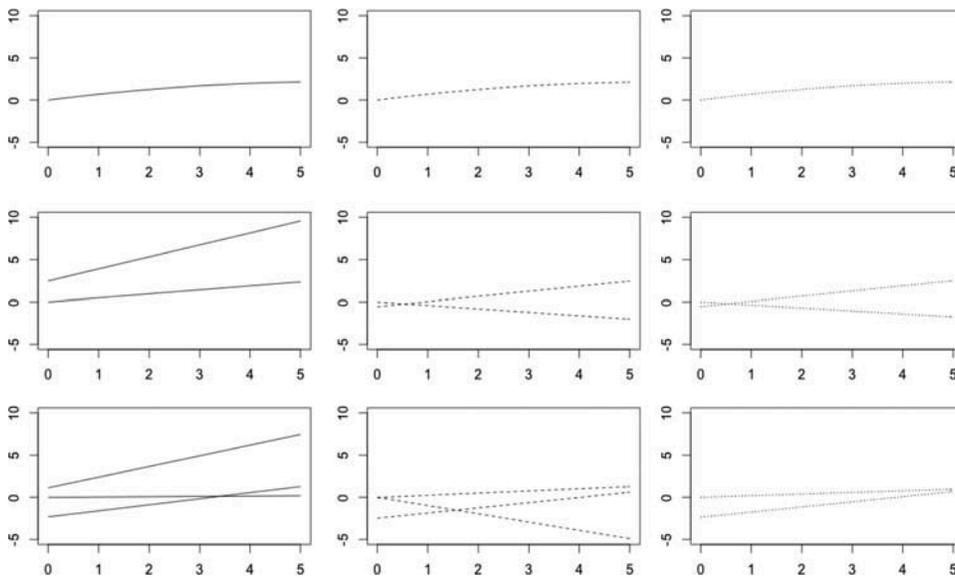


FIGURE 4 Fitted values under categorical data treatment. Columns 1 through 3 represent three-, five-, and seven-response category scales, respectively. Rows 1–3 represent hierarchical generalized linear models (HGLMs), growth mixture mixed models (GMMMs), and latent class growth analysis (LCGA), respectively.

of response categories for GMMM with the categorical treatment was a two-class linear growth model with a random intercept and slope.

#### LCGA

*Continuous treatment.* Under LCGA with the continuous data treatment, there were no estimation problems for models that assumed two latent classes. For three classes assumed, no solution was obtained for the 3-point response

scale (recall that this was also the case using GMMM). The best fitting model to these data regardless of the number of response categories was a quadratic growth model. Three classes were favored for all models for which estimates could be obtained. With the 3-point response scale, a three-class model could not be estimated, thus the selected model was a two-class model.

*Categorical treatment.* Estimation of the LCGA models with the categorical data treatment was slightly more

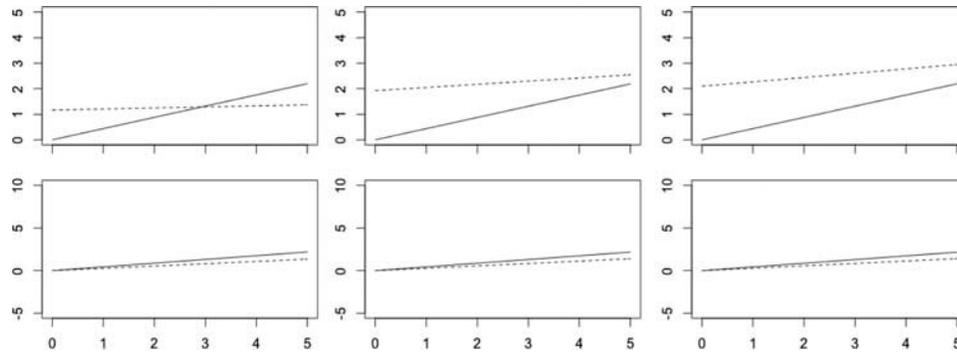


FIGURE 5 Fitted values for the two-part model. Columns 1 through 3 represent three-, five-, and seven-response category scales, respectively. Rows 1 and 2 represent continuous and categorical data treatment, respectively. The solid line represents the  $u$  part and the dashed line represents the  $m$  part of the model.

difficult than LCGA with continuous treatment. The selected model across all numbers of response categories was one that assumed linear growth. If a quadratic slope was added to the models, no estimates could be obtained. This result was consistent across the three numbers of response categories. Three classes were favored for all estimated models, with the exception of the 7-point response scale model, for which the adjusted LMR statistic suggested two classes were adequate to describe the data under this model.

### Two-Part Latent Growth Model

For the dichotomous response of the two-part model,  $u_{ij}$ , parameter estimates were similar across all numbers of response categories and data treatments, as expected given that the dichotomous variable was treated identically across all models, as categorical and with an estimated random intercept and random linear slope. For all numbers of response categories and data treatments, a linear growth model provided the best fit for the conditional response,  $m$ , with a random intercept and slope.

Given three response categories, the correlation between the random intercepts of the two model parts if  $m_{ij}$  was treated as continuous was .33 (90% CI = [.06, .59]) and if treated as categorical was .33 (90% CI = [.08, .57]). Given five response categories, the correlation between the random intercepts of the two model parts if  $m_{ij}$  was treated as continuous was .71 (90% CI = [.46, .95]) and if  $m_{ij}$  was treated as categorical was .79 (90% CI = [.53, 1.05]). Given seven response categories and continuous treatment of  $m_{ij}$ , the correlation between the random intercepts of the two model parts was .61 (90% CI = [.39, .85]). With  $m_{ij}$  evaluated as a categorical outcome, the correlation was .68 (90% CI = [.68, .92]). Across the numbers of response categories, the correlations were slightly higher if  $m_{ij}$  was treated as categorical. Across all results, these correlations suggest a moderately strong association between the tendency to engage in drinking and the frequency of engagement at age

15. Across the different number of response categories, given the same data treatment, parameter estimates were generally consistent.

Similar to the other methods that treated the response variable as continuous, a two-part growth model was fit assuming heterogeneity of the residual variances by wave for the conditional continuous response. Unlike the results based on other methods, likelihood ratio tests suggested that the residual variances were equal by wave under a two-part growth model for the outcome based on five and seven response categories,  $\chi^2(2) = 1.64, p = .44$ , and  $\chi^2(2) = 1.27, p = .53$ , respectively. For three response categories, the test suggested heterogeneity of variance by wave,  $\chi^2(2) = 34.05, p < .001$ . These results suggest that if treating the outcome as a continuous variable, a two-part growth model might be preferred over the other methods considered here, particularly if the number of response categories is at least five.

## DISCUSSION

The use of ordinal rating scales in survey research often creates a data analysis problem that is complex relative to those involving continuous data. Further, some behavioral data might also include a large proportion of zeros, and thus, might not be well matched to statistical methods for ordinal response data that assume an underlying normal response. Methods for ordinal data might also be difficult to implement in practice, leading some researchers to treat ordinal data as continuous. Indeed, such a strategy has been shown to be useful if there are four or more response categories. However, if a response departs far from a normal distribution, treating the response as continuous could bias parameter estimates no matter how many response categories are observed (Rhemtulla et al., 2012). This study investigated the utility of various longitudinal methods that were applied to nonsymmetric, ordinal response data with different numbers of response categories and that also included many zeros at each measurement occasion.

## Number of Response Categories and Treatment of Outcome Variable

In the drinking example provided, no direct comparisons could be made between treating the response variable as continuous versus categorical, although some comments should be made about the parameter estimates obtained under the different treatments. Under HLM, parameter estimates varied with respect to the number of response categories, as expected (see Figure 3). Under HGLM the estimates varied less, as shown in Figure 4, in which the estimates are nearly indistinguishable across the numbers of response categories. Thus, assuming a hierarchical model is preferred to other methods, the categorical treatment provided the most consistent results, and so, might be preferred over treating the data as continuous. Relative to other methods, the HLM/HGLM method might have underestimated drinking levels of some adolescents.

Assuming multiple latent classes by using GMMM, the results suggested a greater degree of variation in drinking relative to that suggested by HLM or HGLM. Treating the response as continuous under GMMM, the results suggested a wide range in drinking frequency at age 15 with remarkably less variation by age 20, a finding that was consistent across the numbers of latent classes and response categories. Further, results from analysis of the 5- and 7-point response scales resulted in relatively higher initial drinking levels than were suggested from results based on the 3-point response scale. These results were counter to those obtained if the response was treated as categorical, however, in which initial levels were highest for the 3-point scale and relatively low for the 5- and 7-point scales. Similar patterns in the data resulted from LCGA, a finding that was not surprising given that the variances of the random coefficients under GMMM were small and in some cases not statistically different from zero.

As the original data were collected on an ordinal 7-point scale, we considered as a reference the models that treated the original 7-point response as categorical. Generally, it appears that there is not a great loss in information if a 5-point scale was used. Speaking to whether categorical outcomes could be treated as continuous without a loss of information given five or more response categories, the example presented here suggests that for data that include a high proportion of zeros, there can be a difference in the interpretation of results from the two treatments. Most notable is the difference in the variation of drinking level estimates for models that included latent classes. Specifically, treating the response as continuous suggested less heterogeneity in drinking relative to treating the response as categorical.

In general, if an outcome variable is treated as categorical, the parameter estimates are easily interpreted, such as the intercept of a longitudinal model as the log-odds of engaging in the outcome at the given time point. This interpretability might be a benefit of treating an ordinal variable as categorical. Under a continuous data treatment,

the parameter estimates become less interpretable, as the response is actually measured on a scale that was intended to relate to a latent scale. For example, by treating the response as continuous, the parameter estimates from our empirical example did not offer clear information, such as estimates of an intercept that relates to an individual's expected alcohol use at a given time. The estimates can be interpreted only relative to other time points (e.g., frequency of use increases with age).

The two-part latent growth model appears to have produced relatively consistent results across the two data treatments and numbers of response categories relative to the other methods. That is, parameter estimates were relatively less sensitive to the number of response categories and to whether the conditionally positive responses were treated as categorical or continuous. Interestingly, the estimated residual variances for the conditionally positive responses were best fit by a model that assumed homogeneity of variance across occasions. Under all other methods, the residual variances were relatively dissimilar across the three time points. This finding might suggest the need to consider two-part growth models for data that include a sizable proportion of zeros. However, such conclusions should be drawn with caution, as the results are limited to the empirical data used in this study and might not be generalizable to other problems. The two-part growth model was not evaluated by Feldman et al. (2009) but might be a reasonable approach to handle a high proportion of zeros, an attribute of data that is common in psychological and behavioral data.

The limitations of this study might include the use of empirical data without the use of simulated data to more fully evaluate the utility of the methods considered. The empirical data evaluated here were measured on ordinal scales and included many zeros. The type of distributions that generated such data might be difficult to mimic, particularly given that different types of zeros could be present in empirical data. That is, a zero response might denote that an individual did not engage in the behavior at a particular occasion but has been known to engage at other times. A zero response could also denote no engagement for an individual who does not ordinarily engage in the behavior. Thus, using an empirical example might afford a more realistic way to evaluate methods for such data because the underlying mechanism that generated the zeros is not likely to be known. Future studies could still benefit from simulated data for which different scenarios could be studied.

This study used different methods for handling data with lots of zeros, including most that were evaluated by Feldman and colleagues, in addition to a two-part latent growth model. Although the methods considered here might handle the characteristics of the data in different ways, we do not believe that any of these approaches is yet ideal for analyzing ordinal data with many zeros. Even the two-part latent growth model, which was intended to handle the high proportions of zeros, has an assumption of normality for the

positive values, and was originally intended to handle continuous, not categorical data. Thus, more research is needed to define an appropriate tool for handling such data.

In summary, despite the availability of advanced methods for the analysis of longitudinal categorical data, applications of these models to real data can present challenges. Whether to use continuous or categorical methods might not always be clear. One contributing factor could be that at this point in time there is no fit statistic that allows for a direct comparison between a model that treats a variable as continuous and a corresponding model that treats the same variable as categorical. Most fit statistics, including the widely used AIC and Bayesian information criterion, are based on a log-likelihood function that is defined differently for continuous and discrete distributions, making it difficult to draw conclusions about which approach might be most appropriate for a given data set.

### ACKNOWLEDGMENTS

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill and funded by Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health Web site (<http://www.cpc.unc.edu/addhealth>). No direct support was received from Grant P01-HD31921 for this article.

### REFERENCES

- Bentler, P., & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16*, 78–117. doi:10.1177/0049124187016001004
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147–158. doi:10.1037/0033-2909.101.1.147
- DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling, 9*, 327–346. doi:10.1207/S15328007SEM0903\_2
- Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5, and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47*, 309–326. doi:10.1111/j.2044-8317.1994.tb01039.x
- Duan, N., Manning, W. G., Jr., Morris, C. N., & Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics, 1*, 115–126.
- Feldman, B., Masyn, K., & Conger, R. (2009). New approaches to studying problem behaviors: A comparison of methods for modeling longitudinal, categorical adolescent drinking data. *Developmental Psychology, 45*, 652–676. doi:10.1037/a0014851
- Harris, K. M., & Udry, J. R. (1994–2008). *National Longitudinal Study of Adolescent Health (Add Health), 1994–2008* (ICPSR21600-v11). Ann Arbor, MI: Interuniversity Consortium for Political and Social Research. doi:10.3886/ICPSR21600.v11
- 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
- Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological methods, 11*(1), 36.
- Laska, M. N., Pasch, K. E., Lust, K., Story, M., & Ehlinger, E. (2009). Latent class analysis of lifestyle characteristics and health risk behaviors among college youth. *Prevention Science, 10*, 376–386. doi:10.1007/s11121-009-0140-2
- Li, F., Duncan, T. E., Duncan, S. C., & Acock, A. (2001). Latent growth modeling of longitudinal data: A finite growth mixture modeling approach. *Structural Equation Modeling, 8*, 493–530. doi:10.1207/S15328007SEM0804\_01
- Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity within factor mixture models. *Psychological Methods, 10*, 21–39. doi:10.1037/1082-989X.10.1.21
- Muthén, B., & Muthén, L. (2000). The development of heavy drinking and alcohol-related problems from ages 18 to 37 in a U.S. national sample. *Journal of Studies on Alcohol, 61*, 290–300.
- Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Nylund, K., Bellmore, A., Nishina, A., & Graham, S. (2007). Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development, 78*, 1706–1722. doi:10.1111/j.1467-8624.2007.01097.x
- Olsen, M. K., & Schafer, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association, 96*, 730–745. doi:10.1198/016214501753168389
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.
- Rhemtulla, M., Brosseau-Liard, P., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under sub-optimal conditions. *Psychological Methods, 17*, 354–373. doi:10.1037/a0029315
- Su, L., Tom, B. D., & Farewell, V. T. (2009). Bias in 2-part mixed models for semicontinuous longitudinal data. *Biostatistics, 10*, 374–389. doi:10.1093/biostatistics/kxn044
- 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. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In S. Menard (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis* (pp. 373–385). Burlington, MA: Elsevier. doi:10.1080/01621459.1996.10476679