A primer on distributional assumptions and model linearity in repeated measures data analysis

Yadira Peralta [a], Nidhi Kohli [a, ✉] & Chun Wang [b]

[a]Department of Educational Psychology, College of Education & Human Development, University of Minnesota
[b]University of Washington

Abstract ∎ Repeated measures data are widely used in social and behavioral sciences, e.g., to investigate the trajectory of an underlying phenomenon over time. A variety of different mixed-effects models, a type of statistical modeling approach for repeated measures data, have been proposed and they differ mainly in two aspects: (1) the distributional assumption of the dependent variable and (2) the linearity of the model. Distinct combinations of these characteristics encompass a variety of modeling techniques. Although these models have been independently discussed in the literature, the most flexible framework – the generalized nonlinear mixed-effects model (GNLMEM) – can be used as a modeling umbrella to encompass these modeling options for repeated measures data. Therefore, the aim of this paper is to explicate on the different mixed-effects modeling techniques guided by the distributional assumption and model linearity choices using the GNLMEM as a general framework. Additionally, empirical examples are used to illustrate the versatility of this framework.

Keywords ∎ repeated measures data, distributional assumptions, model linearity.. **Tools** ∎ SAS, R.

✉ nkohli@umn.edu

YP: 0000-0003-4823-6939; *NK*: 0000-0003-4690-2854; *CW*: 0000-0003-2695-9781

Introduction

Correlated data structure arises as a consequence of research design or as an inherent characteristic of the population (Fitzmaurice, Laird, & Ware, 2011). Repeated measures data are one particular type of correlated data where measurements of the same individuals are gathered over time and/or in different experimental conditions (Fitzmaurice et al., 2011). The assumption of independence holds at the between-subjects level (inter-individual or level-2), but the repeated measures of the same individual over time are dependent rather than independent. The within-subject (intra-individual or level-1) observations are correlated and such correlation must be taken into account in order to get valid inferences from the model (Fitzmaurice et al., 2011). An example of this study design in educational research is the widely used Early Childhood Longitudinal Study—Kindergarten (ECLS-K) dataset where a nationally representative cohort of children who entered kindergarten in 1998-1999 school year were followed

through middle school (Tourangeau, Nord, Lê, Sorongon, & Najarian, 2009). During that period, data on students' background and academic achievement were collected on seven measurement occasions, where the level-1 consists of time (seven instances of data collected for each student) and the level-2 of students (data gathered across students). The data collected on the same student across seven time-points are dependent and likely correlated (within-subject or level-1 correlation), whereas the data from different students (between-subjects or level-2) are assumed to be independent.

A major advantage of repeated measures data is that it can capture patterns of change over time, that is, the growth trajectory of the individuals and the independent variables that predict the trajectory (Fitzmaurice et al., 2011; Stroup, 2013). Several statistical models have been proposed to analyze this type of data. Researchers in diverse fields such as education, sociology, psychology, statistics, and biostatistics have contributed to the development of new methods for the analysis of repeated measures

data. Although there is abundant literature showing a specific type (or types) of models that work well for the analysis of different kinds of repeated measures datasets, there is lack of a didactic paper that synthesizes the connections and distinctions among these models. Hence, oftentimes, users of these models are left wondering which of these unconnected modeling alternatives to choose to analyze their data at hand. However, model differences mainly come from two elements: (1) the distributional assumptions of the outcome variable, and (2) the model linearity to describe the underlying trajectory over time, which depends on the functional form.[1] Having a general scheme to learn about statistical models for repeated measures data considering these two main elements can provide a better understanding of the modeling options. This in turn could encourage researchers to use a wider variety of statistical techniques when appropriate.

Initially, Laird and Ware (1982) proposed to use linear mixed-effects models (LMEMs) for continuous, normal data to estimate both the average trajectory (i.e., fixed-effects) and the between-subject variation (i.e., random-effects), where normality was assumed at both levels of the errors (within- and between-subjects). Nevertheless, the amalgamation of random effects (both at level-1 and level-2) within the generalized linear models (GLMs) framework allowed the analysis of non-normal repeated measures data, introducing the generalized linear mixed-effects models; GLMEMs (Stiratelli, Laird, & Ware, 1984; Zeger, Liang, & Albert, 1988). Note that LMEMs are a special case of GLMEMs when the normality assumption of the outcome variable holds, just as the traditional linear model is a special case of the GLM when errors follow a normal distribution.

Earlier models mainly allowed for fixed and random effects that enter the model in a linear fashion, assuming level-1 and level-2 errors to be normally distributed. However, there were some developments in the field of pharmacokinetics that used a specific nonlinear model to describe repeated measures data (see e.g., Sheiner & Beal, 1980). Lindstrom and Bates (1990) proposed a general nonlinear mixed-effects model (NLMEM) for the analysis of repeated measures data. This approach introduced a general framework to use *intrinsically* nonlinear functions (e.g., exponential; see formal definition in the next section) to model changes over time of continuous data, assuming normal distribution at both levels of error. Furthermore, the LMEM can be considered a special case of the NLMEM (Lindstrom & Bates, 1990).

The combination of non-normal data and nonlinear

trajectories over time led to the extension of GLMEMs to include nonlinear (and linear) fixed and random effects (see e.g., Payne, 2014; Vonesh, 2012; Vonesh, Chinchilli, & Pu, 1996; H. Wu & Zhang, 2006). Such combination defined a general class of models, the generalized nonlinear mixed-effects models (GNLMEMs). The GNLMEMs bring together all different models for repeated measures data described above; all of them could be derived as special cases of this general framework because the differences between these approaches are due mostly to (1) the normality assumption of the outcome variable, and (2) the (non)linearity of the model. However, despite the practicality of this resourceful scheme, there has been little discussion around it in the social and behavioral sciences. This lack of dissemination intensifies the use of LMEMs under the assumption of normality when the nature of the outcome variable requires other modeling strategies. For instance, a wide majority of analyses in psychological research of count data are still assuming normality when the Poisson distribution would be preferred (Vives, Losilla, & Rodrigo, 2006).

The main objective of this article, therefore, is to describe the most popular mixed-effects models subsumed under the GNLMEM framework as a combination of two features: the distributional assumption and the model linearity to describe the mean trajectory of the dependent variable. To elaborate, the data at hand could be either normally or non-normally distributed, and the trend exhibited by the data could be linear, nonlinear, or intrinsically nonlinear. Combinations of data-type (e.g., continuous, binary, categorical, continuous and nonnegative, and count) and data-behavior (linear and nonlinear in parameters) will determine the model choice for appropriately describing the underlying structure. That is, modeling techniques are conditioned to the choice of distributional assumption and model linearity. In addition, the GNLMEM is the most general framework for repeated measures data analysis. Thus, we use it to enclose and derive from it the rest of the statistical approaches. We acknowledge that temporal dependencies in the within-subject variability (level-1 errors) are possible for each of the models discussed in the article. However, a discussion of this feature is out of the scope of the study.

Our intent is to communicate to a broader audience of methodologists, substantive researchers and practitioners that the GNLMEM framework constitutes a resourceful tool that assembles diverse statistical techniques for the description of a variety of repeated measures data. We advocate that having this broad framework of statistical models will help disseminate the different modeling options for

---

[1]Another option to model repeated measures data is within the nonparametric and semiparametric frameworks where the mean trajectory is not specified *apriori* but rather estimated by using smoothing procedures (Fitzmaurice et al., 2011). Nonparametric and semiparametric models are out of the scope of this article.

these types of data. We note that the LMEMs, the NLMEMs, the GLMEMs, and the GNLMEMs have been discussed and presented elsewhere separately (e.g., Davidian & Giltinan, 1995; Stroup, 2013; Vonesh, 2012; Vonesh et al., 1996; H. Wu & Zhang, 2006); however, our contribution is unique in that we integrate different modeling approaches for repeated measures data under the GNLMEM scheme. Researchers can significantly benefit from having a general scheme that reminds them of all the modeling possibilities for these types of data. Available software packages for different submodels of GNLMEMs are also provided for broad audience with different preferences.

The organization of this paper is as follows. In Section 2 we present the GNLMEM framework, and use it to derive other mixed-effects models classified by two main features: (1) the distributional assumptions of the outcome variable, and (2) the model linearity characterized by the functional form used to describe temporal change. In Section 3 we briefly discuss the estimation capabilities in different software packages. We then present empirical examples in Section 4 to illustrate the practical utility of viewing mixed-effects models as a combination of their foremost characteristics using the GNLMEM as an umbrella approach. And finally we reemphasize the importance of having a flexible umbrella framework for repeated measures data in Section 5.

Model

Following Vonesh et al. (1996), Stroup (2013), and Anderson, Verkuilen, and Johnson (2013), the GNLMEM can be expressed as

$$
\begin{aligned}
\boldsymbol{y}_i | \boldsymbol{b}_i &\sim G(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \\
g(\boldsymbol{\mu}_i | \boldsymbol{b}_i) &= \boldsymbol{\eta}_i \\
\boldsymbol{\eta}_i &= f(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i) \\
\boldsymbol{b}_i &\sim N(\boldsymbol{0}, \boldsymbol{\Phi})
\end{aligned}
\tag{1}
$$

where:

- $\boldsymbol{y}_i$ is a $n_i \times 1$ vector of repeated measures on subject i (n_i is the number of time points for subject i);
- $\boldsymbol{b}_i$ is a $r \times 1$ vector of r subject-specific random effects, which follows a normal distribution with mean $\boldsymbol{0}$ and (co)variance $\boldsymbol{\Phi}$;
- $G(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is a general distribution function with mean $\boldsymbol{\mu}_i$ and variance-covariance matrix $\boldsymbol{\Sigma}_i$;
- $g(\boldsymbol{\mu}_i | \boldsymbol{b}_i)$ is a link function that connects the mean of the response variable, $\boldsymbol{\mu}_i$, with the (non)linear predictor, $\boldsymbol{\eta}_i$, given the vector of random effects, $\boldsymbol{b}_i$;
- $\boldsymbol{\eta}_i$ is the (non)linear predictor, which depends on the subject only through its dimension n_i;
- $f(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$ is a general $n_i \times 1$ (non)linear vector function;
- $\boldsymbol{X}_i$ is a $n_i \times p$ design matrix of p independent variables;

- $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects.

This model represents a comprehensive framework to analyze correlated data. For instance, the general distribution $G(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ could be either normal or non-normal (e.g., binomial, gamma, Poisson), which also corresponds to different forms of the link function, $g(\boldsymbol{\mu}_i | \boldsymbol{b}_i)$. The functional form, $f(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$, can be either linear or nonlinear (see definitions of model linearity in a subsequent subsection). For example, mixed-effects models assuming normality for both levels of the error structure, which have been studied extensively, represent a special case of the GNLMEM framework. In this case $G(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is the multivariate normal density, and $g(\boldsymbol{\mu}_i | \boldsymbol{b}_i)$ is the identity link. Due to the versatility and flexibility of the GNLMEM, it provides a series of modeling options for repeated measures data from which researchers can choose according the type of data at hand and its behavior. Special cases derived from Model (1) along with examples of their applications are presented below categorized by its main features: (1) the distributional assumption of the response variable, and (2) the model linearity.
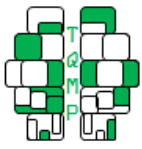
Distributional Assumptions

If the distribution of the outcome variable is a member of the natural exponential family, a subset of the exponential family, or similar to the exponential family, then the model belongs to the generalized modeling framework (Anderson et al., 2013; McCullagh & Nelder, 1989; Neuhaus & McCulloch, 2011). The generalized modeling framework encompasses a variety of distributional assumptions for the response variable. The choice of the distributional assumption depends entirely on the characteristics of the outcome variable (Anderson et al., 2013; McCullagh & Nelder, 1989), as it will be described below. In the subsequent sections, we provide examples of some common distributional assumptions.

Normal distribution. When the outcome variable is continuous and normally distributed, mixed-effects models as described by Laird and Ware (1982) or Lindstrom and Bates (1990) can be easily obtained from Model (1). We can simply replace the general distribution function "G" in Model (1) by a multivariate normal density function, and choose the link function "g" to be an identity link, such that Model (1) takes the form of

$$
\begin{aligned}
\boldsymbol{y}_i | \boldsymbol{b}_i &\sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) \\
\boldsymbol{\mu}_i &= \boldsymbol{\eta}_i \\
\boldsymbol{\eta}_i &= f(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i) \\
\boldsymbol{b}_i &\sim MVN(\boldsymbol{0}, \boldsymbol{\Phi})
\end{aligned}
\tag{2}
$$

Non-normal distributions. There are several non-normal distributions that could be considered under the

GNLMEM modeling framework, e.g., gamma, exponential, binomial, beta, Poisson, geometric, negative binomial. We will focus on the distribution of the three most common response variables: dichotomous data, count data, and nonnegative response data.

Binomial distribution. The binomial distribution applies to cases where there is a series of trials of independent dichotomous observations. In such cases, the conditional distribution of the vector of repeated measures in Model (1) follows a binomial distribution, i.e., $\boldsymbol{y}_i|\boldsymbol{b}_i \sim$ Binomial$(n_i, \boldsymbol{\pi}_i)$ where $\boldsymbol{\pi}_i$ is a $n_i \times 1$ vector that denotes the probability of success of the repeated measures of subject i. The most common link function is the logit of the mean, or equivalently, the logit of the success probability. In Model (1), this is represented by $g(\boldsymbol{\mu}_i)|\boldsymbol{b}_i) = \text{logit}(\boldsymbol{\pi}_i/(1-\boldsymbol{\pi}_i))$ (Anderson et al., 2013; Stroup, 2013). There is a body of research using generalized models for binary data, also called mixed logit models. For instance, researchers have used these models to analyze the accuracy of language production (Middleton, Chen, & Verkuilen, 2015), memory performance (Murayama, Sakaki, Yan, & Smith, 2014), factors that affect children's language comprehension (Jaeger, 2008), to investigate sources of difficulty in language production and comprehension in children and adults (Arnon, 2010), and to study the effect of different treatments in smoking cessation over time (Hedeker, 2005). Moreover, one-parameter or two-parameter item response models (IRT) could be seen as generalized logistic mixed models as well (Hedeker, 2005).

Poisson distribution. The Poisson distribution is used to model count data, and it is especially useful to model events with low probability of occurrence such as the number of times a student is nominated by his peers as a bully (Anderson et al., 2013). In this case, the conditional distribution of the vector of repeated measures in Model (1) follows a Poisson distribution with parameter $\boldsymbol{\mu}_i$, i.e., $\boldsymbol{y}_i|\boldsymbol{b}_i \sim \text{Pois}(\boldsymbol{\mu}_i)$, where $\boldsymbol{\mu}_i$ refers to both the mean and variance of the distribution. The most common link function is the natural logarithm, $g(\boldsymbol{\mu}_i|\boldsymbol{b}_i) = \log(\boldsymbol{\mu}_i)$ (Anderson et al., 2013; McCullagh & Nelder, 1989; Stroup, 2013). Count data is nonnegative, and usually unbounded (it can take any nonnegative value) and positively skewed (Anderson et al., 2013; Atkins & Gallop, 2007). Examples of variables that have been analyzed assuming a Poisson distribution are marital commitment measured as the number of steps taken towards divorce (Atkins & Gallop, 2007); or school violence measured as the number of bullying incidents or fights a student may experience in a period of time (Huang & Cornell, 2012). Bolger and Patterson (2001) investigated social rejection by peers (measured by years of rejection) as a function of maltreatment, aggression, and

withdrawal. Despite these examples and the fact that count data are highly present in social and behavioral sciences, there is a lack of knowledge and understanding of the use of Poisson models in psychology (Vives et al., 2006).

Gamma distribution. The gamma distribution is helpful to model skewed, continuous non-negative data where the variance is an increasing function of the mean (Anderson et al., 2013; McCullagh & Nelder, 1989). In this case, the conditional distribution of the outcome variable given the random effects in Model (1) will follow a gamma distribution, $\boldsymbol{y}_i|\boldsymbol{b}_i \sim \text{Gamma}(\boldsymbol{\mu}_i, \theta)$, where $\boldsymbol{\mu}_i$ is the mean response vector of subject i and θ is the dispersion parameter from an exponential family parameterization, which for the gamma distribution equals the reciprocal of the shape parameter (Anderson et al., 2013; McCullagh & Nelder, 1989). There are two common link functions for the gamma distribution, the inverse link, $g(\boldsymbol{\mu}_i|\boldsymbol{b}_i) = 1/\boldsymbol{\mu}_i$, and the natural logarithm, $g(\boldsymbol{\mu}_i|\boldsymbol{b}_i) = \log(\boldsymbol{\mu}_i)$. The use of the inverse link suggests an additive model, while the natural logarithm link advocates for a multiplicative one. The choice of the link function depends on the nature of the outcome variable, theoretical considerations, or model fit (Anderson et al., 2013; Stroup, 2013). There are several examples in behavioral sciences that might be analyzed assuming a gamma distribution. For instance, reaction time in cognitive or neuropsychological tests usually follows a non-negatively skewed distribution and thus it is suitable for this model.

Model Linearity

The main purpose of repeated measures data analysis is to accurately capture the pattern of change in the phenomenon of interest (Fitzmaurice et al., 2011). The data structure may either follow linear (straight-line), nonlinear (e.g. quadratic or cubic), or intrinsically nonlinear patterns of change (e.g., exponential or piecewise). Thus the model can be broadly categorized as linear or intrinsically nonlinear in parameters (see definitions below). The choice between linear and nonlinear patterns of change over time can be either based on the physical characteristics of the phenomenon or on an empirical approach (Cudeck & Harring, 2007; Davidian & Giltinan, 1995). The choice of the functional form will be reflected in the term $f(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$ in Model (1), and this will determine the model linearity.

Linear model. Following the definition of Bates and Watts (1988), a linear model is one for which the first-order derivatives of $f(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$ "with respect to any of the parameters are independent of all the parameters" (Bates & Watts, 1988, p. 2). Hence, some nonlinear relationships still can be modeled under the linear model framework. For instance, quadratic, cubic or higher-order polynomials functional forms are examples of models that are still

Table 1 ∎ Schematic representation of models for repeated measures data as a combination of distributional assumption and model linearity

Model	Distributional Assumption	
Linearity	Normal	Non-normal
Linear	LMEM	GLMEM
Nonlinear	NLMEM	GNLMEM

Note. LMEM = linear mixed-effects model; NLMEM = nonlinear mixed-effects model; GLMEM = generalized linear mixed-effects model; GNLMEM = generalized nonlinear mixed-effects model.

linear in parameters but that allow to describe curvilinear relationships (Locascio & Atri, 2011).

When a steady pattern of change does not seem plausible, there is also the option of transforming the raw data and fit a linear model. Data transformation could be the result of several scenarios; for example, to fulfill statistical model assumptions, to make the data easier to visualize, or to have more interpretable results. For instance, time logarithmic transformations are widely used in social sciences (e.g., Shin, Davison, Long, Chan, & Heistad, 2013; Van Breukelen, 2005). Likewise, polynomial or fractional polynomial transformations can be used to model nonlinearity under a linear framework (Long & Ryoo, 2010). However, depending on the transformation used, this might help or actually complicate parameter interpretation (Long & Ryoo, 2010), or additional steps are involved to transform back the estimates to the original scale.

Nonlinear model. Contrary to the linear model, in the intrinsically nonlinear model "at least one of the derivatives of the expectation function with respect to the parameters depends on at least one of the parameters" (Bates & Watts, 1988, p. 32). Some examples of inherent nonlinearity are the exponential function, logistic function, and piecewise function. Nonlinear models are usually related to unsteady change over time, asymptotic behaviors, abrupt increments or decrements, and mostly to the necessity of having estimated parameters easy to interpret in the context of a phenomenon under investigation (Cudeck & Harring, 2007; Davidian & Giltinan, 1995). Examples where researchers have used nonlinear models are language development in children (e.g., Cudeck & Harring, 2007), mathematics achievement over time (e.g., Kohli, Sullivan, Sadeh, & Zopluoglu, 2015), responses to treatment intervention (e.g., Cudeck & Harring, 2007; Kreisman, 2003), and skill acquisition such as reading or verbal learning (e.g., Cudeck, 1996; Cudeck & Harring, 2007; Silverman, Speece, Harring, & Ritchey, 2013).

Interim Summary

Summarizing, distributional assumptions and model linearity are the two principal characteristics that determine the statistical model to describe repeated measures data.

Model choice, enclosing probability distribution and functional form, is contingent on the characteristics and behavior of the data. The general framework of GNLMEMs serves as an umbrella to encompass several modeling alternatives. The LMEM, for instance, is derived from Model (1) when normality is assumed and a linear model is proposed. The NLMEM represents the combination of normal distribution of the outcome variable and an intrinsically nonlinear functional form to describe trajectories. The GLMEM relaxes the normality assumption while still modeling a linear change over time; and the GNLMEM describes non-normal data and nonlinear change over time. Table 1 presents these combinations in a schematic manner.

Estimation Capabilities of Different Software Packages

Parameter estimation is non-trivial for repeated measures data models, especially for NLMEMs and GNLMEMs. The main estimation challenge arises as a consequence of no closed form solution to the marginal likelihood function. Several estimation methods have been proposed to overcome this numerical challenge. In this section we briefly discuss the most common methods and mention their availability in statistical packages. Estimation methods for LMEMs are well documented elsewhere (e.g., Bates & DebRoy, 2004; Fitzmaurice et al., 2011; Laird & Ware, 1982). As such, they will not be the focus of the discussion in this section.

Direct maximization techniques are widely used to estimate models of the GNLMEM class. In this approach, an approximation of the marginal likelihood function is directly maximized. Gauss-Hermite (GH) quadrature and adaptive Gaussian (AGH) quadrature are two methods broadly used for this purpose (Davidian & Giltinan, 2003; Tuerlinckx, Rijmen, Verbeke, & De Boeck, 2006). These methods approximate the integral deterministically by a weighted average of the integrand evaluated at a certain number of quadrature points; such points could be fixed in advance (GH) or determined adaptively at each iteration cycle (AGH). Another type of approximation to the likelihood is the Monte Carlo (MC) integration. In this approach, the numerical approximation is stochastic and it is done by simulating a set

Table 2 ∎ Estimation methods and algorithms for repeated measures data models available in commercial software routines
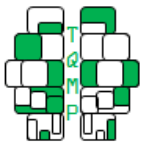
Estimation method and algorithm	Statistical model for repeated measures data			
	LMEM	NLMEM	GLMEM	GNLMEM
Maximum likelihood or restricted maximum likelihood				
GH	SAS-NLMIXED	SAS-NLMIXED	SAS-GLIMMIX SAS-NLMIXED	SAS-NLMIXED
AGH	SAS-NLMIXED HLM7 Stata-gllamm	SAS-NLMIXED R (lme4) nlmer Stata-gllamm	SAS-NLMIXED R (lme4) glmer R (glmmML) glmmML HLM7 Stata-gllamm	SAS-NLMIXED
FIRO	SAS-NLMIXED	SAS-NLMIXED	SAS-NLMIXED	SAS-NLMIXED
NR		R (lme4) nlme		
Laplace	SAS-NLMIXED HLM7	SAS-NLMIXED R (lme4) nlmer	SAS-GLIMMIX SAS-NLMIXED R (lme4) glmer R (glmmML) glmmML HLM7	SAS-NLMIXED
PQL	HLM7		SAS-GLIMMIX R (MASS) glmmPQL HLM7	
MQL			SAS-GLIMMIX	
PL			SAS-GLIMMIX	
REPL			SAS-GLIMMIX	
PLS	R (lme4) lmer			
Bayesian estimation				
MC			R (glmm) glmm	
MCMC	JAGS	JAGS	JAGS	JAGS

Note. LMEM = linear mixed-effects model; NLMEM = nonlinear mixed-effects model; GLMEM = generalized linear mixed-effects model; GNLMEM = generalized nonlinear mixed-effects model. GH = Gauss-Hermite quadrature, AGH = adaptive Gaussian quadrature, FIRO = first-order approximation, NR = Newton-Raphson, Laplace = Laplace approximation, PQL = penalized quasi-likelihood, MQL = marginal quasi-likelihood, PL = pseudo-likelihood, REPL = restricted pseudo-likelihood, PLS = penalized least squares, MC = Monte Carlo integration. R packages to which the functions correspond are in parenthesis.

of values from certain distributions (usually certain priors or density functions) and then approximating the marginal likelihood by a simple average (Tuerlinckx et al., 2006). These techniques (GH, AGH, and MC) are often referred as direct maximization of the "exact" likelihood function, since the maximization process is implemented on the numerical approximation of the likelihood. Additionally, the accuracy of the approximation doesn't depend on the number of observations by subject, as it is the case for some of the techniques described below (Davidian & Giltinan, 2003).

Another popular method for an analytical approximation to the likelihood function is called first-order approximation method (FIRO). FIRO approximates the model of the outcome variable as a Taylor series expansion with respect to the random effects evaluated at zero. Assuming this linear approximation is true, the estimation is performed using available methods for LMEMs (Davidian & Giltinan, 2003; Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006; SAS Institute Inc., 2008). Another alternative is the Laplace approximation, which avoids intractable integration by manipulating the integrand of the likelihood function using a second order Taylor series expansion (Raudenbush, Yang, & Yosef, 2000). Notice that the focus of the direct maximization techniques mentioned in the previous paragraph is to approximate the integral, whereas FIRO approximates the model and Laplace manipulates the integrand of the likelihood function.

Quasi-likelihood approaches emerged as an estimation alternative when sufficient statistics of the data, such as the mean and the variance, are used to construct a function similar to the likelihood but that does not use the exact distributional specification and performs as well or nearly as well as the likelihood function (Neuhaus & McCulloch, 2011; Tuerlinckx et al., 2006). The quasi-likelihood approaches below differ in their way of approximating the likelihood function. For instance, Breslow and Clayton (1993) recommended the use of two estimation techniques for GLMEMs, penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL). Both techniques use a Laplace approximation of the likelihood function. However, while the PQL uses a linear expansion about the current estimate of the fixed effects and about the posterior mode of the random effects; the MQL uses a quadratic expansion about the current estimate of the fixed effects and about the random effects equal to zero (Tuerlinckx et al., 2006). Separately, Wolfinger and O'Connell (1993) introduced a similar methodology that they called pseudo-likelihood (PL), and when the PL estimation is performed using restricted maximum likelihood (REML), it is called restricted pseudo-likelihood (REPL). The main difference between the techniques of Breslow and Clayton and Wolfinger and O'Connell is how they handle the estimation of the dispersion parameter of the exponential family of distributions (Littell et al., 2006).

For a thorough discussion of estimation methods for these models we refer readers to Fitzmaurice et al. (2011), Bates and DebRoy (2004), Davidian and Giltinan (1995, 2003), Tuerlinckx et al. (2006) and L. Wu (2010). Several statistical packages have implemented different estimation methods and algorithms for repeated measures data models. Table 2 shows some of the available routines in different statistical software. To estimate GLMEMs, the `GLIMMIX` SAS procedure implements PL, REPL, PQL, and MQL techniques (Littell et al., 2006), and the `gllamm` routine in Stata uses the AGH algorithm (Rabe-Hesketh, Skrondal, & Pickles, 2002). The HLM7 statistical package offers PQL, Laplace and AGH algorithms to fit LMEMs and GLMEMs (Raudenbush, Bryk, Cheong, Congdon, & de Toit, 2011). Additionally, the `glmmPQL` R function in the `MASS` package also offers PQL estimation for GLMEMs, the `glmer` and `glmmML` R functions in the `lme4` and `glmmML` packages, respectively, offer Laplace and AGH quadrature algorithms. The R functions `nlmer` (in the `lme4` package) and `nlme` (in the `nlme` package) offer Laplace, AGH, and Newton-Raphson algorithms for the estimation of NLMEMs, and the function `lmer` in the `lme4` package estimates LMEMs following penalized least squares principles. The SAS procedure `NLMIXED` implements GH, AGH, FIRO, and Laplace estimation techniques to estimate

GNLMEMs (SAS Institute Inc., 2008). In terms of Bayesian estimation, the `glmm` function in the `glmm` R package implements MC integration to estimate GLMEMs, and the program JAGS uses Markov Chain Monte Carlo (MCMC) simulation where any type of GNLMEMs could be estimated. As summary, the two programs that could potentially cover all models within the GNLMEM framework are SAS and JAGS.

Empirical Examples

To illustrate the practicality of choosing a statistical model to describe repeated measures data based on distributional assumptions and model linearity under the GNLMEM specification in Model (1), we will present two empirical examples: (1) behavioral data from a spatial attention experiment (Jiang & Swallow, 2013), and (2) educational data from the ECLS-K. We first introduce the data and descriptive statistics of the variable of interest, then we choose a model. Two software programs were used to fit the models discussed in this section, R and SAS. The SAS procedure NLMIXED was used in both empirical examples, whereas the function `glmer` from `lme4` R package was used for the first example, and JAGS through R was used for the second example. The commented SAS and R code is provided in the Appendix. In what follows, we discuss the estimation results of the SAS procedure `NLMIXED` because of its flexibility to accommodate a variety of modeling requirements (SAS Institute Inc., 2008) such as the examples in this Section.

Analysis of the Spatial Attention Experiment

The data used for this illustration was collected by Jiang and Swallow (2013). The aim of the original study was to explore spatial attention under different viewer-centered and environmental-centered conditions (Jiang & Swallow, 2013). All participants were students between 18 and 35 years old with normal or corrected-to-normal acuity of vision. They passed a color blindness test, were tested individually, and unaware of the purpose of the experiment.

Participants had to identify a T target among 11 L distractors presented on a computer monitor by using an optical wireless mouse. Their search response time and the color of the T target were recorded. During the training phase, the T target was more often located in one of the quadrants (50% of the time). However, during the testing phase, the T target was randomly displayed over the quadrants. Additionally, participants experienced changes in body and/or head orientation during the testing phase to dissociate viewer-centered from environmental-centered reference frames. Participants completed 384 trials of training and 192 trials of testing. For the purposes of this exemplification, data from the training phase were used since the experiment conditions were more alike (Jiang &
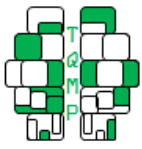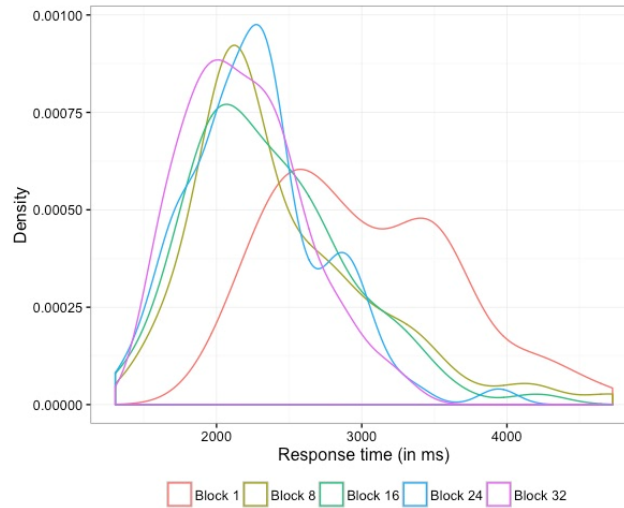
Figure 1 ■ Density plots of RT for five selected blocks for the spatial attention experiment.

Swallow, 2013) and such phase provided more data, which aids model estimation.

Outcome variable. The search response time (RT) data consisted of 32 blocks of 12 trials each per participant. The dependent variable for this illustration is the mean time to respond for each of the 32 blocks. There are 77 participants in total with 32 RT means for each of them.

Descriptive statistics. Table 3 presents descriptive information about the mean RT along the 32 blocks. There is a decreasing RT trend from Block 1 to Block 32 (mean RT of Block 1 =3014 ms, and mean RT of Block 32 = 2190 ms). Likewise, the standard deviation reduces along the blocks ($SD_{Block1} = 627.62, SD_{Block32} = 406.30$, the standard deviation was computed over the 77 RT means at each block). Figure 1 presents density plots for five selected blocks (1, 8, 16, 24, and 32). The distribution of RT by block is positively skewed and it takes positive values only, so statistical analyses are required that take these characteristics of the data into account.

Figure 2 shows 35 individual trajectories randomly selected, and the mean trajectory is added at the top of them. As manifest in many learning experiments, RT exhibits a decline over repeated measures. It is evident that the search RT steadily decreases over blocks in this spatial attention experiment. It is clear that there is individual variation around the average intercept, but it is debatable, though, whether there are also individual differences around the slope.

Choosing a model. Given the descriptive statistics above, the variable under study is continuous, non-negative and positively skewed; and it presents a steady decreasing tra-

jectory. Note that we show a traditional approach to understand the distribution of the data. However, there are other statistical tools that interested readers could find useful such as the `fitdist` function in the `fitdistrplus` R package (Delignette-Muller & Dutang, 2015). The evident non-normality of the spatial attention data will be taken into account by assuming a gamma model, which seems plausible for this example for two reasons: (1) the variance of RT seems to decrease as mean RT decreases, signaling a potential relationship between the mean and variance, and (2) RT is a continuous non-negative variable (Anderson et al., 2013). Initial exploratory analyses allowed us to conclude that a linear trajectory was satisfactory to capture the trend in data. That is, using descriptive statistics and the nature of the variable of interest, we propose a linear model assuming non-normality to analyze the RT outcome variable. Such model falls under the category of GLMEMs in the characterization presented in Table 1. Moreover, following the general GNLMEM framework in Model (1), the gamma linear mixed-effects model can be written as:

$$\begin{aligned}
\boldsymbol{y}_i | \boldsymbol{b}_i &\sim \text{Gamma}(\boldsymbol{\mu}_i, \theta) \\
\log(\boldsymbol{\mu}_i) &= \boldsymbol{\eta}_i \\
\boldsymbol{\eta}_i &= \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i \\
\boldsymbol{b}_i &\sim N(\boldsymbol{0}, \boldsymbol{\Phi})
\end{aligned} \tag{3}$$

where $\boldsymbol{y}_i$ is a 32×1 vector of the search RTs on subject i, $\boldsymbol{X}_i$ is a 32×2 design matrix, $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ is the vector of fixed effects, $\boldsymbol{b}_i$ is a 2×1 vector of random effects associated to the intercept and slope, $\boldsymbol{Z}_i$ is a 32×2 design matrix associated to $\boldsymbol{b}_i$, $\boldsymbol{\eta}_i$ is the linear predictor, $g(\boldsymbol{\mu}_i | \boldsymbol{b}_i) = \log(\boldsymbol{\mu}_i)$
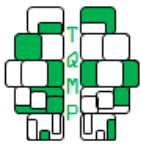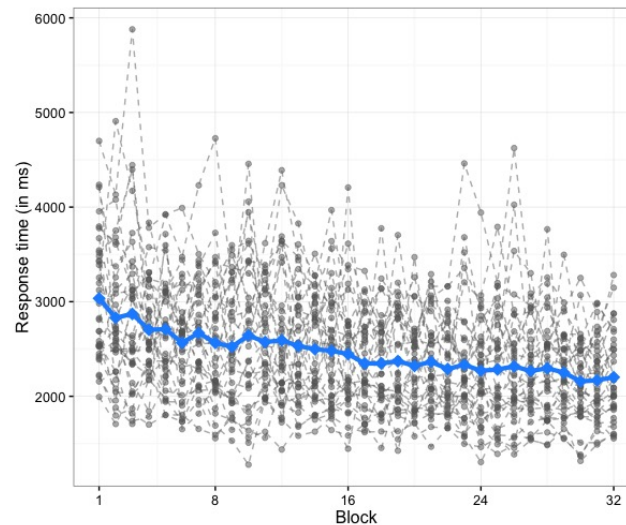
Figure 2 ■ Individual RT trajectories and observed mean of 35 randomly selected participants for the spatial attention experiment.

represents the natural logarithm link function, and θ represents the dispersion parameter. Mixed models incorporating a random effect for the intercept only (see Model 1 in Table 4) and considering random effects for both intercept and slope (see Model 2 in Table 4) were analyzed.

Results. Estimation results of this model are shown in Table 4. Figure 3 shows that the estimated mean trajectory closely resembles the observed mean trajectory in the data. Relying on the BIC fit statistics to compare covariance models (since the mean trajectory was the same), the variability in data could be adequately captured by the random intercept and slope model. Thus, only results from the model with random intercept and slope (Model 2 in Table 4) will be discussed next. However, notice that the variance of the random slope and the covariance between the random intercept and random slope were smaller than 1e-4, so they will be omitted in the discussion. For completeness purposes both model estimation results are presented in Table 4.

The parameter estimate of β_0 represents the average (model-scale) search RT ($\hat{\beta}_0 = 7.92, p < 0.001$), and of β_1 the average linear rate of change in ms per block ($\hat{\beta}_1 = -0.008, p < 0.001$). Notice that these estimates can be translated into the data-scale. For example, the intercept is given by $e^{\hat{\beta}_0+\hat{\beta}_1} = 2718.76$, and the factor of change is $e^{\hat{\beta}_1} = 0.99$, which means that the mean trajectory is actually decreasing 0.99 ms per block. The estimate of φ_{11} denotes the random variation around the intercept due to individual differences ($\hat{\varphi}_{11}$=0.03). Summarizing, on average,

the search RT steadily decreases without much variability around the rate of change. Differences between individuals are present around the starting RT.

Analysis of ECLS-K Data

The Early Childhood Longitudinal Study – Kindergarten Class of 1998-1999 (ECLS-K) is a nationally representative longitudinal study of U.S. students that started kindergarten in the 1998-1999 academic year. Students were followed over a period of nine years and data were collected at seven occasions: fall and spring of kindergarten and first grade, and spring of third, fifth, and eighth grades (time was coded accordingly as 0, 0.5, 1, 1.5, 3.5, 5.5, and 8.5, respectively). Information was collected from kindergartners, parents, schools and teachers across the U.S. The ECLS-K was designed to provide information about kids early school experiences and development during elementary and middle school, as well as their relationship with other factors at the individual or community levels (Tourangeau et al., 2009). The analytic sample for this illustration consists of a random sample of 400 subjects selected from students whose mathematics achievement scores were complete ($N = 2, 300$).

Outcome Variable. The dependent variable in this example is mathematics IRT (item response theory) scores. IRT scoring was used to generate comparable scores across grades with the purpose of capturing growth over time (Tourangeau et al., 2009).
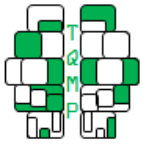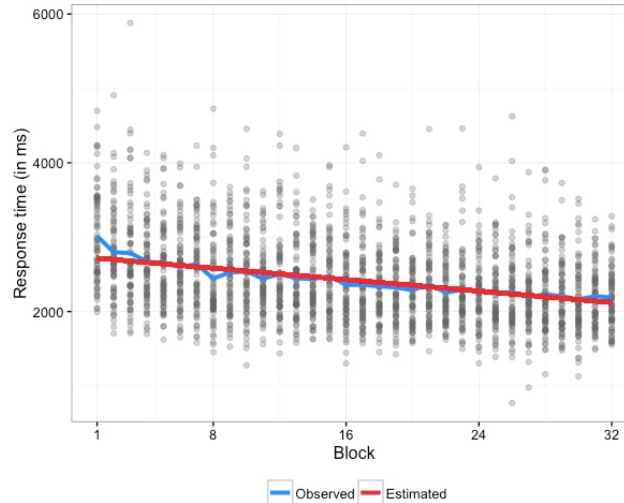
Figure 3 ■ Observed and estimated mean trajectories for the gamma mixed-effects model for the spatial attention experiment.

Descriptive statistics. Table 5 presents descriptive statistics for the math IRT scores at each grade. There is an increasing overall mean and dispersion in the data, and skew and kurtosis values are under normality assumptions. The pattern of change over time is illustrated in Figure 4, which shows 35 individual trajectories chosen at random along with their observed mean. The variable under examination presents an increasing trend across time. However, two distinct growth phases can be observed. At earlier stages the increment in math achievement is greater than later in time. Thus, a linear trajectory does not seem plausible. In addition, it is clear that there is individual variation along the math achievement trajectory. Thus, modeling techniques need to consider these characteristics in the data.

Choosing a model. It seems plausible to assume a normal distribution in this example, but a linear trend seems unlikely. The data features segmented trajectories. A suitable option for these types of data is the piecewise mixed-effects model, which allows for different phases of growth or trajectories over time to be modeled. An important element of this type of models is the knot or change point, which characterizes the time point at which the trajectory changes from one phase to another (Kohli, Hughes, Wang, Zopluoglu, & Davison, 2015). Following Bates and Watts (1988), nonlinearity definition, this model is an example of nonlinear models. The flexibility of piecewise models allows different functional forms to describe each stage or phase of change. For the purposes or this illustration and based on the observed mean growth in Figure 4, we de-

cided to model a linear-linear trajectory with random effects for all model parameters. In summary, based in descriptive statistics and the nature of the outcome variable, we propose a nonlinear model assuming normality to analyze the mathematics IRT scores. This modeling combination results in a NLMEM, as characterized in Table 1. In addition, following the general GNLMEM modeling framework described in Model (1), the piecewise mixed-effects model can be specified as.

$$
\begin{aligned}
&\boldsymbol{y}_i | \boldsymbol{b}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) \\
&\boldsymbol{\mu}_i = \boldsymbol{\eta}_i \\
&\boldsymbol{\eta}_i = f(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i) \\
&\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\Phi})
\end{aligned}
\tag{4}
$$

where f is a nonlinear functional form described by

$$
f_{ij} =
\begin{cases}
\beta_{1i} + \beta_{2i}\, t_{ij} & t_{ij} \le \gamma_i \\
\beta_{1i} + \beta_{2i}\, \gamma_i + \beta_{3i}\, (t_{ij} - \gamma_i) & t_{ij} > \gamma_i
\end{cases}
\tag{5}
$$

for individual i at the jth time point (Kohli, Sullivan, et al., 2015). In equations (4) and (5), $\boldsymbol{y}_i$ is a 7×1 vector of math IRT scores for subject i $\boldsymbol{X}_i$, is a 7×4 design matrix, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \gamma)'$ is the vector of fixed effects, where β_1 represents the average intercept, β_2 the average slope before the change point, β_3 the average slope after the change point, and γ the average knot or change point. $\boldsymbol{b}_i$ is a 4×1 vector of random effects associated to the intercept, two slopes, and the knot. $\boldsymbol{Z}_i$ is a 7×4 design matrix associated to $\boldsymbol{b}_i$, $\boldsymbol{\eta}_i$ is the linear predictor, $g(\boldsymbol{\mu}_i|\boldsymbol{b}_i) = \boldsymbol{\mu}_i$ represents the identity link, and it is assumed $\boldsymbol{\Lambda}_i = \sigma^2 \boldsymbol{I}$.

Figure 4 ■ Individual math IRT score trajectories and observed mean of 35 randomly selected participants for the ECLS-K data.
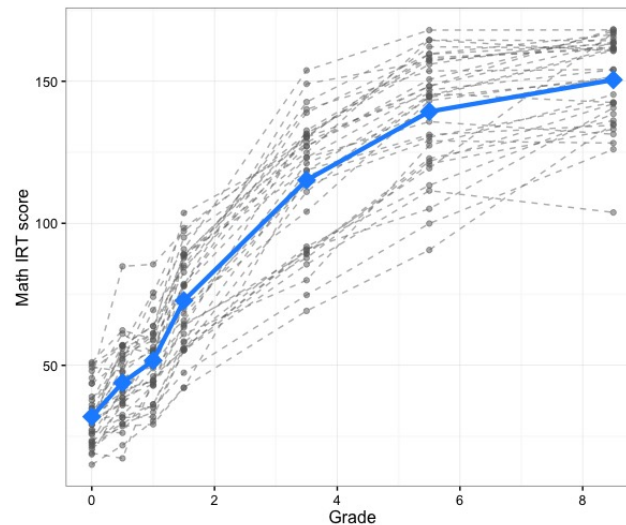

Results. The estimation results of Model (4) are presented in Table 6. Figure 5 illustrates the observed and estimated mean trajectories over time. The findings suggest that mathematics achievement evolves in different phases over time. The mean initial mathematics achievement was of 27.31 points. The initial slope was of 22.86 mathematics IRT scores per academic year, and the second slope of 5.72. That is, the findings showed a faster increase during the first stage of the trajectory followed by a slow down during the second phase of growth. Moreover, there was significant individual variability around the mean trajectory in all model parameters. Variance parameters associated to intercept, first and second slopes were 57.19, 37.91, and 5.11, suggesting that there were more individual differences from kindergarten to fourth grade. After the change point, individual differences around the second slope were smaller. The change point was located around fourth grade. This means that the mathematics developmental trajectory switched at that point. Likewise, individual differences were found around the mean change point as suggested by a variance of 0.81.

In summary, we have presented two illustrative examples on choosing a statistical model based on different combinations of data-type and data-behavior to analyze repeated measures data. In addition, model representations can be derived from the same general specification of mixed-effects models considering the most flexible framework, GNLMEMs.

Discussion

The aim of this article was to present a comprehensive statistical framework for describing different statistical models for a variety of repeated measures data. Distributional assumption and model linearity encompass the principal characteristics of modeling alternatives. We argue that characterizing models for repeated measures data by these two main features adds clarity and simplicity to the discussion of mixed-effects models. In addition, the class of GNLMEMs as defined in Model (1) represents a general scheme from which a variety of specific models can be derived (e.g., Table 1). The particular choice of the model depends on the characteristics of the data at hand. For example, if the response variable is categorical, then modeling the phenomenon of interest as a binomial or multinomial stochastic process would be appropriate. One example is when investigating memory performance by means of whether subjects recall items or not (e.g., Murayama et al., 2014). If we have a count outcome variable, assuming a Poisson model would be suitable. The analysis of abnormal behavior such as conduct disorder, depression or anxiety measured by the number of symptoms subjects experienced over a period of time (e.g., Lahey, Loeber, Burke, Rathouz, & McBurnett, 2002) requires the statistical model to consider the nature of the outcome variable, for which the Poisson distribution would be the most appropriate. Moreover, the behavior exhibited by the data also plays an important role in the modeling strategy choice. Sometimes linearity in parameters between the outcome variable and

Figure 5 ■ Observed and estimated mean trajectories for the piecewise mixed-effects model for the ECLS-K data.
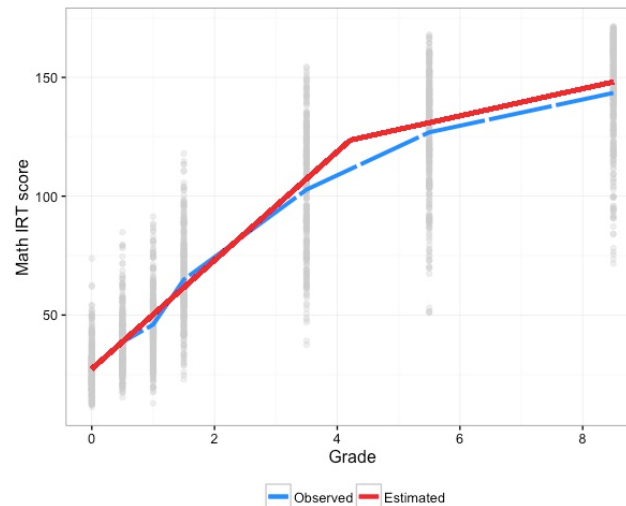
the covariate(s) will be sufficient to summarize the data, others may require a nonlinear model. For instance, learning experiments will often display a nonlinear trajectory to model reading speed or language development (e.g., Cudeck & Harring, 2007).
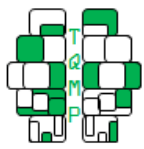
Broadly speaking, in the context of repeated measures modeling, researchers' main focus is to describe the behavior displayed by the data to the best of their knowledge. We advocate that considering modeling alternatives as a combination of two main components (distributional assumption and model linearity) will enable researchers to expand the scope of their modeling choices. We illustrated this by showing how different models for distinct data-type and data-behavior combinations can be derived from the GNLMEM to summarize change over time. However, it is still a researcher's task to get to know and explore the data at hand in order to choose the most appropriate model specification.

Despite the fact that the generalized framework for mixed-effects models has been discussed for over a couple of decades in the literature, still more dissemination is needed (Stroup, 2013). This is especially the case for models with distributional assumptions other than the normal and binomial. For instance, a wide lack of understanding of the Poisson model has been documented in applied psychological research (Vives et al., 2006). This reinforces the need of a broader exposure of generalized mixed-effects models, since both researchers and practitioners will greatly benefit from a wider insight of the range of options to model non-normal data. In addition, describing non-steady rate of change through nonlinear-

ity in parameters as defined by Bates and Watts (1988) has been discussed separately in the literature. But nonlinear change over time is found frequently in social and behavioral sciences. We therefore believe that researchers should be aware of the available modeling options above and beyond normal and binomial models, as well as all the possible combinations of the underlying stochastic process and the form of the trajectory over time.

Estimating a mixed-effects model is not an easy task. Issues regarding sample size, functional form, distributional assumption, or starting values, for example, may impact the estimation results or even convergence. Discussions of relevant estimation issues are given by Davidian and Giltinan (1995, 2003), Vonesh and Chinchilli (1997), L. Wu (2010), and Stroup (2013), for example. Additionally, the definition of the GNLMEM provided in this article considers the random effects to be normally distributed. In theory, other distributional assumptions could be considered and some efforts have been done to estimate mixed-effects models with non-normal random effects (e.g., Liu & Yu, 2008; Nelson et al., 2006). However, this option is still not readily available in statistical software.

Summarizing, separate efforts to model repeated measures data have resulted in distinct statistical modeling approaches whose differences lie mainly in distributional assumptions and model linearity. In addition, those approaches can be derived from the flexible GNLMEM umbrella framework. We argue that learning about the general GNLMEM scheme and viewing modeling alternatives as a combination of two main characteristics will provide researchers with a broader collection of statistical mod-

els to choose from. Because the generalized framework of mixed-effects models still remains underutilized, we hope this exposition will exhort researchers to make greater use of diverse modeling techniques when appropriate based on the nature of the data.

Authors' note

We would like to thank Yuhong V. Jiang of University of Minnesota for graciously providing the behavioral data from a spatial attention experiment used in one of our illustrative examples.

References

Anderson, C. J., Verkuilen, J., & Johnson, T. R. (2013). *Applied generalized linear mixed models?: Continuous and discrete data*. Washington: Book in preparation.

Arnon, I. (2010). Rethinking child difficulty: The effect of np type on children's processing of relative clauses in hebrew. *Journal of Child Language*, *37*(1), 27–57. Retrieved from http://doi.org/10.1017/S030500090900943X

Atkins, D. C., & Gallop, R. J. (2007). Rethinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models. *Journal of Family Psychology*, *21*(4), 726–735. Retrieved from http://doi.org/10.1037/0893-3200.21.4.726

Bates, D. M., & DebRoy, S. (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis*, *91*(1), 1–17. Retrieved from http://doi.org/10.1016/j.jmva.2004.04.013

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. Hoboken, New Jersey, USA: John Wiley & Sons.

Bolger, K. E., & Patterson, C. J. (2001). Developmental pathways from child maltreatment to peer rejection. *Child Development*, *72*(2), 549–568. Retrieved from http://doi.org/10.1111/1467-8624.00296

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*(421), 9–25.

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research*, *31*(3), 371–403. Retrieved from http://doi.org/10.1207/s15327906mbr3103_6

Cudeck, R., & Harring, J. R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology*, *58*, 615–637. Retrieved from http://doi.org/10.1146/annurev.psych.58.110405.085520

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. Boundary Row, London, UK: Chapman and Hall.

Davidian, M., & Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, *8*(4), 387–419. Retrieved from http://doi.org/10.1198/1085711032697

Delignette-Muller, M. L., & Dutang, C. (2015). Fitdistrplus: An r package for fitting distributions. *Journal of Statistical Software*, *64*(4), 1–34.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis (2nd ed.)* Hoboken, New Jersey, USA: John Wiley & Sons.

Hedeker, D. (2005). Generalized linear mixed models. In *Encyclopedia of statistics in behavioral science*. Washington: J. Wiley and sons. Retrieved from http://doi.org/10.1002/0470013192.bsa251

Huang, F. L., & Cornell, D. G. (2012). Pick your poisson: A tutorial on analyzing counts of student victimization data. *Journal of School Violence*, *11*(3), 187–206. Retrieved from http://doi.org/10.1080/15388220.2012.682010

Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. Retrieved from http://doi.org/10.1016/j.jml.2007.11.007

Jiang, Y. V., & Swallow, K. M. (2013). Body and head tilt reveals multiple frames of reference for spatial attention. *Journal of Vision*, *13*(13), 1–11. Retrieved from http://doi.org/10.1167/13.13.9

Kohli, N., Hughes, J., Wang, C., Zopluoglu, C., & Davison, M. L. (2015). Fitting a linear-linear piecewise growth mixture model with unknown knots: A comparison of two common approaches to inference. *Psychological Methods*, *20*(2), 259–275. Retrieved from http://doi.org/10.1037/met0000034

Kohli, N., Sullivan, A. L., Sadeh, S., & Zopluoglu, C. (2015). Longitudinal mathematics development of students with learning disabilities and students without disabilities: A comparison of linear, quadratic, and piecewise linear mixed effects models. *Journal of School Psychology*, *53*, 105–120. Retrieved from http://doi.org/10.1016/j.jsp.2014.12.002

Kreisman, M. B. (2003). Evaluating academic outcomes of head start: An application of general growth mixture modeling. *Early Childhood Research Quarterly*, *18*(2), 238–254. Retrieved from http://doi.org/10.1016/S0885-2006(03)00027-9

Lahey, B. B., Loeber, R., Burke, J., Rathouz, P. J., & McBurnett, K. (2002). Waxing and waning in concert: Dy-

namic comorbidity of conduct disorder with other disruptive and emotional problems over 7 years among clinic-referred boys. *Journal of Abnormal Psychology*, *111*(4), 556–567. Retrieved from http://doi.org/10.1037//0021-843X.111.4.556

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*(4), 963–974. Retrieved from http://doi.org/10.2307/2529876

Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics, 46*(3), 673–687. Retrieved from http://doi.org/10.2307/2532087

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *Sas for mixed models (second edi)*. Cary, NC, USA: SAS Institute Inc.

Liu, L., & Yu, Z. (2008). A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine*, *27*, 3105–3124. Retrieved from http://doi.org/10.1002/sim

Locascio, J. J., & Atri, A. (2011). An overview of longitudinal data analysis methods for neurological research. *Dementia and Geriatric Cognitive Disorders Extra*, *1*(1), 330–357. Retrieved from http://doi.org/10.1159/000330228

Long, J., & Ryoo, J. (2010). Using fractional polynomials to model non-linear trends in longitudinal data. *The British Journal of Mathematical and Statistical Psychology*, *63*(1), 177–203. Retrieved from http://doi.org/10.1348/000711009X431509

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models (second edi)*. Washington: Chapman and Hall.

Middleton, E. L., Chen, Q., & Verkuilen, J. (2015). Friends and foes in the lexicon: Homophone naming in aphasia. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *41*(1), 77–94. Retrieved from http://doi.org/10.1037/a0037778

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type i error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. journal of experimental psychology. *Learning, Memory, and Cognition*, *40*(5), 1287–1306. Retrieved from http://doi.org/10.1037/a0036914

Nelson, K. P., Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J., Parzen, M., & Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics*, *15*(1), 39–57. Retrieved from http://doi.org/10.1198/106186006X96854

Neuhaus, J., & McCulloch, C. (2011). Generalized linear models. *Computational Statistics*, *3*(5), 407–413.

Payne, R. W. (2014). Hierarchical generalized nonlinear models. In *In statistical modelling in biostatistics and bioinformatics* (pp. 244–244). Berlin: Spriger. Retrieved from http://doi.org/10.1007/978-3-319-04579-5

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, *2*(1), 1–21.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & de Toit, M. (2011). *Hlm 7: Linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, S. W., Yang, M.-L., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order , multivariate laplace approximation. *Journal of Computational and Graphical Statistics*, *9*(1), 141–157.

SAS Institute Inc. (2008). The NLMIXED procedure for SAS/STAT User's Guide (Version 9.2). Cary.

Sheiner, L. B., & Beal, S. L. (1980). Evaluation of methods for estimating population pharmacokinetics parameters. i. michaelis-menten model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*, *8*(6), 553–71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7229908

Shin, T., Davison, M. L., Long, J. D., Chan, C. K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences*, *23*(1), 92–100. Retrieved from http://doi.org/10.1016/j.lindif.2012.10.002

Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the simple view of reading. *Scientific Studies of Reading, (17*, 1–26. Retrieved from http://doi.org/10.1080/10888438.2011.618153

Stiratelli, R., Laird, N., & Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics*, *40*(4), 961–971. Retrieved from http://doi.org/10.2307/2531147

Stroup, W. W. (2013). *Generalized linear mixed models: Modern concepts, methods and applications*. Washington: Taylor and Francis Group.

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., & Najarian, M. (2009). *Early childhood longitudinal study , kindergarten class of 1998-99 (ecls-k): Combined user's manual for the ecls-k eighth-grade and k-8 full sample data files and electronic codebooks (nces)*. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Tuerlinckx, F., Rijmen, F., Verbeke, G., & De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *The British Journal of Mathematical and*
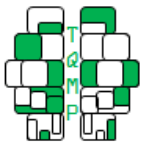
Statistical Psychology, *59*(2), 225–255. Retrieved from http://doi.org/10.1348/000711005X79857

Van Breukelen, G. J. P. (2005). Psychometric modeling of response speed and accuracy with mixed and conditional regression. *Psychometrika*, *70*(2), 359–376. Retrieved from http://doi.org/10.1007/s11336-003-1078-0

Vives, J., Losilla, J.-m., & Rodrigo, M.-f. (2006). Count data in psychological applied research. *Psychological Reports*, *98*(3), 821–835. Retrieved from http://doi.org/10.2466/PR0.98.3.821-835

Vonesh, E. F. (2012). *Generalized linear and nonlinear models for correlated data: Theory and applications using sas*. Cary, NC, USA: SAS Institute Inc.

Vonesh, E. F., & Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. New York, N.Y.: Marcel Dekker, Inc.

Vonesh, E. F., Chinchilli, V. M., & Pu, K. (1996). Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics*, *52*(2), 572–587. Retrieved from http://doi.org/10.2307/2532896

Wolfinger, R. D., & O'Connell, M. (1993). Generalized linear mixed-models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, *4*, 233–243.

Wu, H., & Zhang, J.-T. (2006). *Nonparametric regression methods for longitudinal data analysis: Mixed-effects modeling approaches*. Hoboken, New Jersey, USA: John Wiley & Sons.

Wu, L. (2010). *Mixed effects models for complex data*. Boca Raton, FL, USA: Chapman & Hall.

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, *44*(4), 1049–1060. Retrieved from http://doi.org/10.2307/2531734
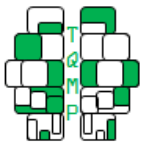
Appendix

SAS code for the gamma linear mixed-effects model

```
*Define data name and convergence options;
PROC NLMIXED DATA=data_long MAXITER=10000 NOAD QPOINTS=20 GCONV=0;
TITLE "Gamma model: Random intercept & slope";
*Individual generalized linear models with a gamma distribution were fitted
*Mean of individual parameter estimates were taken as initial values;
PARMS  Beta0=7.91  Beta1=-0.008 k=65.15 phi11=0.03 phi22=0.000016 phi21=0.00028;
*Define individual parameters;
B0i = Beta0 + b0;
B1i = Beta1 + b1;
*Define the linear predictor - block denotes the independent variable;
lin_pred = B0i + B1i*block;
*The natural logarithm is being used as the link function given the gamma parameterization
   of NLMIXED;
mu = exp(lin_pred);
*theta denotes the scale parameter;
*k denotes the shape parameter;
theta = mu/k;
*Define the gamma model;
MODEL MRT ~ gamma(k,theta);
*Define boundaries for parameters that should be non-negative;
BOUNDS k phi11 phi22 >= 0;
*Vector of random effects follows a normal distribution;
RANDOM b0 b1 ~ normal([0,0],[phi11,phi21,phi22]) subject=id;
RUN;
```

SAS code for the piecewise mixed-effects model

```
*Define data name and convergence options;
PROC NLMIXED DATA=data_long MAXITER=5000 NOAD QPOINTS=20 GCONV=0;
TITLE "Piecewise model: Random knot, intercept & slope";
* Initial values;
PARMS  Beta1=26.51  Beta2=25.72  Beta3=5.97 Gamma=3.56
       s2e=52.78 t11=9.8 t21=1.48 t22=7.26 t31=-0.92 t32=0.08 t33=3.56
```

```
        t41=0.08 t42=-0.55 t43=-0.61 t44=0.75;
*Define individual parameters;
B1i = Beta1 + b1;
B2i = Beta2 + b2;
B3i = Beta3 + b3;
*Gammai denotes the random knot;
Gammai = Gamma + b4;
*Define the piecewise mean trajectory;
mu = B1i + B2i*time;
IF (time > Gammai) THEN DO;
        mu = B1i + B2i*Gammai + B3i*(time - Gammai);
END;
*Outcome variable follows a normal distribution centered on mean trajectory with variance
    s2e;
MODEL math_score ~ normal(mu,s2e);
*Cholesky decomposition of the covariance matrix of random effects;
phi11 = t11*t11;
phi21 = t21*t11;
phi22 = t21*t21 + t22*t22;
phi31 = t31*t11;
phi32 = t31*t21 + t32*t22;
phi33 = t31*t31 + t32*t32 + t33*t33;
phi41 = t41*t11;
phi42 = t41*t21 + t42*t22;
phi43 = t41*t31 + t42*t32 + t43*t33;
phi44 = t41*t41 + t42*t42 + t43*t43 + t44*t44;
* Vector of random effects follows a normal distribution;
RANDOM b1 b2 b3 b4 ~ normal([0,0,0,0],[phi11,phi21,phi22, phi31, phi32, phi33, phi41, phi42,
    phi43, phi44]) subject=id;
*Recover parameters of the covariance matrix of random effects;
ESTIMATE 'phi11' t11*t11;
ESTIMATE 'phi21' t21*t11;
ESTIMATE 'phi22' t21*t21 + t22*t22;
ESTIMATE 'phi31' t31*t11;
ESTIMATE 'phi32' t31*t21 + t32*t22;
ESTIMATE 'phi33' t31*t31 + t32*t32 + t33*t33;
ESTIMATE 'phi41' t41*t11;
ESTIMATE 'phi42' t41*t21 + t42*t22;
ESTIMATE 'phi43' t41*t31 + t42*t32 + t43*t33;
ESTIMATE 'phi44' t41*t41 + t42*t42 + t43*t43 + t44*t44;
RUN;
```

R code for the gamma linear mixed-effects model using glmer function in the lme4 package

```
# Declare  initial  values as a  list
fixef <- c(7.91, -0.008)
theta <- c(0.03, 0.0028, 0.000016)
ini_val <- list("fixef" = fixef,
                "theta" = theta)
# Use the glmer function to  define  the  random−effects model
# following  a gamma distribution  with the  logarithmic  link  function
gamma_model <- glmer(MRT ~ 1 + block + (1 + block | id),
                    data=data_long,
                    family = Gamma(link = "log"),
                    start = ini_val)
```

R code for the piecewise mixed-effects model using JAGS

```r
# Define  function  to pass  to JAGS for  model estimation
# X is  the  design  matrix
# Y is  the  outcome variable  in  wide format
Bayes_PW_LL <- function(X, Y){
  # Bayesian model  definition
  pw_ll_rjags <- "model{
  for (i in 1:Nind) {
   for (j in 1:Ntime) {
     y[i,j] ~ dnorm(mu_y[i,j], prec_y)
     mu_y[i,j] <- b.ind[i, 1] + b.ind[i, 2] * (min(x[i,j], b.ind[i, 4])) + b.ind[i,
     3] * (max(0, x[i,j]- b.ind[i, 4]))
   }
  }
  ### Priors
  # Level 1 precision
  prec_y ~ dgamma(0.001, 0.001)
  # MVN priors for random-effects of the betas
  for (i in 1:Nind){
   b.ind[i, 1:4] ~ dmnorm(mu_b[1:4], tau_b[1:4,1:4])
  }
  # Prior for the fixed-effects
  # minT and maxT are the minimum and maximum time points
  # minCP and maxCP are the minimum and maximum values in time where the
  # change-point could occur. That is minCP is the second to the first time
  # point, and maxCP is the second to the last time point.
  mu_b[1] ~ dnorm(0, 0.001)
  mu_b[2] ~ dnorm(0, 0.001)
  mu_b[3] ~ dnorm(0, 0.001)
  mu_b[4] ~ dnorm(g,tau_g)T(minT,maxT)
  g ~ dunif(minCP, maxCP)            # change-point
  sigma_g ~ dunif(0, (maxT-minT)/4) # standard deviation of change-point
  tau_g <- 1/(sigma_g * sigma_g)
  # Prior for the var-cov of the random-effects
  # The variance has an inverse Wishart prior
  tau_b[1:4,1:4] ~ dwish(Omega[1:4,1:4], 5)
}
```
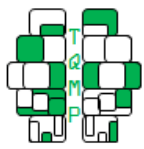
Citation

Tables 3 to 6 follows.

Table 3 ■ Descriptive statistics for mean response time ($N = 77$) for the spatial attention experiment

Block	Mean	S.D.	Min.	Max.	Range	Skew	Kurtosis
				Descriptive Statistic			
1	3013.63	627.62	1993.08	4699.17	2706.08	0.51	−0.46
2	2798.07	605.35	1707.17	4908.33	3201.17	0.72	0.68
3	2786.10	736.39	1708.33	5879.17	4170.83	1.36	2.73
4	2673.69	541.56	1700.83	4134.17	2433.33	0.52	−0.23
5	2644.23	568.04	1623.33	4104.17	2480.83	0.67	−0.18
6	2602.22	536.89	1652.67	3990.83	2338.17	0.49	−0.31
7	2636.05	558.15	1578.92	4230.00	2651.08	0.30	−0.34
8	2444.89	629.35	1453.50	4727.50	3274.00	1.26	1.66
9	2526.58	533.71	1529.92	3675.00	2145.08	0.31	−0.85
10	2537.34	620.60	1277.50	4456.67	3179.17	0.63	0.30
11	2443.67	456.17	1562.50	3618.33	2055.83	0.64	−0.06
12	2502.81	617.12	1437.75	4389.17	2951.42	0.98	0.55
13	2449.08	541.23	1582.50	4200.00	2617.50	0.85	0.44
14	2438.47	478.27	1626.67	3506.67	1880.00	0.31	−1.04
15	2477.25	556.14	1628.50	3968.33	2339.83	0.65	−0.14
16	2358.32	538.11	1303.00	4207.42	2904.42	0.76	0.63
17	2359.28	555.86	1571.75	4395.00	2823.25	1.17	1.58
18	2339.54	498.69	1454.42	4101.67	2647.25	0.99	1.29
19	2324.25	455.00	1425.00	3704.17	2279.17	0.40	−0.32
20	2288.61	459.69	1576.67	3650.00	2073.33	0.82	0.62
21	2365.29	485.12	1466.67	4455.00	2988.33	1.13	2.72
22	2257.43	394.71	1620.83	3163.33	1542.50	0.47	−0.74
23	2289.84	515.85	1498.33	4461.67	2963.33	1.60	3.53
24	2269.92	475.30	1304.42	3941.67	2637.25	0.65	0.79
25	2244.83	480.89	1391.08	3790.83	2399.75	0.76	0.19
26	2262.13	585.81	770.83	4624.17	3853.33	1.12	3.00
27	2194.36	466.42	979.50	3874.17	2894.67	0.72	1.62
28	2233.47	540.85	1158.17	3907.50	2749.33	1.06	1.26
29	2204.87	459.06	1478.33	3729.17	2250.83	0.79	0.59
30	2164.81	493.68	1128.50	3515.00	2386.50	0.66	0.18
31	2205.67	402.69	1490.00	3240.83	1750.83	0.42	−0.59
32	2189.86	406.30	1555.00	3281.67	1726.67	0.51	−0.32
Total	2422.70	565.37	770.83	5879.17	5108.33	0.93	1.34

Note. Kurtosis reported in this table is sample kurtosis.

Table 4 ■ Estimation results of the gamma mixed-effects models for the spatial attention experiment

Parameter	Estimate		S.E.
Model 1			
β_0	7.873	***	0.009
β_1	−0.001	***	0.000
ϕ	0.032		
k	47.486	***	1.360
Model 2			
β_0	7.916	***	0.015
β_1	−0.008	***	0.000
ϕ_{11}	0.026		
ϕ_{22}	0.000		
ϕ_{21}	0.000		
	49.456	***	1.448
BIC			
Model 1	36067		
Model 2	36041		

Note. ***: $p < 0.001$. Mean slopes are reported up to three decimals to better represent parameter estimates. S.E. = standard error.

Table 5 ■ Descriptive statistics for mathematics IRT scores for ECLS-K data

	Descriptive Statistic						
Grade	Mean	S.D.	Min.	Max.	Range	Skew	Kurtosis
0	27.56	9.45	11.36	73.84	62.48	1.06	1.76
0.5	38.63	11.93	15.43	84.85	69.42	0.89	1.10
1	45.97	14.11	12.85	91.42	78.57	0.63	0.48
1.5	64.89	17.33	22.97	117.97	95.00	0.39	0.11
3.5	102.79	23.97	37.47	154.49	117.02	−0.16	−0.55
5.5	126.90	23.50	50.87	168.01	117.14	−0.71	0.15
8.5	143.47	21.73	71.65	171.39	99.74	−1.01	0.53

Table 6 ■ Estimation results of the piecewise mixed-effects model for ECLS-K data

Parameter	Estimate		S.E.
β_1	27.31	***	0.48
β_2	22.86	***	0.39
β_3	5.72	***	0.21
γ	4.212	***	0.07
φ_{11}	57.19	***	5.83
φ_{22}	37.91	***	4.04
φ_{33}	5.11	***	1.19
φ_{44}	0.81	***	0.12
φ_{21}	34.22	***	3.49
φ_{31}	−3.46	*	1.59
φ_{32}	0.13		1.38
φ_{41}	−4.32	***	0.52
φ_{42}	−4.50	***	0.58
φ_{43}	−0.87	**	0.31
σ^2	53.13	***	2.09
BIC	21198		

Note. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. S.E. = standard error.