International Journal of Psychological Research
ISSN: 2011-2084
ijpr@usbmed.edu.co
Universidad de San Buenaventura
Colombia

Hsieh, Chueh-An; von Eye, Alexander
The Best of Both Worlds: A Joint Modeling Approach for the Assessment of Change across Repeated Measurements.
International Journal of Psychological Research, vol. 3, núm. 1, 2010, pp. 177-210
Universidad de San Buenaventura
Medellín, Colombia

International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Hsieh, C., Von Eye, A., (2010). The Best of Both Worlds: A Joint Modeling Approach for the Assessment of Change across Repeated Measurements. *International Journal of Psychological Research, 3* (1), 177-210.

The Best of Both Worlds: A Joint Modeling Approach for the Assessment of Change across Repeated Measurements.

Lo mejor de ambos mundos: una propuesta de modelamiento combinado para la evaluación del cambio a lo largo de mediciones repetidas.

Chueh-An Hsieh and Alexander von Eye
Michigan State University, East Lansing, USA

ABSTRACT

The usefulness of Bayesian methods in estimating complex statistical models is undeniable. From a Bayesian standpoint, this paper aims to demonstrate the capacity of Bayesian methods and propose a comprehensive model combining both a measurement model (e.g., an item response model, IRM) and a structural model (e.g., a latent variable model, LVM). That is, through the incorporation of the probit link and Bayesian estimation, the item response model can be introduced naturally into a latent variable model. The utility of this IRM-LVM comprehensive framework is investigated with a real data example and promising results are obtained, in which the data drawn from part of the British Social Attitudes Panel Survey 1983-1986 reveal the attitude toward abortion of a representative sample of adults aged 18 or older living in Great Britain. The application of IRMs to responses gathered from repeated assessments allows us to take the characteristics of both item responses and measurement error into consideration in the analysis of individual developmental trajectories, and helps resolve some difficult modeling issues commonly encountered in developmental research, such as small sample sizes, multiple discretely scaled items, many repeated assessments, and attrition over time

Key words: Bayesian inference, item response model, latent growth curve analysis, simulation, generalized linear latent and mixed model

RESUMEN

La utilidad de los métodos Bayesianos en la estimación de modelos estadísticos complejos es innegable. Desde un punto de vista Bayesiano, el presente artículo busca demostrar la capacidad de los métodos Bayesianos y proponer un modelo exhaustivo que combina un modelo de medición y un modelo estructural. La utilidad de este método combinado se investiga usando datos reales tomados de una encuesta sobre actitudes sociales. El método combinado permite extraer las características de las respuestas a los ítems como de los errores en la medición para el análisis individual de trayectorias del desarrollo. Tales resultados permiten resolver asuntos que se presentan en investigación en psicología del desarrollo, e.g., tamaños de muestra pequeños, evaluaciones repetidas, etc.

Palabras clave: Inferencia Bayesiana, modelo de respuesta al ítem, análisis de curva de crecimiento latente, simulación, modelo linear generalizado latente y mixto.

Artículo recibido/Article received: December 15, 2009/Diciembre 15, 2009, Artículo aceptado/Article accepted: March 15, 2009/Marzo 15/2009

Dirección correspondencia/Mail Address:
Chueh-An Hsieh and Alexander von Eye, Michigan State University, East Lansing, USA Email: hsiehchu@msu.edu

1. LONGITUDINAL DATA ANALYSIS

1.1 Covariance structure analysis (CSA)

The use of growth models in social, behavioral, and educational research has increased rapidly, for it answers important research questions such as the study of psychological, social development and the process of learning. Already it is well known that growth models can be approached from several perspectives via the formulation of equivalent models and can provide identical estimates for a given data set (e.g., Bauer, 2003; Chou, Bentler, and Pentz, 1998; Curran, 2003; Engel, Gattig, and Simonson, 2007; Hox and Stoel, 2005; Willett and Sayer, 1994). For instance, a model can be constructed as a standard two-level hierarchical linear model (HLM), where the repeated measures are positioned at the lowest level, and treated as nested within the individuals (e.g., Singer, 1998; Steele and Goldstein, 2007). Equally, a model can be constructed as a structural equation model (SEM), in which latent variables are used to account for the relations between the observed variables; hence, it has the name latent growth curve (LGC) analysis (e.g., Muthén, 2002).

It is this mean and covariance structure that makes it possible to specify exactly the same model as an HLM or LGC, because the fixed and random effects in the HLM correspond to the mean and covariance structure of the latent variables in the LGC analysis. However, although several key differences remain between these two models, the discrepancies are rapidly disappearing (Preacher et al., 2008; Raykov, 2007). One of the major differences is that, in the HLM, time is treated as a fixed explanatory variable, whereas it is introduced in the LGC model via the factor loadings. The consequence is that the HLM is essentially a univariate approach with time points treated as observations of the same variable, whereas the LGC model essentially takes a multivariate approach with each time point treated as a separate variable (e.g., Bauer, 2003; Curran, 2003; Hox and Stoel, 2005; Muthén, 2002; Preacher, Wichman, MacCallum, and Briggs, 2008; Raudenbush and Bryk, 2002; Willett and Sayer, 1994).

When the outcome measurements are on a discrete scale, however, the application of conventional growth curve models will introduce a potentially significant bias in the analysis and subsequent inferences (Curran, Edwards, Wirth, Hussong, and Chassin, 2007). Currently, there are two major modeling strategies which allow for the explicit incorporation of categorical repeated data in growth curve models. One strategy is to use the nonlinear multilevel model (e.g., Diggle, Heagerty, Liang, and Zeger, 2002; Gibbons and Hedeker, 1997; Raudenbush, Johnson, and Sampson, 2003), and the other is to use the nonlinear structural equation model (e.g., Jöreskog, 2002; Muthén, 1983, 1984, 1996, 2002). As Curran et al. (2007) and Vermunt (2007) state, when fitting measurement models to empirical data of the type commonly encountered in developmental research, such as small sample sizes, multiple discretely scaled items, many repeated assessments, and attrition over time, both models become quite complex and have difficulty achieving convergence[1].

1.2 A unified modeling approach

In order to accommodate this, we bridge the gap by resorting to an integrative modeling framework: using the derivative of the generalized linear latent and mixed model (GLLAMM; Skrondal and Rabe-Hesketh, 2004) in a single analytic session. That is, strengthened by attributes of the item response model (e.g., Lord and Novick, 1968) and latent variable model (e.g., Muthén, 2002), we can incorporate multiple categorical measurement models in the growth curve analysis. For instance, in item response models, the generalized linear model (GLIM) formulation is typically used. Through a commonly used link function, either a logit or probit, the conditional probability of a particular response given the latent trait can easily be specified. The classical application of these models is in the literature on educational testing and psychometrics, where the subscript *i* represents an item or question in a test and the responses are scored as correct (1) or incorrect (0) for dichotomous items. In this setting, η_j represents the latent ability of person *j*, and the model is parameterized as $\text{logit}[Pr(Y_{ij}=1|\eta_j)]=\alpha_i(\eta_j - \beta_i)$ or $\text{probit}[Pr(Y_{ij}=1|\eta_j)]=\alpha_i(\eta_j - \beta_i)$, corresponding to a unidimensional two-parameter logistic item response model or a unidimensional two-parameter normal ogive model. Here, the abilities can be interpreted either as logits or probits of the probability of a correct response to a particular item[2]. Item difficulty parameters (β_i) are defined as the location of inflection points in an item characteristic curve (ICC) along the same scale as the latent ability (η); thus, when the participant's ability equals the item difficulty, he or she will have a 50% chance of correctly answering an item or endorsing an item, whereas the α_i are the slopes of ICCs at their inflection points, which can be considered the degree to which item response varies with the underlying latent

---

[1] Specifically, with categorical response variables, when there are more than two or three latent variables with random effects, relying on the untestable assumption that these random coefficients come from a multivariate normal distribution, the integrals appearing in the likelihood function are hard to analytically determine and need to be solved using approximation methods (Vermunt, 2007).

[2] For instance, if subjects have the abilities of ±1, the implication is that in the logit scale there are 73% and 27% probabilities of correctly answering an item, say, the item 1 (since $\beta_1=0$ and $\alpha_1=1$). Similarly, in the probit scale the probabilities of correctly answering the item 1 for the same person are 84% and 16%, respectively.

construct: this helps determine how well the item discriminates between subjects with different abilities (e.g., Birnbaum, 1968; Lord and Novick, 1968).

As a useful version of the random coefficient model, a single-domain latent growth curve analysis was adopted in the present study, in which individuals were assumed to differ not only in their intercepts, but also in other aspects of their trajectory over time in terms of a unidimensional latent variable, the general attitude toward abortion (e.g., Byrne and Crombie, 2003; Skrondal and Rabe-Hesketh, 2008). As the chronological ordering of responses and the clustering of responses within individuals are two important features of longitudinal data, in order to accommodate this mean and covariance structure, a longitudinal model must allow for dependence among responses on the same subject (e.g., Everitt, 2005; Skrondal and Rabe-Hesketh, 2004). As a two-factor model, the latent growth curve model can be formulated as, $\eta_{kj}=\lambda_{ok}\zeta_{oj}+\lambda_{1k}\zeta_{1j}+\varepsilon_{kj}$ where η_{kj} the propensity to hold a positive attitude toward abortion at the k^{th} occasion for an individual j (measured by four dichotomous items at each point of time), is the focus of the study; ζ_{0j} and ζ_{1j} are the true initial level and shape factors; and ε_{kj} represents the level-1 residuals in the structural model. The data that we analyzed are time-structured and balanced in occasions: all subjects are measured at the same set of time points, k=1, …, 4. In addition, the loadings for the intercept factor ζ_{0j} are fixed at $\lambda_{0k}=1$ ($\forall\, k$), and the loadings for the shape factor ζ_{1j} are set equal to λ_{1k}. As Meredith and Tisak (1990) suggest, it is feasible to model a nonlinear growth or decline trajectory using a two-factor model with free factor loadings for ζ_{1j}. To make the model identifiable, we set $\lambda_{11}=0$ and $\lambda_{14}=1$ and estimate the coefficients for the intermediate time points[3].

As for the link function, given the similarities between logit and probit of these two models (i.e., both models can be treated as sample cases of the wider class of generalized linear models, GLIMs), either model will in most applications give identical substantive conclusions (Liao, 1994; Stefanescu, Berger, and Hershberger, 2005). Normally, by multiplying a factor of $\pi/\sqrt{3}$, we can go from one set of estimates to the other[4]. However, when we have heavy tails in the distribution of observations, estimates from the logit and probit models can differ substantially (Amemiya, 1981). Moreover, one of the Markov chain Monte Carlo (MCMC) sampling algorithms, the direct Gibbs sampling (Albert, 1992; Chib and Greenberg, 1995; Gelfand et al., 1990; Patz and Junker, 1999a) has been implemented only for normal ogive item response models, requiring the use of a process called data augmentation (Albert and Chib, 1997; Fox, 2007; Jackman, 2000; Kim and Bolt, 2007; Stefanescu et al., 2005). Therefore, in the model estimated via Bayesian Markov chain Monte Carlo, using Gibbs sampling as implemented in WinBUGS 1.4 (Spiegelhalter, Thomas, Best, and Lunn, 2003), the probit link[5] is considered the more appropriate function.

With the longitudinal design, mathematically, the response model can now be written as:

$\text{probit}[Pr(Y_{kij}=1|\eta_{kj})]=\alpha_{ki}(\eta_{kj}-\beta_{ki})$, where subscript k represents the different occasions. In the present study, assuming that the assumption of strong measurement invariance holds (Meredith and Teresi, 2006; Sayer and Cumsille, 2001), we then impose equality for each of the item parameters over time[6] (i.e., assuming that neither item difficulty nor item discrimination varies across different points in time), which further reduces α_{ki} to α_i and β_{ki} to β_i from the above mathematical formula. If the invariance of the factor structure fails to hold over time, the consequence is that the difference in means may be partially attributable to differences in the scale of a factor's (Blozis, 2007). Thus, through the estimated item characteristic curves (ICCs) for a unidimensional two-parameter item response model, this unified modeling approach can be specified as:

$Pr(Y_{kij}=1|\eta_{kj})=\exp(\nu_{kij})/(1+\exp(\nu_{kij}))=\Phi(\nu_{kij})$, where ν_{kij} is the linear predictor (i.e., $\alpha_i(\eta_{kj}-\beta_i)$), $\Phi(\cdot)$ is the standard normal cumulative distribution function (i.e., the inverse probit link), and again, η_{kj} can be replaced by $\tau_k+\lambda_{0k}\zeta_{0j}+\lambda_{1k}\zeta_{1j}+\varepsilon_{kj}$. As the model becomes more complex, for identification purposes we exclude the intercept (τ_k) from the structural models, fix one of the constants in the response model, and set the first item difficulty parameter (β_1) equal to zero. By doing so, we enforce other individual-level covariates (e.g., gender, age, and religious status) to affect the response via the latent variable only (Skrondal and Rabe-Hesketh, 2004).

1.3 Objectives of the study

The usefulness of Bayesian methods in estimating complex statistical models is undeniable. Bayesian data analysis is seen as having a range of advantages, such as an intuitive probabilistic interpretation of the parameters of

---

[3] This level and shape model (LS) is equally useful regardless of whether the developmental trajectory is linear or nonlinear (Raykov and Marcoulides, 2006)

[4] Or, multiplying a factor lying somewhere between 1.6 and 1.8 (Amemiya, 1981).

[5] In addition, a useful feature of the probit model is that it can be used to yield tetrachoric correlations for the clustered binary outcomes, and polychoric correlations for ordinal responses (Hedeker, 2005).

[6] However, for most applications in which the aim is to ensure fairness and equity, a stronger assumption of strict factorial invariance is necessary, that is, equal factor loadings, intercepts, and equivalent residual variances (specific factor plus error variable) across different occasions (Meredith and Teresi, 2006).

interest, the efficient incorporation of prior information into empirical data analysis, model averaging, and model selection (Best, Spiegelhalter, Thomas, and Brayne, 1996; Maier, 2001; Rupp, Dey, and Zumbo, 2004; Western, 1999). Additionally, unlike the maximum likelihood estimation requiring large samples to approximate the sampling distribution of the MLE when making statistical inference, Bayesian inference can be considered a plausible way to deal with small sample size studies (Congdon, 2005; Lee and Wagenmakers, 2005; Zhang et al. 2007). Thus, the present study aim is to demonstrate the capacity of Bayesian methods, and propose a comprehensive model combining both a measurement model (e.g., an item response model, IRM) and a structural model (e.g., a latent variable model, LVM). That is, through the incorporation of the probit link and Bayesian estimation, the item response model can be naturally introduced into a latent variable modeling framework.

Despite a large number of components requiring attention when selecting an appropriate statistical model, this study restricts its focus to the following issues: (1) model formulation: how Bayesians explicitly incorporate multiple repeated measures of discrete scale into a growth curve model. In order to differentially weight individual items, and examine developmental stability and change over time, one specific model, an IRM-LGC is presented. This model combines the two-parameter normal ogive item response model (e.g., Lord and Novick, 1968) and latent growth curve analysis (e.g., Meredith and Tisak, 1990); (2) missing data compensation: as the results from the full dataset[7] (the one containing missing outcomes) do not differ systematically from the complete cases in unconditional models, the unprovable missing data generation mechanism, missing at random (MAR; Rubin, 1987), seems sustainable[8]. Being an alternative estimation method, the Bayesian inference explicitly models missing outcomes and handles them as extra parameters to estimate (Gelman and Hill, 2007; May, 2006; Patz and Junker, 1999b; Spiegelhalter et al., 2003). Thus, the incorporation of individual-level auxiliary predictors makes it trivial to use the Bayesian approach to effectively estimate missing values in a conditional model (Carrigan, Barnett, Dobson and Mishra, 2007; Gelman and Hill, 2007).

2. THE DERIVATIVE OF THE GENERALIZED LINEAR LATENT AND MIXED MODEL (GLLAMM)

As a class of multilevel latent variable models, the GLLAMM encompasses the response model and the structural model (Skrondal and Rabe-Hesketh, 2003; 2004). In the present study, a unified modeling approach, IRM-LGC, is proposed, where the former pertains to the response model (i.e., the generalized linear model, GLIM) and the latter relates to the structural model. When we incorporate random effects in the underlying continuous latent construct(s) (i.e., augment GLIMs via the inclusion of random effects in the latent variables; hence the name 'generalized linear mixed models'-GLMMs), and regress latent variables upon other latent variables or covariates, this unified model is expanded as the generalized linear latent and mixed model (GLLAMM). Thus, with longitudinal designs, the data are of a multilevel nature with a set of categorically scored items nested within each person on each measurement occasion, which causes the response model, the structural model, and three-level indices (k=1, …T; i=1, …I; j=1,…,n) to be the required elements.

2.1 The response model

Standard use of a latent growth curve analysis typically considers a single manifest indicator at each measurement occasion, in which each response is a function of time and constitutes the first level of the measurement model. However, taking such an approach fails to capitalize on one of the capacities inherent in the structural equation models (SEM): not only ignoring the relations between multiple indicators and the underlying latent construct, but also dismissing information about the psychometric properties of manifest variables (Sayer and Cumsille, 2001). On the contrary, when we incorporate multiple indicators of discretely scaled measurement into the model, a second-order factor structure is utilized to investigate the developmental trajectory over time, which allows the researcher to evaluate the factorial invariance of the latent constructs across waves, and permits the separation of time-specific error and measurement error (Blozis, 2007; Sayer and Cumsille, 2001).

As mentioned earlier, through the estimated ICCs the multivariate random-coefficient probit regression model for dichotomous responses can be expressed as

$Pr(Y_{kij}=1|\eta_{kj})=\exp(v_{kij})/(1+\exp(v_{kij}))=\Phi(v_{kij})$, where v_{kij} is the linear predictor; $\Phi(\cdot)$ is the standard normal cumulative distribution function, and η_{kj} are the latent ability for person j at occasion k^{th}. Using matrix formulation, the vector of linear predictors can now be written as follows (Zheng and Rabe-Hesketh, 2007):

$$v_{kij} = \eta_{kj} Z \alpha_i - X(\alpha\beta)_i \qquad (2.1)$$

(k=1,…, T; i=1,…, I; j=1,…, n), where Z and X are the design matrices, α_i and $(\alpha\beta)_i$ are the corresponding vectors

---

[7] The overall nonresponse rate is 12.1%, (323-284)/323.

[8] The corresponding significance value associated with Little's MCAR test (Little, 1988) is .222, indicating that the data are missing completely at random (MCAR).

associated with item parameters, and again, η_{kj} represents the latent ability across the entire study span. Since a multivariate random coefficient probit regression for dichotomous responses is utilized in the present study, where v_{kij} represents the vector associated with the probit function of a correct response for item *i* and person *j* at the k^{th} occasion, and $\eta_{kj}Z\alpha_i$ and $X(\alpha\beta)_i$ are the respective random and fixed components, analogously, using Equation (2.1), the unidimensional two-parameter normal ogive response model for four dichotomous items can be written as follows:

$$\begin{bmatrix} v_{k1j} \\ v_{k2j} \\ v_{k3j} \\ v_{k4j} \end{bmatrix} = \left(\eta_{kj}\right) I_{4\times 4} \begin{bmatrix} 1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \alpha_2\beta_2 \\ \alpha_3\beta_3 \\ \alpha_4\beta_4 \end{bmatrix} \qquad (2.2)$$

(k=1,…, T; i=1,…, 4; j=1,…, n). For identification purposes, we fix the first item discrimination parameter (α_1) to 1 and the first item difficulty parameter (β_1) to zero, which results in $\alpha_1\beta_1=0$. By doing so, we enforce level-2 covariates at the structural model to affect the response through the latent variable (η_{kj}) only (Skrondal and Rabe-Hesketh, 2004).

2.2 The structural model: the two-stage formulation

Perhaps the most intuitively appealing way of specifying a growth curve model is to link it to two distinct questions about change: one entails the starting position (level) and the other involves the overall true change across the entire study span (shape), each arising from a specific level in a natural hierarchy; this is called two-stage model formulation (Rabe-Hesketh and Skrondal, 2008; Singer and Willett, 2005).

2.2.1 The level-1 structural model

When taking the perspective of latent response formulation, change is now modelled in the repeated latent constructs and no longer in the observed scores, making it possible for the error in the measurement model to be partitioned into time-specific error and measurement error. As Blozis (2007) puts it, serving as the subject of analysis, the latent variable encompasses time-specific error without the confounding influence of measurement error. This is because at each point of time, a common factor is assumed to account for the dependencies among a set of categorically scored items and allow for the decomposition of the error variances not attributable to growth (i.e., time-specific error variances). Using LISREL notation, the level-1 structural model can be expressed as follows:

$$\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \\ \eta_{3j} \\ \eta_{4j} \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{bmatrix} + \begin{bmatrix} 1 & \lambda_{11} \\ 1 & \lambda_{12} \\ 1 & \lambda_{13} \\ 1 & \lambda_{14} \end{bmatrix} \begin{bmatrix} \zeta_{0j} \\ \zeta_{1j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \varepsilon_{3j} \\ \varepsilon_{4j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & \lambda_{12} \\ 1 & \lambda_{13} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \zeta_{0j} \\ \zeta_{1j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \varepsilon_{3j} \\ \varepsilon_{4j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & \lambda_{12} \\ 1 & \lambda_{13} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \zeta_{0j} \\ \zeta_{1j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \varepsilon_{3j} \\ \varepsilon_{4j} \end{bmatrix}$$

or

$$\eta_{kj}=\lambda_{0k}\zeta_{0j}+\lambda_{1k}\zeta_{1j}+\varepsilon_{kj}=\zeta_{0j}+\lambda_{1k}\zeta_{1j}+\varepsilon_{kj} \qquad (2.3)$$

(k=1, …, T; j=1,…, n; λ_{11}=0, λ_{14}=1). Since the repeated measures (η_{kj}) have been extracted from the item response model through the probit link, converting the expected response to the linear predictor, Equation (2.3) is the structural model. As before, the term η_{kj} refers to the measures of an individual *j* at time *k,* and is a function of intercepts (τ_k), latent variables (representing the underlying initial status (ζ_{0j}) and the relative growth or decline trajectory (ζ_{1j})), and time-specific disturbance residuals (ε_{kj}). The τ_k term is typically constrained to zero, yielding a simplified model structure. Also, if there is a significant amount of variation to be explained, analysis can proceed in a stepwise manner by adding time varying covariates (TVCs) to Equation (2.3), as time-specific predictors of the repeated measures. In order to model a nonlinear growth or decline trajectory, we adopt the suggestion of Meredith and Tisak (1990): fixing all λ_{0k} equal to 1 and setting λ_{11} and λ_{14} to be 0 and 1, for model identification purposes. Unlike the assumption typically made in SEM (ε_{kj} are identically and independently normally distributed with mean (0) and variance (ψ)), we allow disturbance residuals at the level-1 structural model to be time-heteroskedastic and to vary across different occasions. That is, these time-specific error variances are distributed heteroscedastically over time within-person. Because the random-effect (η_{kj}) can be further represented by the variances of ζ_{0j} and ζ_{1j} at the second level of structural model, the LGC represents one kind of the random-effect models, and

$$\Theta_{\varepsilon} = \begin{bmatrix} \sigma^2_{\varepsilon 1} & 0 & 0 & 0 \\ 0 & \sigma^2_{\varepsilon 2} & 0 & 0 \\ 0 & 0 & \sigma^2_{\varepsilon 3} & 0 \\ 0 & 0 & 0 & \sigma^2_{\varepsilon 4} \end{bmatrix} (2.4)$$

2.2.2 The level-2 structural model

The level-2 structural model allows us to distinguish the growth or decline trajectories among individuals using their specific growth parameters, such as the true initial status and the change rate, implying that we can examine unobserved heterogeneity in growth curves by studying inter-individual variation in growth parameters. As declared by Singer and Willett (2005), an appropriate level-2 model entails the four following characteristics: (1) the level-2 outcomes are the level-1 individual growth parameters; (2) the level-2 model can be written in separate formulae, one for each level-1 growth parameter; (3) each formula specifies a relationship between the individual growth parameter and time-invariant covariates (TICs), and (4) each level-2 formula must contain the stochastic component: for those individuals who share a common predictor could vary in their specific change trajectories; hence the name random coefficient models. Similarly, the level-2 LGC model can be expressed as

$$\begin{bmatrix}\zeta_{0j}\\ \zeta_{1j}\end{bmatrix}=\begin{bmatrix}v_{00}\\ v_{10}\end{bmatrix}+\begin{bmatrix}\gamma_{01}\\ \gamma_{11}\end{bmatrix}\begin{bmatrix}\omega_{1j}\end{bmatrix}+\begin{bmatrix}\gamma_{02}\\ \gamma_{12}\end{bmatrix}\begin{bmatrix}\omega_{2j}\end{bmatrix}+\begin{bmatrix}\gamma_{03}\\ \gamma_{13}\end{bmatrix}\begin{bmatrix}\omega_{3j}\end{bmatrix}+\ldots+\begin{bmatrix}\upsilon_{0j}\\ \upsilon_{1j}\end{bmatrix} \text{ or}$$

$$\zeta_{0j}=v_{00}+\gamma_{01}\omega_1+\gamma_{02}\omega_2+\gamma_{03}\omega_3+\gamma_{04}\omega_1*\omega_2+\gamma_{05}\omega_1*\omega_3+\gamma_{06}\omega_2*\omega_3+\gamma_{07}\omega_1*\omega_2*\omega_3+\upsilon_{0j} \tag{2.5}$$

$$\zeta_{1j}=v_{10}+\gamma_{11}\omega_1+\gamma_{12}\omega_2+\gamma_{13}\omega_3+\gamma_{14}\omega_1*\omega_2+\gamma_{15}\omega_1*\omega_3+\gamma_{16}\omega_2*\omega_3+\gamma_{17}\omega_1*\omega_2*\omega_3+\upsilon_{1j} \tag{2.6}$$

(j=1,…, n), where Equations (2.5) and (2.6) represent regression equations among latent variables, one for each level-1 growth parameter. Proposed in the present study (as model 1 in both data sets), in a fully conditional model, the ζ_{0j} and ζ_{1j} factors with v_{00} and v_{10} as corresponding intercepts, are predicted by gender (ω_1), age (ω_2), religious status (ω_3), and all two-way and three-way interactions with residuals υ_{0j} and υ_{1j}, respectively. In addition, υ_{0j} and υ_{1j} are usually assumed to have a bivariate normal distribution with zero mean and unstructured covariance matrix. That is, controlling for the predictors of change, the residual variances and covariance of true initial level and shape are distributed as follows.

$$\psi=COV(\upsilon)=\begin{bmatrix}\sigma_{\upsilon 0}^2 & \sigma_{\upsilon 01}\\ \sigma_{\upsilon 10} & \sigma_{\upsilon 1}^2\end{bmatrix} \tag{2.7}$$

2.2.3 The composite model

As an analogous model formulation, a reduced form (Equation (2.8)) can be specified by substituting Equations (2.5) and (2.6) into the level-1 model (2.3), postulating that the general attitude simultaneously depends on: (1) the corresponding factor loadings for each level-1 predictor, the latent growth parameters; (2) the level-2 predictors, such as gender, age, and religious status; (3) the cross-level interaction (i.e., the corresponding factor loadings of each level-1 predictor multiplied by the level-2 predictors), and (4) the composite residual variances.

$$\begin{aligned}\eta_{kj}&=\zeta_{0j}+\lambda_{1k}\zeta_{1j}+\varepsilon_{kj}\\
&=(v_{00}+\gamma_{01}\omega_1+\gamma_{02}\omega_2+\gamma_{03}\omega_3+\gamma_{04}\omega_1*\omega_2+\gamma_{05}\omega_1*\omega_3+\gamma_{06}\omega_2*\omega_3+\gamma_{07}\omega_1*\omega_2*\omega_3+\upsilon_{0j})\\
&\quad+\lambda_{1k}(v_{10}+\gamma_{11}\omega_1+\gamma_{12}\omega_2+\gamma_{13}\omega_3+\gamma_{14}\omega_1*\omega_2+\gamma_{15}\omega_1*\omega_3+\gamma_{16}\omega_2*\omega_3+\gamma_{17}\omega_1*\omega_2*\omega_3+\upsilon_{1j})+\varepsilon_{kj}\\
&=(v_{00}+\gamma_{01}\omega_1+\gamma_{02}\omega_2+\gamma_{03}\omega_3+\gamma_{04}\omega_1*\omega_2+\gamma_{05}\omega_1*\omega_3+\gamma_{06}\omega_2*\omega_3+\gamma_{07}\omega_1*\omega_2*\omega_3+\upsilon_{0j})\\
&\quad+(\lambda_{1k}v_{10}+\lambda_{1k}\gamma_{11}\omega_1+\lambda_{1k}\gamma_{12}\omega_2+\lambda_{1k}\gamma_{13}\omega_3+\lambda_{1k}\gamma_{14}\omega_1*\omega_2+\lambda_{1k}\gamma_{15}\omega_1*\omega_3+\lambda_{1k}\gamma_{16}\omega_2*\omega_3\\
&\quad+\lambda_{1k}\gamma_{17}\omega_1*\omega_2*\omega_3+\lambda_{1k}\upsilon_{1j})+\varepsilon_{kj}\\
&=(v_{00}+\lambda_{1k}v_{10})+(\gamma_{01}\omega_1+\gamma_{02}\omega_2+\gamma_{03}\omega_3+\gamma_{04}\omega_1*\omega_2+\gamma_{05}\omega_1*\omega_3+\gamma_{06}\omega_2*\omega_3+\gamma_{07}\omega_1*\omega_2*\omega_3)\\
&\quad+(\lambda_{1k}\gamma_{11}\omega_1+\lambda_{1k}\gamma_{12}\omega_2+\lambda_{1k}\gamma_{13}\omega_3+\lambda_{1k}\gamma_{14}\omega_1*\omega_2+\lambda_{1k}\gamma_{15}\omega_1*\omega_3+\lambda_{1k}\gamma_{16}\omega_2*\omega_3\\
&\quad+\lambda_{1k}\gamma_{17}\omega_1*\omega_2*\omega_3)+(\upsilon_{0j}+\lambda_{1k}\upsilon_{1j}+\varepsilon_{kj})\end{aligned} \tag{2.8}$$

(k=1, …, T; j=1,…, n; λ_{11}=0, λ_{14}=1). Thus, using the same notation as before, a unified modeling approach combining both the item response model and the latent growth curve analysis can be written as below, which is the derivative of the generalized linear latent and mixed model (GLLAMM), since the model contains both fixed (ω_j) and random effects (υ_{0j} and υ_{1j}), and has latent abilities (η_{kj}) being regressed upon other factors and observed covariates.

$$\begin{aligned}\text{probit}[Pr(Y_{kij}=1|\omega_j,\upsilon_{0j},\upsilon_{1j})]&=v_{kij}=\alpha_i(\eta_{kj}-\beta_i)=\alpha_i(\zeta_{0j}+\lambda_{1k}\zeta_{1j}-\beta_i)\\
&=\alpha_i((v_{00}+\lambda_{1k}v_{10})+(\gamma_{01}\omega_1+\gamma_{02}\omega_2+\gamma_{03}\omega_3+\gamma_{04}\omega_1*\omega_2+\gamma_{05}\omega_1*\omega_3+\gamma_{06}\omega_2*\omega_3+\gamma_{07}\omega_1*\omega_2*\omega_3)\\
&\quad+(\lambda_{1k}\gamma_{11}\omega_1+\lambda_{1k}\gamma_{12}\omega_2+\lambda_{1k}\gamma_{13}\omega_3+\lambda_{1k}\gamma_{14}\omega_1*\omega_2+\lambda_{1k}\gamma_{15}\omega_1*\omega_3+\lambda_{1k}\gamma_{16}\omega_2*\omega_3+\lambda_{1k}\gamma_{17}\omega_1*\omega_2*\omega_3)\\
&\quad+(\upsilon_{0j}+\lambda_{1k}\upsilon_{1j}+\varepsilon_{kj})-\beta_i)\end{aligned} \tag{2.9}$$

$$(i=1,\ldots I;\ k=1,\ldots,T;\ j=1,\ldots,n;\ \lambda_{11}=0,\ \lambda_{14}=1).$$

Furthermore, with a sampling distribution assumption being imposed, $Y_{kij}=1|\pi_{kij}$ ~binomial(1, π_{kij}) (2.10), where $\pi_{kij}=Pr(Y_{kij}=1|\omega_j,\upsilon_{0j},\upsilon_{1j})$, Raudenbush and Bryk (2002) concluded that this GLLAMM can be categorized into three subcomponents: (1) the level-1 sampling model (Equation (2.10)); (2) the link function (Equation (2.9) left-hand side), and (3) the structural model (Equation (2.9) right-hand side).

In short, although many other techniques have been developed to capitalize on the special features of longitudinal research, the latent growth curve analysis provides a broad class of statistical methods which are highly flexible in model articulation, provide enhanced statistical power for testing hypotheses, and demonstrate greater correspondence between the statistical model and the traditional theory underpinning developmental trajectories (Preacher et al., 2008). That is, the LGC model is capable of allowing straightforward examination of intraindividual change as well as interindividual variability. More importantly, apart from its capabilities that lead to greater understanding and discernment of the developmental trajectory, LGC modeling is appealing for making possible inquiry into the antecedents, processes, and consequences of change (Willett and Sayer, 1994).

3. BAYESIAN INFERENCE

3.1 Estimating statistical complex models using the Markov chain Monte Carlo (MCMC)

Bayes' theorem is the centerpiece of Bayesian inference, $f(\Omega|Y)=f(Y|\Omega)*f(\Omega)/\int f(Y|\Omega)f(\Omega)d\Omega$, where $f(\Omega|Y)$ denotes the posterior probability density function, Y denotes the observed response data, and Ω represents the unknown model parameters and latent measurement variables. The "$f(.)$" can be replaced by "$p()$" and "$\sum$" can take the place of "$\int$" when we have data and parameters of a discrete nature. This posterior density can be used to determine model parameter estimates; the quantity $f(Y|\Omega)$ denotes the likelihood function of the model parameters given the response data (Y), and $f(\Omega)$ is the prior distribution for the model parameters, representing the relative likelihoods of particular parameter values before accessing the data (Kim and Bolt, 2007). In contrast to the frequentist perspective, from a Bayesian point of view the parameters of interest are treated as random quantities and probabilities are assigned to each model that possibly could have generated the data (Johnson, Sinharay, and Bradlow, 2007). Described as a normalizing constant, the quantity in the denominator, $\int f(Y|\Omega)f(\Omega)d\Omega$, is a scaling factor and makes the posterior density a proper one (i.e., non-negative densities which integrate to one). However, this normalizing constant is usually unknown and hard to analytically determine. Thus, Bayes' theorem can be expressed as $f(\Omega|Y) \propto f(Y|\Omega)*f(\Omega)$, indicating that the joint posterior density is proportional to the product of the likelihood function and the prior density for the model parameters, which makes for an appealing feature associated with the Bayesian approach: that the posterior distributions obtained include information regarding both the observational setting and the substantive domain.

When the model becomes complex, this joint posterior distribution tends to become numerically or analytically intractable. This is because calculating this posterior density typically requires a large summation and/or multidimensional integrals. In order to solve this intractability problem, the use of Monte Carlo integration was revisited by Bayesian statisticians in the late 1980s. A random sequence or chain is generated, such that in the long run each parameter value occurs with a frequency proportional to $f(\Omega|Y)$. In addition, the chain is generated so that each value in the sequence depends only on its immediate predecessor, which under certain conditions makes it a finite order Markov process (Kim and Bolt, 2007; Rupp et al., 2004; Thompson, Palmer, and Moreno, 2006; Western, 1999). Possessing these two properties, this sampling procedure is named the Markov chain Monte Carlo (MCMC), the goal of which is to reproduce the joint posterior distribution through simulation (e.g., Jackman, 2000; Kim and Bolt, 2007; Lynch and Western, 2004; Patz and Junker, 1999b). By sampling enough observations, researchers could obtain a general description of the posterior distribution, such as the expected a posteriori (*EAP*; the mean of the posterior density), maximum a posteriori (*MAP*; the mode of the posterior density), posterior standard deviation (*PSD*; standard deviation of the posterior density), the 95% credible interval, etc.

3.2 Sampling procedures

The mechanism by which sampling is conducted varies depending on the known features of the posterior distribution, $f(\Omega|Y)$. In general, various types of sampling algorithms are considered within MCMC, two of which are the Metropolis-Hastings algorithm (e.g., Chib and Greenberg, 1995) and the Gibbs sampling (e.g., Casella and George, 1992). Also known as rejection sampling, the key to the former is trying to find a suitable candidate-generating density (i.e., a proposal density, $q(\underline{\theta}^*|\underline{\theta}_t)$) to use for suggesting a new value, $\underline{\theta}^*$, given the current value in the chain $\underline{\theta}_t$. The choice of a proposal distribution affects the efficiency of the algorithm: a good choice of proposal distribution will make the chain converge quickly to the long-run probabilities; however, a poor choice of proposal distribution will leave the chain stuck while generating parameter values and slow down the convergence of the sequence (Thompson et al., 2006). Usually, the Metropolis-Hastings algorithm is needed when estimating logistic item response models, for the complete conditional distributions are not of a known distribution form (Kim and Bolt, 2007). To make the Markov chain reach convergence reasonably fast, Patz and Junker (1999b) suggested the use of Metropolis-Hastings within Gibbs (MHwG) for the two- and three-parameter logistic model as well as the generalized partial credit model. Interested readers may refer to Chib and Greenberg, 1995) for further discussion.

As a special case of the Metropolis-Hastings, the Gibbs sampling involves cycling through smaller subsets of parameters of the model[9] and using the current estimate of the full conditional posterior distribution as the proposal density (Casella and George, 1992; Chib and Greenberg, 1995; Fox, 2007; Gelfand et al., 1990; Patz and Junker, 1999a, 1999b; Thompson et al., 2006). In a situation of

---

[9] The parameters may themselves be univariate or multivariate, such as sampling from the full conditional posterior distributions of each unknown or blocks of unknowns; however, such technique as parameter expansion, updating parameters in blocks instead of one by one, has a dramatic impact on computational efficiency and helps improve the mixing rate of Markov chains (Dunson et al., 2005; Fox, 2007).

updating one parameter at a time, the univariate conditional distribution, $f(\Omega_k|Y, \Omega_{-k})$, represents the posterior distribution of a single model parameter Ω_k, conditional upon the data and all other model parameters Ω_{-k}. Being a "divide and conquer" strategy, sometimes the Gibbs sampler may be inefficient, moving slowly over the parameter space (Western, 1999); however, due to its use of known conditional distributions for simulation, this setup helps reduce multidimensional problems to a series of univariate calculations and it is easier to simulate draws (Casella and George, 1992; Jackman 2000; Patz and Junker, 1999a, 1999b). Interested readers may refer to Gelman et al. (2003) for an introduction to the Gibbs sampler in multilevel models.

3.3 Specification of priors

As noted earlier, the posterior distributions from the Bayesian inferences depend not only on the data through the likelihood function but also on the prior density (e.g., Western, 1999): thus, the specification of prior distributions for each of the model parameters (such as the latent variables and the parameters from the measurement and structural models) plays an important role in the Bayesian approach. That is, unlike those of the frequentists, Bayesian methods provide a clear channel for us to incorporate prior information, which helps increase the statistical power of the analysis and contributes to the accumulation of scientific findings (Hsieh and Maier, 2009). Moreover, based on Bayes' law, whenever our prior is uniformly distributed in the region where the likelihood function is located, the posterior distribution for the Bayesian function is nearly proportional to the likelihood function (Gill, 2002; Maier, 2001; Rice, 1995).

Furthermore, as sample sizes increase, priors are generally asymptotically irrelevant, and the estimates obtained from the Bayesian and frequentist should approach an identical value (Dunson, Palomo, and Bollen, 2005; Lynch and Western, 2004; Western, 1999).

In this sense, the Bayesian method can be treated as a direct alternative to the maximum likelihood estimates (MLEs) for parameter estimation when using non-informative priors.

A long-running debate in Bayesian inference revolves around the choice between subjective priors and objective priors, in which the subjective priors indicate the inclusion of existing subject-matter knowledge, and objective priors remove any subjectivity from the analysis. Although the role of the prior diminishes as sample size increases, inferences may be sensitive to the choice of the prior (Gill, 2002; Kim and Bolt, 2007).

In practice, there is a preference for objective reference priors, for they resolve the dispute between Bayesian and likelihood approaches, which results in proper but diffuse priors as a popular choice (Lynch and Western, 2004).

However, informative subjective priors allow researchers to build on previous research, and can be justified on the basis of opinion elicited from scientific specialists, archival materials, and the weight of established evidence (e.g., Lee and Wagenmakers, 2005). As Congdon (2005) suggests, one may carry out an evidence synthesis using forms of meta-analysis to set an informative prior, but this is beyond the scope of the present study.

Seeing that the prior densities are needed to define the posterior distribution, it is desirable to select conjugate priors whenever possible. Adopting conjugate priors implies that the distribution of the posterior is already known and of the same form as the prior density, which makes sampling in MCMC to be computationally efficient (Johnson et al., 2007; Kim and Bolt, 2007; Rupp et al., 2004).

In other words, in assigning noninformative priors to the model parameters of interest, the researcher allows the data to provide as much information as possible by themselves. Thus, in subsequent analyses, non-informative and conjugate priors are used for most important parameters[10].

For instance, the means of initial level and shape are estimated using normal distribution priors, and two kinds of noninformative priors are used for the variance of measurement error: the inverse gamma prior and the uniform distribution prior (Gelman and Hill, 2007).

In regard to the covariance matrix of the random effect parameters, the conjugate prior, the inverse Wishart distribution is adopted. The complete specification of different priors can be found in Table 1.

3.4 Monitoring the Markov chain(s) and evaluating the model goodness of fit

For the model estimated via the Bayesian Markov chain Monte Carlo (MCMC), using Gibbs sampling (Chib and Greenberg, 1995; Gelfand et al., 1990; Patz and Junker, 1999a) as implemented in WinBUGS 1.4 (Spiegelhalter et al., 2003), the 'burn-in' period for the MCMC chain(s) was determined using the methods

---

[10] In order to facilitate the model identification, a normal prior with tight precision, $\beta \sim N(0,1)$, was utilized for item difficulty parameters, and a truncated normal prior, $N(0, 1.0E\text{-}02)I(0, \ )$ was adopted for item discrimination parameters.

proposed by Geyer (1992) and Gelman and Rubin (1992) . As Geyer (1992) claims, generating one single long chain is more efficient in using the simulation output, although it leads to more complex Monte Carlo standard error expressions[11]. As opposed to running one long sequence, Gelman and Rubin (1992) argue that, to monitor convergence of the Gibbs sampler, it is important to run multiple chains using a range of different starting values (Seltzer, Wong, and Bryk, 1996). Thus, in the present study we perform a sensitivity analysis by contrasting these two methods.

Table 1. Different types of prior used in the present study.

Measurement model	
Parameter	Baseline priors
$\beta2$ $\beta3$ $\beta4$	dnorm(0, 1)
$\alpha2$ $\alpha3$ $\alpha4$	dnorm(0, 1.0E-02)I(0,)
Structure model	
Parameter	Non informative priors
$S2$ $S3$	dnorm(0, 1.0E-4)
μ_L μ_S	dnorm(0, 1.0E-4)
$\begin{pmatrix}\sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2\end{pmatrix}^{-1}$	dwish$\left[\begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix},2\right]$
$1/\sigma_{et}^2$	(1) $1/\sigma_{et}^2$ ~dgamma(.001, .001) (2) σ_{et}^2 ~dunif(0,1.0E04)

Regardless of the method, we need to specify a series of observations as each state in a Markov chain. In order to begin the sampling process, we need an initial set of values, treated as the starting values for the model parameters. They can be generated either by random variables or obtained from existing maximum likelihood-based estimation programs whenever possible. However, as noted by Kim and Bolt (2007) and Thompson et al. (2006), the choice of starting values may influence the sequence of values produced, and successive values may be highly correlated in the early stage of the chain. In this case, simulated values cannot be treated as a random sample from the posterior distribution. Thus, it is common to disregard a number of the initial iterates, treat them as the burn-in period, and estimate the posterior distribution using the remaining iterates. That is, in order to ensure that each chain has converged to its stationary distribution and stable parameter estimates have been obtained, one normally allows for a burn-in period of some length, and makes use of the subsequent simulated states to construct the posterior distribution (e.g., Kim and Bolt, 2007; Patz and Junker, 1999b).

Several methods have been proposed for model comparison, based on Bayesian principles; for instance, Spiegelhalter and his colleagues (2002) proposed the deviance information criterion (*DIC*), which includes many features of classical model assessment, such as requiring accurate predictions and penalizing complexity. Being composed of two major elements, the posterior distribution of the deviance (i.e., -2*log-likelihood, denoted as *D*) and an effective number of parameters (p_D), mathematically, the *DIC* is defined as $DIC=\overline{\mathrm{D}(\Omega)}+p_D$, where Ω denote all the parameters under the assumed model, and $\overline{\mathrm{D}(\Omega)}$ is a measure of lack of fit, representing an estimated average discrepancy between model and data; p_D accounts for the expected decrease in deviance attributable to the added parameters of the more complex model (Fox, 2007; Li, Bolt, and Fu, 2006).

As the model diagnosis and evaluation criterion, estimation of the *DIC* index can be requested from the WinBUGS program: a smaller *DIC* represents a better fit of the model. In addition, a difference of less than five or ten units between models does not provide sufficient evidence for favoring one model over another (Spiegelhalter et al., 2003).

In addition to the *DIC*, the posterior predictive check (*PPC*) is a criterion used for assessing the model goodness of fit (Gelman, Carlin, Stern, and Rubin, 2003).

Mathematically, the posterior predictive distribution can be written as: $p(\mathrm{Y}^{\mathrm{rep}}|\mathrm{Y})=\int p(\mathrm{Y}^{\mathrm{rep}}|\Omega)p(\Omega|\mathrm{Y})d\Omega$, where $\mathrm{Y}^{\mathrm{rep}}$ denotes replicated values of Y, and Ω represents all model parameters (e.g., Sinharay and Stern, 2003). The integral defining the posterior predictive distribution consists of two parts: the sampling distribution ($p(\mathrm{Y}^{\mathrm{rep}}|\Omega)$) and the posterior distribution for model parameters ($p(\Omega|\mathrm{Y})$). That

---

[11] Because the posterior distributions are constructed from simulated samples, errors in the estimates can be attributed to the standard deviation of the posterior as well as the sampling error. Here, the sampling error is referred to as the Monte Carlo standard error (*MCSE*) (Patz and Junker, 1999b; Spiegelhalter et al., 2003).

is, the posterior predictive distribution takes the following two uncertainties into account: sampling uncertainty and model uncertainty (Lynch and Western, 2004; Rupp et al., 2004; Western, 1999). The rationale behind posterior predictive checks involves simulating data under the model stated in the null hypothesis and comparing the features of these replicated data with the observed ones. This approach grants the researcher a wide range of fit statistics; the overall discrepancy statistics utilized in the present study is the Bayesian chi-square, $T(Y_{kij}, \Omega_{kij})=\sum (Y_{kij}-E(Y_{kij} |\Omega_{kij}))^2/var(Y_{kij} |\Omega_{kij})$, the sum of squares of the outfit measures[12].

Specifically, being a quantitative measure of lack of fit, with simulated iterates generated from the posterior distribution, the Bayesian *p* value (also known as the PPP-value) can be assessed by comparing the observed T(Y) to the replicated $T(Y^{rep})$, and defined as $p=Pr(T(Y^{rep})\geq T(Y)|Y)$, where this tail-area probability (or *p*-value) is estimated from the simulation as the proportion of the *N* replications for which $T(Y^{rep})\geq T(Y)$, and can be interpreted as the probability of observing extreme data conditional on the model (Lynch and Western, 2004; Sinharay and Stern, 2003; Sinharay, Johnson, and Stern, 2006). Thus, any systematic discrepancy between the replications and observed data reflects the implausibility of the data under the model, and suggests that the presumed model doesn't fit the data well (Li et al., 2006; Lynch and Western, 2004; Sinharay and Stern, 2003; Sinharay et al., 2006). Usually, the PPP-value under the correct model tends to be closer to .5; however, if the posterior predictive *p* values are extreme, being close to zero, one, or both (depending on the nature of the discrepancy measure), it is clear that the observed response would be unlikely to occur provided that the null hypothesis is true (Sinharay and Stern, 2003; Sinharay et al., 2006).

4. PRACTICAL ILLUSTRATION: FIT OF IRM-LGC TO THE ABORTION DATA USING WinBUGS1.4

a. Measures and data sources

As part of the investigation of British social attitudes, the data represent the responses to seven items concerning attitudes toward abortion by a selected panel sample of 410 from the years 1983 to 1986. For each item, respondents were asked if they agreed that the law should allow abortion: where 1 stands for "agree" and 0 otherwise. These seven items are listed in Table 2[13].

However, when we perform a confirmatory factor analysis (CFA) to examine the underlying construct using the software of Mplus (Muthén and Muthén, 2006), we find these seven items do not measure the same thing: that is, these items do not form a unidimensional construct. As a simplified demonstration, we decide to focus on participants' general attitudes toward abortion (measured by the bottom four items in Table 2) and remove the extreme circumstance factor from subsequent analyses (Skrondal and Rabe-Hesketh, 2004).

By doing so, the assumption of local independence can be entertained, and gamma change can be ruled out through conducting a CFA on the scale at four time periods[14].

That is, a single underlying latent variable helps explain the whole association between the responses to different items by an individual, and all items load onto this single latent factor across the entire study span.

Table 2. The seven items concerning attitudes to abortion on the British Social Attitudes Panel Survey, 1983-1986

Here are a number of circumstances in which a woman might consider an abortion. Please say whether or not you think the law should allow an abortion in each case. Should abortion be allowed by law?

Extreme circumstance factor:
1. [Risk] the woman's health is seriously endangered by the pregnancy.
2. [Rape] the woman became pregnant as a result of rape.
3. [Defect] there is a strong chance of a defect in the baby.

General attitude factor:
4. [Financial] the couple cannot afford any more children.
5. [Marriage] the woman is not married and does not wish to marry the man.
6. [Couple] the couple agree that they do not wish to have the child.
7. [Woman] the woman decides on her own she does not wish to have the child.

---

[12] Even though it has advantages over standard applications of fit statistics, however, this chi-square-type measure should be interpreted with great caution. This is because in IRTmodel checking, it is not a suitable discrepancy measure and fails to detect the problems with inadequate psychometrics models (Sinharay et al., 2006).

[13] Data were supplied by the UK Data Archive. Neither the original data collectors nor the archive bear any responsibility for the analyses.

[14] In Golembiewski et al.'s triumvirate conceptualization of longitudinal change (1976), they claim that the true change (a.k.a. the alpha change) can only be inferred from observed scores in a situation when there are no beta and gamma changes, where beta change is defined as the change resulting from the respondent's recalibration of the measurement scale over time, and gamma change refers to as a fundamental change concerning the respondent's understanding and perception of the latent constructs of primary interest.

Moreover, the breakdown analyses and response pattern for complete cases and available cases can be found from Table 3 to Table 6. In our analyses, only approval or disapproval responses were counted as valid and other responses were treated as item non-response, which results in 284 respondents giving complete responses for all four years. However, if the responses of "don't know" and "no answer" are included, we finally have a usable sample of 323 cases.

As observed in the response pattern for each data set, it is found that in the contingency table we have a few response patterns with large frequencies and many response patterns with small frequencies, which implies that the data form a rather sparse contingency table and the asymptotic normality of the maximum likelihood estimator cannot be obtained, since in both data sets some of the 2^4 possible response patterns are not observed. Thus, when frequentist methods are adopted, all kinds of problems associated with this sparseness such as statistical inference and hypothesis testing should be constantly borne in mind (Knott, Albanese, and Galbraith, 1990; Fienberg and Rinaldo, 2007).

Table 3. *Breakdown table for the restricted data*

Latent variable outcomes			Attitude 1983	Attitude 1984	Attitude 1985	Attitude 1986
Gender	Female(0)	*n*	160	160	160	160
		Mean	.261	-.208	.262	.439
		SD	1.709	1.649	1.710	1.592
	Male (1)	*n*	124	124	124	124
		Mean	.349	-.069	.494	.860
		SD	1.856	1.630	1.806	1.573
Age	Senior (0)	*n*	141	141	141	141
		Mean	.126	-.319	.161	.527
		SD	1.702	1.593	1.792	1.661
	Junior (1)	*n*	143	143	143	143
		Mean	.470	.022	.563	.717
		SD	1.827	1.67	1.697	1.526
Religion	Yes (0)	*n*	182	182	182	182
		Mean	.095	-.417	.124	.375
		SD	1.840	1.538	1.742	1.567
	No (1)	*n*	102	102	102	102
		Mean	.664	.333	.791	1.064
		SD	1.586	1.711	1.698	1.556
Total		*N*	284	284	284	284
		Mean	.299	-.147	.364	.623
		SD	1.771	1.640	1.753	1.595

Note.1. Each of these three explanatory variables were dichotomized as follows: gender (0: female vs. 1: male), age (0: elder (>40) vs. 1: young respondents (<=40)), and religious status (0: have religion vs. 1: no religion).

Table 4. *Breakdown table for the full data.*

Latent variable outcomes			Attitude 1983	Attitude 1984	Attitude 1985	Attitude 1986
Gender	Female(0)	*n*	180	180	180	180
		Mean	.256	-.312	.169	.386
		SD	1.577	1.808	1.588	1.629
	Male (1)	*n*	143	143	143	143
		Mean	.419	-.283	.343	.798
		SD	1.721	1.758	1.708	1.607
Age	Senior (0)	*n*	157	157	157	157
		Mean	.153	-.410	.026	.411
		SD	1.664	1.878	1.667	1.680
	Junior (1)	*n*	166	166	166	166
		Mean	.493	-.195	.454	.718
		SD	1.608	1.689	1.595	1.572
Religion	Yes (0)	*n*	204	204	204	204
		Mean	.032	-.475	.012	.349
		SD	1.554	1.741	1.618	1.602
	No (1)	*n*	119	119	119	119
		Mean	.836	.001	.648	.946
		SD	1.670	1.824	1.610	1.615
Total		*N*	323	323	323	323
		Mean	.328	-.299	.246	.569
		SD	1.642	1.783	1.642	1.630

Note.1. Each of these three explanatory variables were dichotomized as follows: gender (0: female vs. 1: male), age (0: elder (>40) vs. 1: young respondents (<=40)), and religious status (0: have religion vs. 1: no religion).

Table 5. *Frequencies of the response patterns observed for the 1983-1986 panels (complete cases)*

1983			
Response pattern	Observed frequencies	Response pattern	Observed frequencies
1111	95	1001	8
0000	70	0010	8
1000	20	1100	7
1110	19	0111	4
0011	12	0110	4
1010	10	1101	3
1011	10	0101	3
0100	9	0001	2
1984			

Response pattern	Observed frequencies	Response pattern	Observed frequencies
0000	121	1010	6
1111	70	1101	5
1000	20	0011	4
1110	14	0001	4
0100	10	0111	3
0010	8	1001	2
1100	8	0110	1
0101	7	1011	1

1985

Response pattern	Observed frequencies	Response pattern	Observed frequencies
1111	96	1011	6
0000	86	0101	5
1000	21	0010	5
1110	19	1010	4
0111	9	0110	4
1100	9	1101	3
0011	8	0001	2
0100	7		

1986

Response pattern	Observed frequencies	Response pattern	Observed frequencies
1111	107	1010	6
0000	72	1101	5
1110	32	0110	3
1100	17	0011	3
0111	12	1011	2
1000	9	0001	1
0100	8		
0010	7		

Table 6. Frequencies of the response patterns observed for the 1983-1986 panels (available cases)

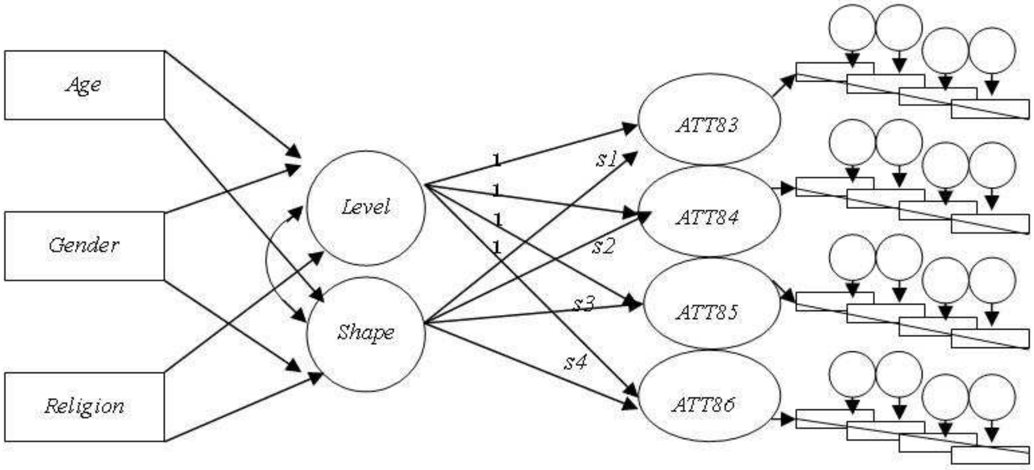
1983

Response pattern	Observed frequencies	Response pattern	Observed frequencies
1111	102	1001	8
0000	85	1100	8
1110	21	9999	5
1000	21	0111	4
0011	14	0110	4
1010	13	1101	3
1011	10	0101	3
0100	10	0001	2

0010	10		
1984			
Response pattern	Observed frequencies	Response pattern	Observed frequencies
0000	134	1010	7
1111	73	1101	5
1000	24	0001	5
1110	17	0011	4
9999	13	0111	3
0100	11	1001	2
1100	8	1011	1
0010	8	0110	1
0101	7		
1985			
Response pattern	Observed frequencies	Response pattern	Observed frequencies
1111	99	1011	6
0000	93	0101	5
9999	23	0010	5
1110	21	1010	4
1000	21	0110	4
1100	10	1101	3
0111	9	0001	2
0011	9		
0100	9		
1986			
Response pattern	Observed frequencies	Response pattern	Observed frequencies
1111	117	1010	6
0000	85	0110	4
1110	36	1011	3
1100	18	9999	3
0111	12	0011	3
1000	12	0001	1
0100	9		
0010	8		
1101	6		

The sampling method was a multi-stage design with multiple separate stages of selection, where selecting respondents were nested within addresses, addresses within polling districts, polling districts within constituencies, and constituencies within the electorate (The British Social Attitudes Panel Survey, 1983-1986). Given that a key task of an annual series survey was to look at trends and changes in attitudes over time, a longitudinal rather than a repeated cross-sectional design was adopted (McGrath and Waterton, 1986; Wiggins, Ashworth, O'Muircheartaigh, and Galbraith, 1990). In the present study, we aim to extend our concentration on the methodological issues: that is, the proposal and evaluation of an IRM-LGC hybrid model. Because a growth curve analysis is utilized to model the process of change, the estimation of growth profiles is represented by the parameters of initial level and shape, along with other explanatory variables. A conceptual modeling framework is depicted in Figure 1, which follows the graphic conventions in Muthén and Muthén (2006).

Figure 1. Path diagram of a four-wave IRM-LGC model



4.2 Unconditional models

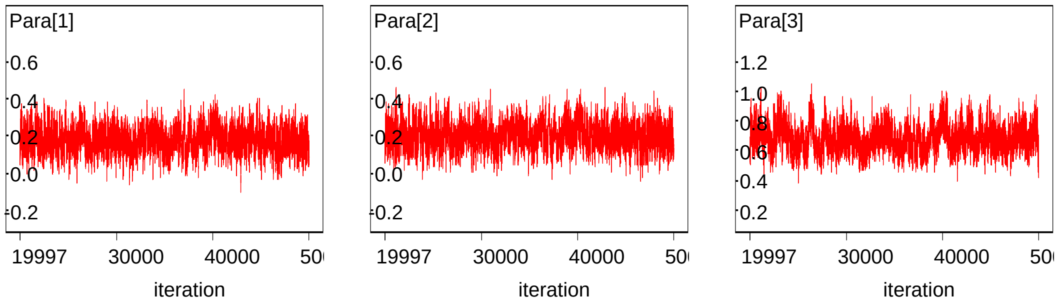
In order to examine the robustness of the obtained Bayesian results, the monitoring of three independent chains with overdispersed initial values and the convergence assessment of one single long chain were performed. It was found that the results from these two approaches were close to each other within at least one decimal place: in the situation of running three independent chains, the first 20,000 iterations were discarded as burn-in for each chain, which results in a total of an additional 30,003 iterations for the three chains and they were carried out to define the posterior distribution of each parameter. Similarly, for a single long chain, we used a burn-in period of 19,998, with parameter estimates based on the 50,000 subsequent iterations (see Figures 2-3). The output was summarized on the basis of the remaining 30,003 iterations.

Generally, the simulation should be run until the Monte Carlo standard error associated with each parameter is within an acceptable range, say, less than 5% of the sample standard deviation (Dunson et al., 2005; Kim and Bolt, 2007; Spiegelhalter et al., 2003). However, compared to the results obtained from the multiple-chain approach, it is found that the Monte Carlo errors are not all less than 5% of the sample standard deviation when we adopt one single long chain to generate the simulated sample. While using multiple independent chains, however, most of the Gelman-Rubin statistics, with the potential scale reduction factor (PSRF), approximately approach one for each quantity of interest (Gelman and Rubin, 1992), which indicates the reaching of convergence (see Figures 4-6). Thus, in subsequent analyses we adopt Gelman and Rubin's suggestion and monitor the model convergence using three independent chains with over-dispersed starting values.

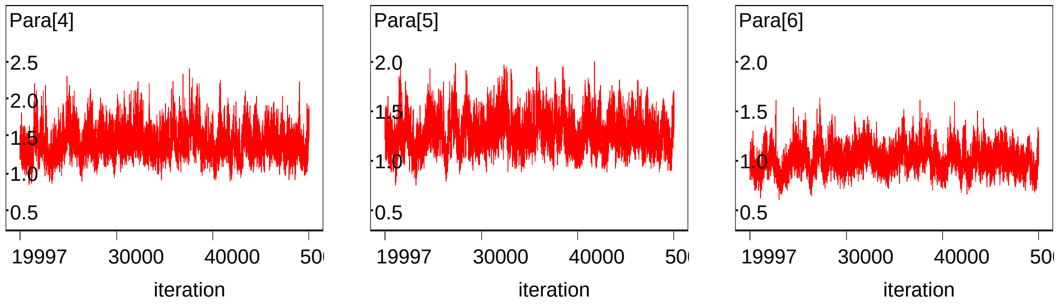
Based on the results from Table 7, in considering a few candidate models, it is found that all of them provide convergent substantive interpretation; thus, according to the model goodness of fit index (i.e., *DIC*), we take the model in the column on the extreme right, the one with the probit link and uniform prior for level-1 residual variances, as an example of the adequate representation of the data. Again, the results of parameter estimates and associated standard deviations from the complete data set (*n*=284) are given in Table 8 (the right panel), where we see that the estimated discrimination parameters $\hat{a}_i$ for item 2 and item 3 are both greater than 1 and larger than for the other two items, indicating that item 2 and item 3 better discriminate person ability than do items 1 and item 4.

Figure 2. *Time series plots for the restricted data (uniform prior for varying residuals): single long chain*

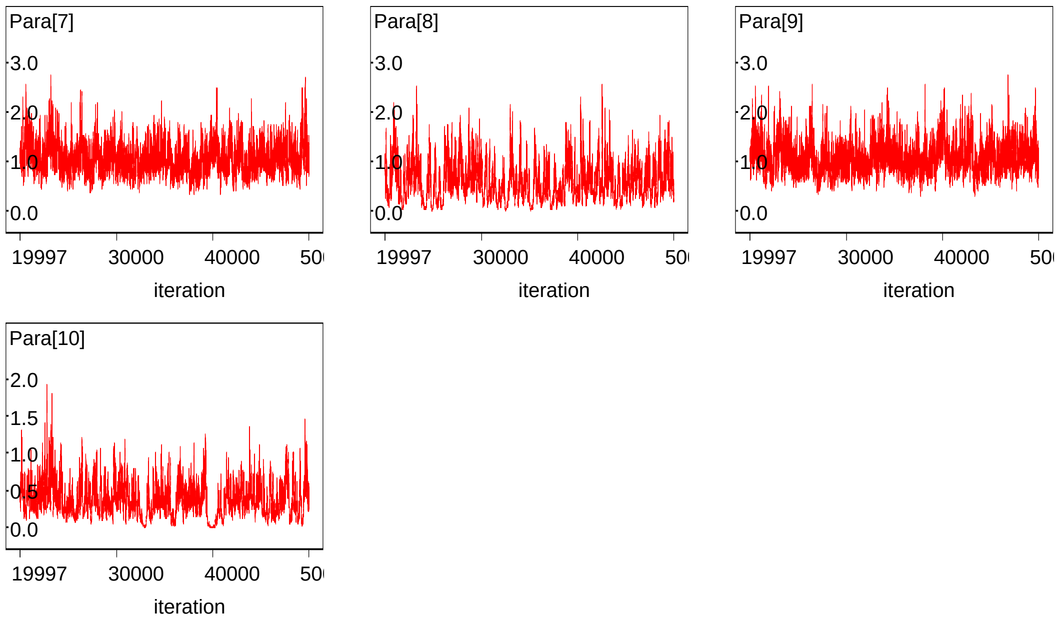
Item difficulty parameter estimates:



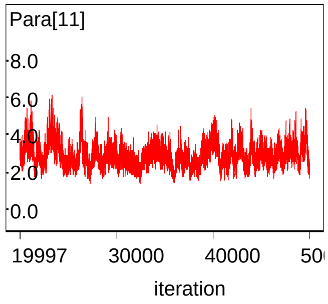
Item discrimination parameter estimates:

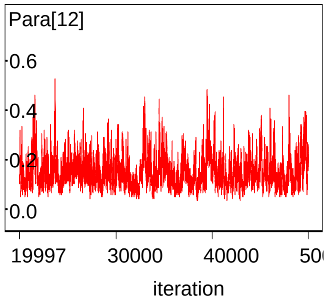


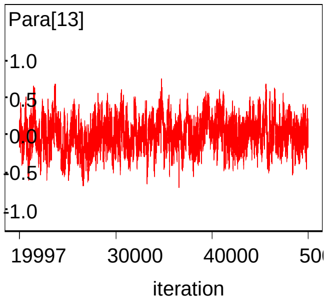
Level1 varying residuals:



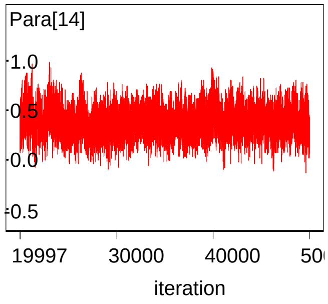
Level2 variance-covariance structure:

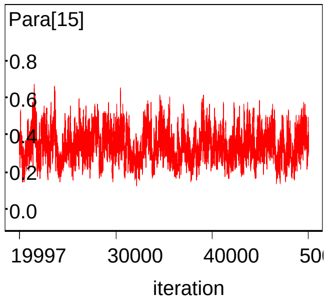




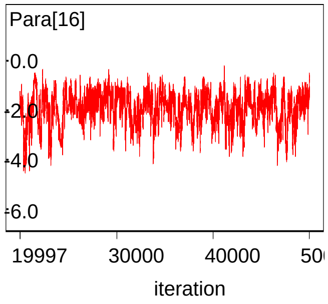


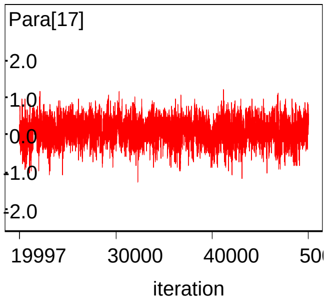
Initial level and shape:





Coefficients of time variable





Posterior predictive check: Bayesian p-value

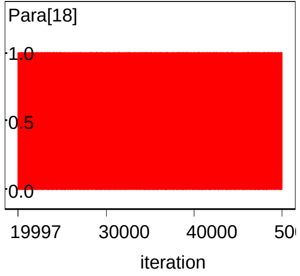


Figure 3. Kernel density for the restricted data: one single long chain (excerpted).

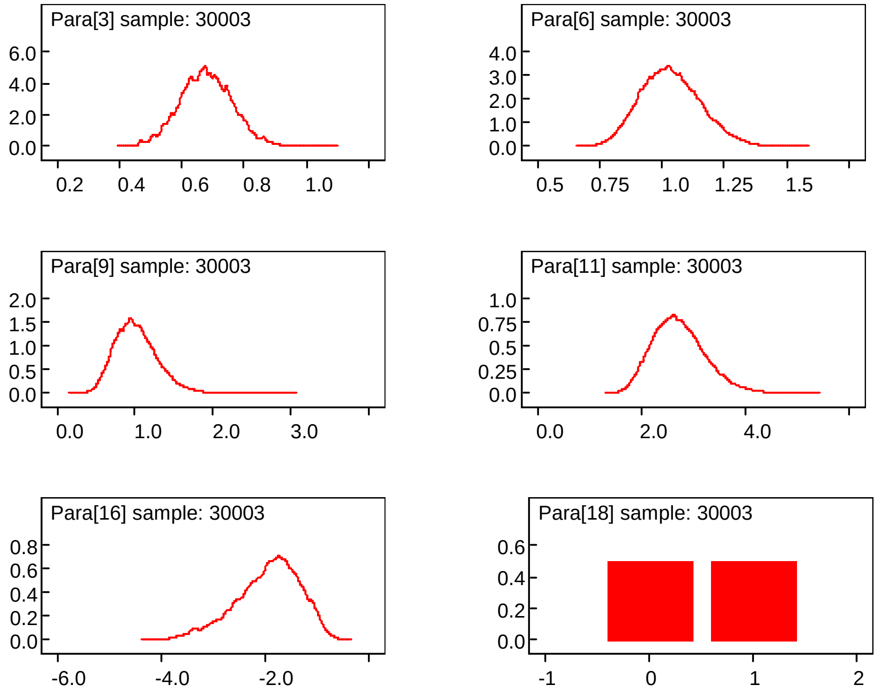
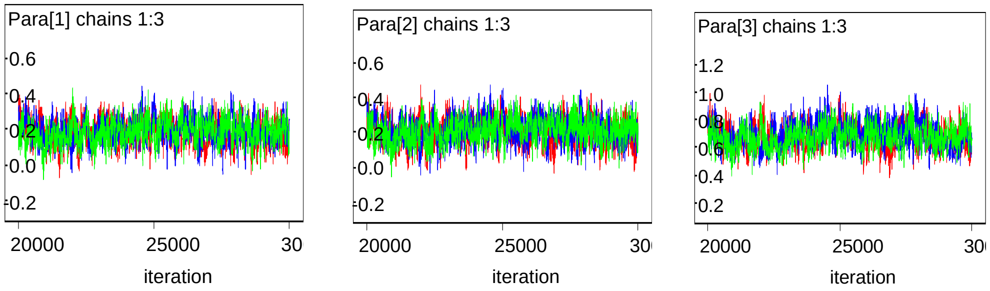
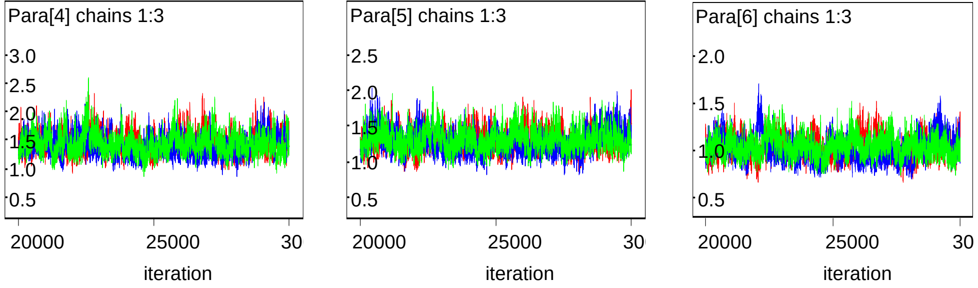


Figure 4. Time series plots for the restricted data (uniform prior for varying residuals): three independent chains.

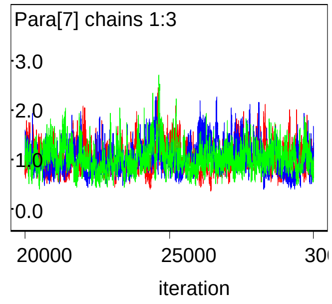
Item difficulty parameter estimates:

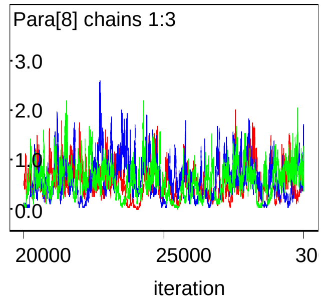


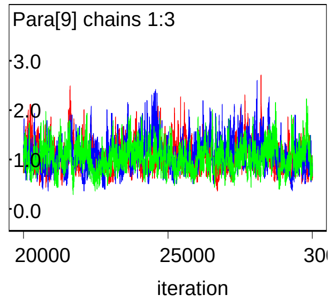
Item discrimination parameter estimates:

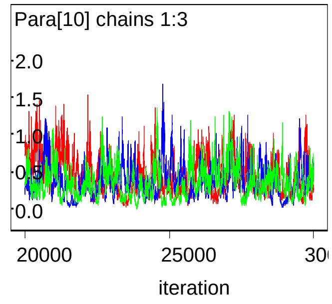


Level 1 varying residuals:

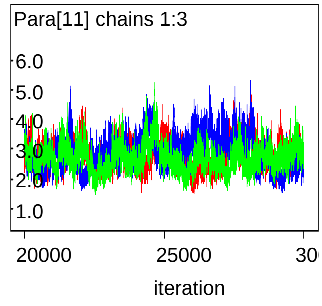


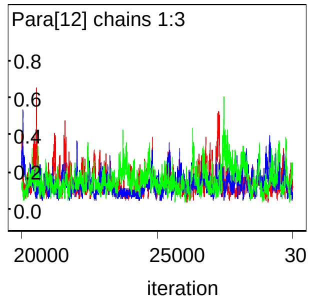


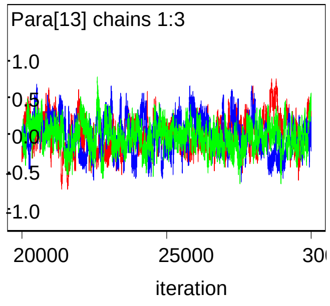




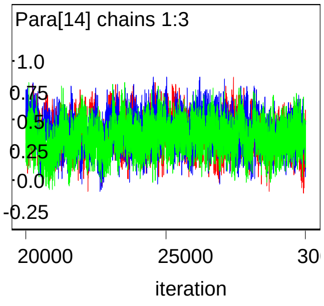
Level2 variance-covariance structure:

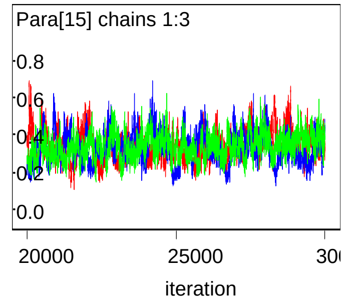




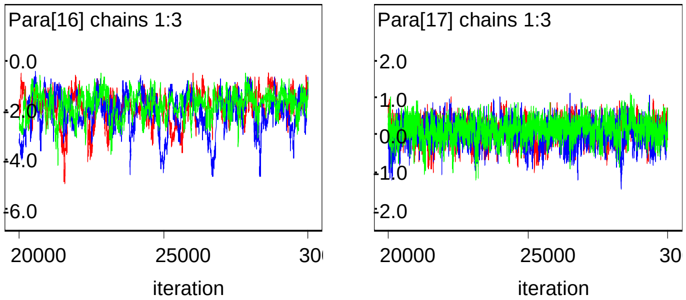


Initial level and shape:





The coefficients of time variable:



Posterior predictive check: Bayesian p-value

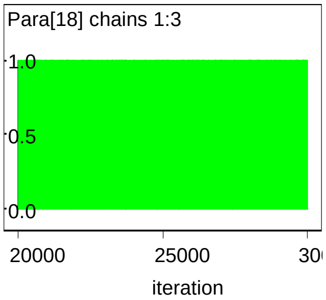


Figure 5. *Kernel density for the restricted data: three independent chains (excerpted).*

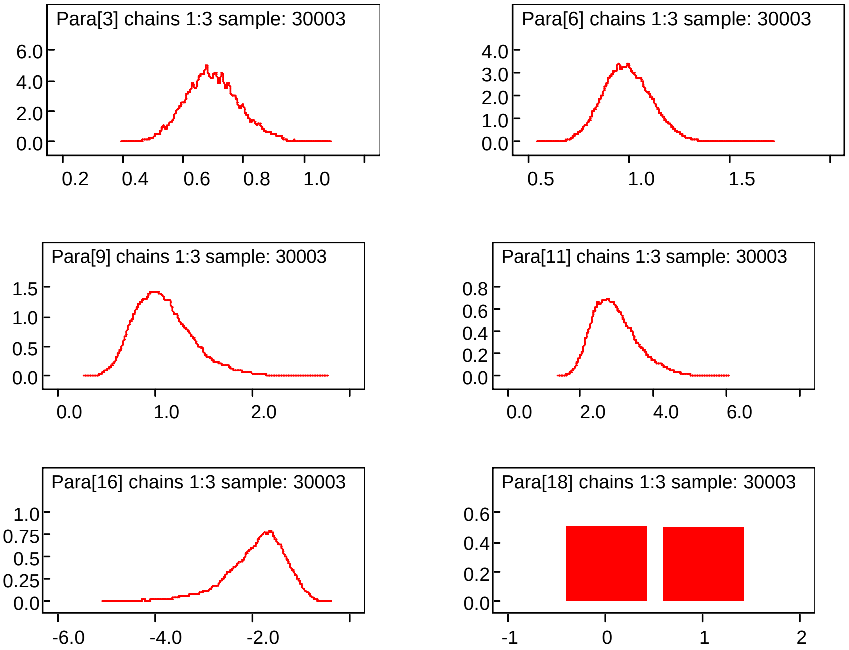
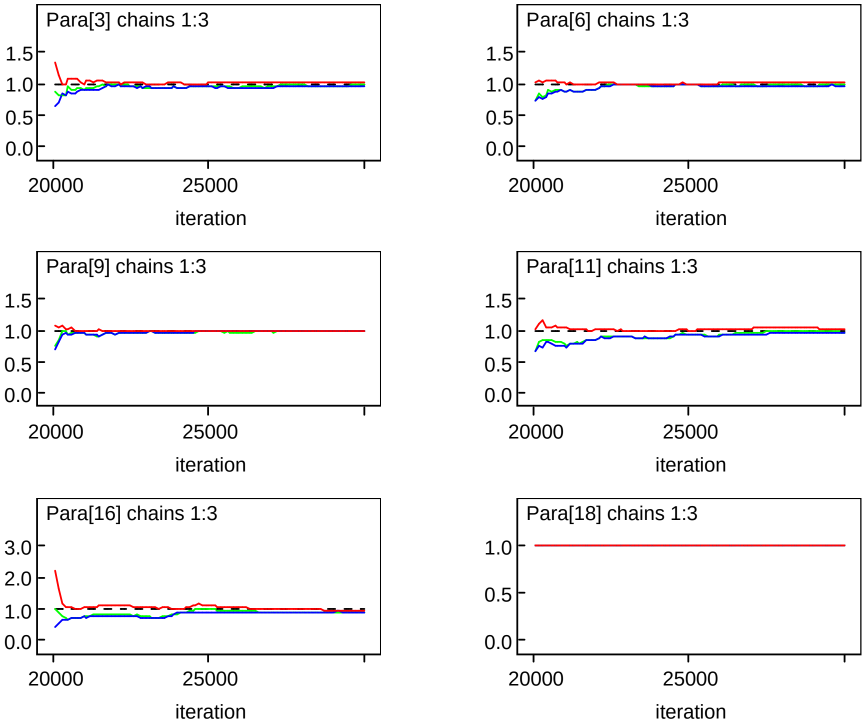


Figure 6. *Gelman-Rubin statistic for the restricted dataset: three independent chains (excerpted).*



This is because greater discrimination indicates a stronger relationship between an item and the underlying latent trait; hence, we would say that the "marriage" and "couple" items are more closely related to holding a positive attitude to abortion than are the "financial" and "'woman" items. As for the item difficulty parameter estimates, the estimated difficulty parameter $\hat{b}_i$ associated with item 4 is the largest among the four, indicating that the "woman makes the abortion decision herself" is the hardest item to endorse. In other words, the endorsement of this item reflects a higher level of propensity to hold a generally positive attitude toward abortion than do other items, such as "financial", "marriage", and "couple" items.

As for the substantive interpretation of the latent growth or decline trajectory, the empirical result shows that without controlling any explanatory variables, a mean growth curve emerges with a true initial level of .392 (p<.01) and a change rate of .336 (p<.01). The significant variation between the respondents around these mean values ($\hat{\sigma}_L^2$=2.953 and $\hat{\sigma}_S^2$=.144) implies that, overall, these subjects start their growth process at different phases and go on to change at different rates, which not only reveals systematic difference in the change trajectory among participants but also suggests true variation remaining in both the initial status and rate of change, indicative of the need for additional time-invariant predictors (e.g., Singer and Willett, 2005). The correlation between the initial level and the growth rate is -.021 ($\hat{\sigma}_{LS} / \left(\hat{\sigma}_L \cdot \hat{\sigma}_S\right)$, ns), implying that the initial level has no predictive power for the growth rate. The level-1 varying residual variances, describing the measurement fallibility in general attitudes toward abortion over time (their estimated values are 1.077, .581, 1.095, and .391, respectively, being statistically significant at the first, third, and fourth points of time), suggest that the existence of additional outcome variation at level-1 may be further explained by other time-varying predictors. Finally, it is found that a piecewise linear growth trajectory exists (i.e., the estimated slopes for four repeated assessments are $s1$=0 (fixed), $s2$=-2.072 (p<.01), $s3$=.061 (ns) and $s4$=1(fixed)) in terms of participants' general attitudes toward abortion.

Table 7. *The parameter estimates of IRM-LGC model for the restricted data.*

Note. 1. Multiplying by a factor of 1.701. 2.*$p<.1$ (1.6449), **$p<.05$ (1.96), ***$p<.01$ (2.5758).

	Non informative priors α ~ dnorm (0, 1.0E-02)I(0,) and β~ dnorm(0,1)				Non informative priors α ~ dnorm (0, 1.0E-02)I(0,) and β~ dnorm(0,1)			
	Probit link		Logit link[1]		Probit link			
	gamma priors for varying residuals (~dgamma(.001, .001))				gamma priors for varying residuals (~ dgamma(.001, .001))		uniform priors for varying residuals (~ dunif(0, 1.0E04))	
	Bayesian-one single long chain (30,000 iterations, 20,000 burn-in)				Bayesian-three independent chains (30,000 iterations, 20,000 burn-in)			
	Estimate (*EAP*)	*SD*	Estimate (*EAP*)	*SD*	Estimate (*EAP*)	*SD*	Estimate (*EAP*)	*SD*
$\beta1$	.000	---	.000	---	.000	---	.000	---
$\beta2$	.201***	.071	.167**	.071	.186***	.066	.185***	.069
$\beta3$	.223***	.070	.195***	.072	.210***	.068	.210***	.069
$\beta4$	.636***	.071	.662***	.094	.677***	.088	.699***	.090
$\alpha1$	1.000	---	1.000	---	1.000	---	1.000	---
$\alpha2$	1.600***	.182	1.449***	.186	1.441***	.185	1.384***	.197
$\alpha3$	1.514***	.165	1.319***	.155	1.304***	.161	1.256***	.161
$\alpha4$	1.200***	.119	1.054***	.123	1.038***	.124	.995***	.121
$S1$	.000	---	.000	---	.000	---	.000	---
$S2$	-2.174***	.586	-2.522***	.804	-2.517***	.686	-2.072***	.744
$S3$	.084	.253	.079	.302	-.002	.292	.061	.289
$S4$	1.000	---	1.000	---	1.000	---	1.000	---
μ_L	.375***	.109	.383***	.140	.405***	.132	.392***	.135
μ_S	.271***	.054	.286***	.072	.276***	.064	.336***	.089
σ_L^2	2.159***	.284	2.908***	.483	2.742***	.487	2.953***	.623
σ_S^2	.136***	.049	.144***	.040	.143**	.058	.144**	.061
ρ_{LS}	-.076	.180	-.137	.165	-.017	.191	-.021	.214
σ_{e1}^2	.856***	.210	1.005***	.243	1.007***	.258	1.077***	.307
σ_{e2}^2	.157	.206	.086	.197	.183	.287	.581	.387
σ_{e3}^2	.873***	.192	1.061***	.281	1.057***	.270	1.095***	.304
σ_{e4}^2	.071	.099	.181	.190	.170	.189	.391*	.224
Indices	*DIC*=3,329.41; Bayesian *p*-value=.552		*DIC*=3,370.06; Bayesian *p*-value=.488		*DIC*=3,347.52 ; Bayesian *p*-value=.513		*DIC*=3,338.53 ; Bayesian *p*-value=.494	

Table 8. Sensitivity analysis: The parameter estimates of IRM-LGC model for the restricted data

Priors distribution for item parameters: α ~ dnorm (0, 1.0E-02)I(0,) and β~ dnorm(0,1)						
Probit link						
uniform priors for varying residuals (~dunif (0, 1.0E04))						
	One single long chain (50,000 iterations, 19,998 burn-in)			Three independent chains (30,000 iterations, 20,000 burn-in)		
	Estimate (*EAP*)	*SD*	*mcse*^{2}	Estimate (*EAP*)	*SD*	*mcse*
$\beta 1$	0.000	---	---	.000	---	---
$\beta 2$	.182***	.067	0.003	.185***	.069	0.002
$\beta 3$	.205***	.068	0.003	.210***	.069	0.002
$\beta 4$	.679***	.084	0.004	.699***	.090	0.004
$\alpha 1$	1.000	---	---	1.000	---	---
$\alpha 2$	1.427***	.183	0.008	1.384***	.197	0.008
$\alpha 3$	1.307***	.167	0.008	1.256***	.161	0.006
$\alpha 4$	1.035***	.120	0.006	.995***	.121	0.005
$S1$	.000	---	---	.000	---	---
$S2$	-1.940***	.617	0.037	-2.072***	.744	0.038
$S3$	.104	.274	0.008	.061	.289	0.008
$S4$	1.000	---	---	1.000	---	---
μ_L	.370***	.128	0.004	.392***	.135	0.004
μ_S	.333***	.078	0.004	.336***	.089	0.004
σ_L^2	2.73***	.506	0.029	2.953***	.623	0.030
σ_S^2	.144**	.057	0.003	.144**	.061	0.003
ρ_{LS}	-.019	.204	0.010	-.021	.214	0.010
σ_{e1}^2	.996***	.265	0.012	1.077***	.307	0.013
σ_{e2}^2	.546	.348	0.020	.581	.387	0.019
σ_{e3}^2	1.016***	.275	0.013	1.095***	.304	0.012
σ_{e4}^2	.364*	.203	0.011	.391*	.224	0.010
Indices	*DIC*=3,340.25; Bayesian *p*-value=.504			*DIC*=3,338.53 ; Bayesian *p*-value=.494		

Note. 1. *p<.1 (1.6449), **p<.05 (1.96), ***p<.01 (2.5758); 2. *MCSE* stands for Monte Carlo standard error, which can always be reduced by lengthening the chain (Kim and Bolt, 2007).

4.3 Missing longitudinal data compensation

Missing data are unavoidable in almost all serious statistical analyses. Although the way in which the Bayesian estimation compensates for missing data is similar to the multiple imputation (MI) described by Rubin (1987), it extends the MI method by jointly simulating the distributions of variables with missing data as well as with unknown parameters (Carrigan et al., 2007; Patz and Junker, 1999b). Thus, through a fully Bayesian (FB) approach, not only can the missing values be treated as additional parameters to estimate, but these parameter estimates can themselves be marginally integrated from an exact joint posterior distribution for all the parameters of interest (Dunson et al., 2005). For instance, in the context of incomplete longitudinal data, the imputation and analysis models are fully and simultaneously specified in an FB analysis. However, the maximum likelihood method relies on a fully specified model, and its parameter estimates are constructed using likelihood-based approximations (Carrigan et al. 2007; Schafer and Graham, 2002).

In order to explore the influence of the item non-response on estimated parameters, two separate analyses were conducted: one with a reduced data set (for those individuals who have an opinion on every item in all four years), and the other with a full dataset of 323 respondents (Wiggins et al., 1990). As mentioned earlier, because WinBUGS treats missing values as additional parameters which need to be estimated, for those respondents with incomplete survey responses, handling missing data in this way helps to improve the reliability of inference for individual latent growth or decline trajectories (May, 2006; Patz and Junker, 1999b). Thus, in the present study, the paper by Wiggins and his colleagues (1990) serves as guidance in selecting explanatory variables, where age, gender, and religious status (treated as fixed at the respondent's 1983 response) were chosen to investigate their influences on the level and shape factors of a latent growth curve analysis.

According to Rubin (1987), there are three potential patterns of missingness: (1) missing completely at random (MCAR), (2) missing at random (MAR), and (3) missing not at random. Although the assumption of MCAR seems statistically retainable in the current study, we instead rely on the MAR assumption (see Table 9), indicating that a systematic difference can be explained by other observed variables (Rubin, 1987). The reason for this is that in longitudinal studies missing values are accumulated over time; in this sense they are easily susceptible to biased results. Therefore, an imputation component was built into the model using the three following auxiliary predictors, gender, age, and religious status, to deal with multivariate missing categorical data at each occasion. Based on the result shown in Table 10, both data sets provide estimates with identical substantial interpretation: there is evidence for an age and religious status interaction in terms of the true initial status. Young people without religious belief tend to have a higher tendency to hold positive attitudes toward abortion; however, the same is not the case for senior people with religious belief. As none of the Bayesian *p*-values is of extreme value, we find no failure of the model: suggesting that the model generates replicate data similar to the observed one.

5. CONCLUDING REMARKS

Obviously, a single-stage analytic strategy is an optimal alternative. In order to model the process of change, our intention is to propose an advanced analytic method which allows for the simultaneous estimation of a measurement model containing a set of categorical items and a latent growth curve analysis. As Bereiter (1963) puts it, one of the problems encountered in measuring change is scalability, in which the comparability of changes from different initial levels is questionable. However, it is expected that this comprehensive framework yields three benefits when the model fits the data well, and Bereiter's concern about scaling can be easily accommodated: (1) the interpretations of item parameters will be invariant to the latent trait distribution of the respondents in question; (2) the interpretations of latent trait parameters will be invariant to the distribution of the test items under consideration; and (3) precision can be approximately obtained in each model parameter estimate and latent variable (e.g., Curran et al., 2007; Dunson et al., 2005; Embretson, 1994; Rasch, 1960; Roberts and Ma, 2006). Finally, our expectation is that through the incorporation of multiple item psychometric models, an overall true score can be generated from this second-order latent growth curve model. That is, each item provides some sources of information, reduces our uncertainty about the examinees, and reflects respondents' positions on the underlying dimension in an adequately reliable way, which further improves the validity and reliability of the growth curve models embedded in the GLLAMM framework (e.g., Bollen, 1989; Curran et al., 2007; Fox, 2007; Preacher et al., 2008; Sayer and Cumsille, 2001; Wiggins et al., 1990).

In practice, however, many applications in educational and psychological testing involve long tests, large samples, response patterns, and high dimensional latent factor structures. As directions for future research, we could consider comparing and contrasting other estimation approaches to implementing the analysis, such as the Gauss-Hermite quadrature procedure with different options controlling the number of quadrature points used for each dimension of the integration (Skrondal and Rabe-Hesketh, 2004)[15], and releasing some strict assumptions, such as the stability of the item parameters over time and among different subpopulations. For instance, in addition to the indirect effects via the latent variable, we could investigate whether there are direct effects of the individual-level covariates on the responses. That is, we could examine whether the differential item functioning (DIF) exists, where the probability of endorsing an item differs among

---

[15] Te Marvelde and his colleagues (2006) argued that for more scales and time points, the adaptive Gauss-Hermite quadrature method may become infeasible, but this requires further investigation.

people with the same ability but distinct characteristics, such as people having the same propensity but being of different gender, and/or ethnicity (e.g., Holland and Wainer, 1993). In the education testing field, such investigation is important, for DIF suggests that participants cannot be fairly assessed by the instrument. Moreover, as stated by Lord (1980), because the latent ability obtained from IRMs are invariant across measures of same construct but with different psychometric properties, the generalizability of this unified model to designs with different item samples (i.e., where these items may be identical, partly overlapping or disjoint) administered on different occasions opens a promising avenue of future research (Curran et al., 2007; Fischer and Seliger, 1997; Patz and Yao, 2007a, 2007b; Roberts and Ma, 2006; Te Marvelde, Glas, Van Landeghem, and Van Damme, 2006).

Table 9. Unconditional models: The parameter estimates of IRM-LGC model for both data sets

	Three independent chains (30,000 iterations, 20,000 burn-in)					
	Complete cases (*n*=284)			Available cases (*n*=323)		
	Estimate (*EAP*)	*SD*	*mcse*^{2}	Estimate (*EAP*)	*SD*	*mcse*
$\beta1$	.000	---	---	.000	---	---
$\beta2$	.185***	.069	0.002	.189***	.066	0.002
$\beta3$	.210***	.069	0.002	.205***	.067	0.002
$\beta4$	.699***	.090	0.004	.724***	.082	0.003
$\alpha1$	1.000	---	---	1.000	---	---
$\alpha2$	1.384***	.197	0.008	1.382***	.171	0.007
$\alpha3$	1.256***	.161	0.006	1.291***	.156	0.006
$\alpha4$	.995***	.121	0.005	1.005***	.111	0.005
$S1$	.000	---	---	.000	---	---
$S2$	-2.072***	.744	0.038	-1.89***	.560	0.027
$S3$	.061	.289	0.008	.110	.261	0.007
$S4$	1.000	---	---	1.000	---	---
μ_L	.392***	.135	0.004	.302**	.122	0.003
μ_S	.336***	.089	0.004	.353***	.076	0.003
σ_L^2	2.953***	.623	0.030	2.957***	.505	0.023
σ_S^2	.144**	.061	0.003	.148**	.059	0.003
ρ_{LS}	-.021	.214	0.010	.029	.202	0.009
σ_{e1}^2	1.077***	.307	0.013	1.019***	.269	0.011
σ_{e2}^2	.581	.387	0.019	.536	.330	0.016
σ_{e3}^2	1.095***	.304	0.012	1.023***	.271	0.010
σ_{e4}^2	.391*	.224	0.010	.324*	.178	0.008
Indices	*DIC*=3,338.53 ; Bayesian *p*-value=.494			*DIC*=3,641.82; Bayesian *p*-value=.500		

Note. 1. **p*<.1 (1.6449), ***p*<.05 (1.96), ****p*<.01 (2.5758); 2. *MCSE*, a type of sampling error, stands for Monte Carlo standard error, which can always be reduced by lengthening the chain (Kim and Bolt, 2007).

Table 10. Conditional models: The parameter estimates of IRM-LGC model for both data sets

Parameter	Restricted data (*n*=284)			Full data (*n*=323)		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
			Measurement part			
β1	.000 (fixed)	.000 (fixed)	.000 (fixed)	.000 (fixed)	.000 (fixed)	.000 (fixed)
β2	.182** (.071)	.180*** (.068)	.173** (.068)	.196*** (.069)	.191*** (.067)	.193*** (.066)
β3	.209*** (.074)	.205*** (.068)	.197*** (.070)	.214*** (.070)	.207*** (.068)	.209*** (.068)
β4	.734*** (.092)	.688*** (.086)	.675*** (.089)	.779*** (.092)	.730*** (.089)	.737*** (.089)
α1	1.000 (fixed)	1.000 (fixed)	1.000 (fixed)	1.000 (fixed)	1.000 (fixed)	1.000 (fixed)
α2	1.309*** (.174)	1.403*** (.181)	1.417*** (.188)	1.282*** (.154)	1.363*** (.168)	1.354*** (.159)
α3	1.173*** (.146)	1.271*** (.149)	1.285*** (.157)	1.2*** (.144)	1.298*** (.158)	1.278*** (.158)
α4	.918*** (.109)	1.005*** (.113)	1.017*** (.120)	.916*** (.098)	1.0*** (.116)	.985*** (.107)
			Structural part			
S1	.000 (fixed)	.000 (fixed)	.000 (fixed)	.000 (fixed)	.000 (fixed)	.000 (fixed)
S2	-1.008*** (.406)	-1.555*** (.594)	-1.915*** (.618)	-1.077*** (.387)	-1.495*** (.491)	-1.827*** (.573)
S3	.173 (.202)	.102 (.256)	.086 (.278)	.182 (.197)	.155 (.233)	.114 (.256)
S4	1.000 (fixed)	1.000 (fixed)	1.000 (fixed)	1.000 (fixed)	1.000 (fixed)	1.000 (fixed)
β1.int	-.366 (.243)	-.197 (.188)	-.180 (.180)	-.378 (.232)	-.252 (.181)	-.231 (.182)
β1.gender	.219 (.384)			.147 (.382)		
β1.age	.606 (.370)	.550** (.273)	.555** (.264)	.520 (.355)	.475* (.259)	.481* (.263)
β1.relig	2.468*** (.675)	1.62*** (.382)	1.613*** (.374)	1.882*** (.606)	1.469*** (.367)	1.507*** (.367)
β1.gen.age	-.0167 (.609)			-.036 (.583)		
β1.gen.rel	-1.12 (.827)			-.488 (.773)		
β1.age.rel	-2.122*** (.797)	-1.252*** (.481)	-1.253*** (.473)	-1.38* (.727)	-.990** (.453)	-1.026** (.463)
β1.gen.age.rel	1.169 (1.063)			.485 (.981)		
β2.int	.388*** (.143)	.314*** (.091)	.344*** (.083)	.387*** (.133)	.336*** (.081)	.369*** (.083)
β2.gender	.514* (.267)	.181 (.126)		.394* (.233)	.159 (.113)	
β2.age	.073 (.219)			.0818 (.194)		
β2.relig	-.047 (.446)			.222 (.391)		
β2.gen.age	-.346 (.377)			-.230 (.315)		
β2.gen.rel	-.171 (.544)			-.322 (.469)		
β2.age.rel	-.174 (.508)			-.443 (.470)		
β2.gen.age.rel	.194 (.686)			.383 (.585)		

σ_L^2	3.012*** (.603)	2.599*** (.474)	2.579*** (.496)	3.16***(.577)	2.766*** (.552)	2.821*** (.500)
σ_S^2	.245** (.122)	.169** (.079)	.144** (.060)	.231** (.109)	.176** (.080)	.147** (.059)
ρ_{LS}	0.173 (.27)	0.107 (.232)	.037 (.213)	.209 (.236)	.126 (.228)	.067 (.210)
σ_{e1}^2	1.167*** (.327)	0.996*** (.276)	.977*** (.266)	1.159*** (.306)	.999*** (.274)	1.014*** (.258)
σ_{e2}^2	1.198*** (.442)	.749* (.404)	.614* (.370)	1.002** (.419)	.645* (.356)	.607* (.337)
σ_{e3}^2	1.158*** (.318)	1.047*** (.293)	1.024*** (.290)	1.155*** (.304)	1.035*** (.287)	1.056*** (.276)
σ_{e4}^2	.419 (.258)	.407* (.225)	.4101* (.227)	.409 (.261)	.359* (.203)	.388* (.233)
Goodness of fit indices	*DIC*=3,336.880; Bayesian *p*=.478	*DIC*=3,340.01; Bayesian *p*=.489	*DIC*=3,341.7 ; Bayesian *p*=.488	*DIC*=3,638.18; Bayesian *p*=.48	*DIC*=3,638.90 ; Bayesian *p*=.495	*DIC*=3,639.29 ; Bayesian *p*=.494

Note. 1. Each number inside the parenthes is stands for the standard deviation of the estimate.
2. *p<.1 (1.6449), **p<.05 (1.96), ***p<.01 (2.5758)

Taken together, the application of IRMs to responses gathered from repeated assessments allows us to take into consideration both the characteristics of item responses and measurement error in the analysis of individual developmental trajectories. We could combine and incorporate other item response models and latent growth models in this comprehensive modeling framework, such as unidimensional/multidimensional and dichotomous/polytomous IRMs, linear/nonlinear, and single/multiple domain(s) LGCs etc. As a simplified demonstration, in the present study we consider the modeling of only a unidimensional latent construct. However, in developmental research one is often interested in the way in which two or more repeatedly followed and interrelated dimensions evolve over time. In order to effectively accommodate a variety of data structures, it is clearly worthwhile to extend to multiple domains through the analysis of random effect regressions[16] (e.g., Cheong, MacKinnon, and Khoo, 2003; Preacher et al., 2008; Raykov, 2007), and simultaneously make use of their interrelationship when we have multiple interrelated dimensions across the entire study period.

REFERENCES

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, 17, 251-269.

Albert, J. H., & Chib, S. (1997). Bayesian analysis of binary and polytomous response data. *Journal of the American Statistical Association*, 88, 669-679.

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature*, 19, 1483-1536.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, 28 (2), 135-167.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 203-212). Madison, University of Wisconsin Press.

Best, N. G., Spiegelhalter, D. J., Thomas, A., & Brayne, C. E. G. (1996). Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society,* Series A, 159 (2), 323-342.

Birnbaum, A. (1968). Test scores, sufficient statistics, and the information structures of tests. In F. M. Lord and M. R. Novick (Ed.). *Statistical theories of mental test scores* (pp. 425-435). Reading, MA: Addison-Wesley Publishing Company.

Blozis, S. (2007). A second order structural latent curve model for longitudinal data. In K. van Montfort, J. Oud, and A. Satorra (Eds.). *Longitudinal Models in the Behavioral and Related Sciences* (pp. 189-214). NJ: Lawrence Erlbaum Associates, Inc.

Bollen, K. A. (1989). *Structural equations with latent variables*. NY: Wiley.

Byrne, B. M., & Crombie, G. (2003). Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics*, 2 (3), 177-203.

Carrigan, G., Barnett, A. G., Dobson, A. J., & Mishra, G. (2007). Compensating for missing data from longitudinal studies using WinBUGS. *Journal of Statistical Software*, 19 (7), 1-17.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46 (3), 164-174.

---

[16] This analysis is one kind of structural equation model with regressions among the latent variables, representing aspects of distinct individual growth curve analyses, each of which is modeled along a particular dimension.

Cheong, J. W., MacKinnon, D. A., & Khoo, S. T. (2003). Investigation of mediational processes using parallel process growth curve modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 10, 238-262.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician,* 49 (4), 327-335.

Chou, C.-P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 5, 247-266.

Congdon, P. (2005). Markov chain Monte Carlo and Bayesian statistics. In B. S. Everitt and D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1134-43). London: John Wiley & Sons.

Curran, P. J. (2003). Have multilevel models been structural equation models all along. *Multivariate Behavioral Research*, 38 (4), 529-569.

Curran, P. J., Edwards, M. C., Wirth, R. J., Hussong, A. M., & Chassin, L. (2007). The incorporation of categorical measurement models in the analysis of individual growth. In T. Little, J. Bovaird, & N. Card (Eds.), *Modeling ecological and contextual effects in longitudinal studies of human development* (pp. 89-120). NJ: Lawrence Erlbaum Associates, Inc.

Diggle, P., Heagerty, P., Liang, K.-Y., & Zeger, S. (2002). *Analysis of longitudinal data* (2nd Ed.). Oxford, England: Oxford University Press.

Dunson, D. B. (2007). Bayesian methods for latent trait modeling of longitudinal data. *Statistical Methods in Medical Research*, 16 (5), 399-415.

Dunson, D. B., Palomo, J., & Bollen, K. (2005). Bayesian structural equation modeling. Technical report # 2005-5. Statistical and Applied Mathematical Sciences Institute. Retrieved 04 February, 2007, from http://www.samsi.info/TR/tr2005-05.pdf.

Embretson, S. E. (1994). Comparing changes between groups: some perplexities arising from psychometrics. In Laveault, D., Zumbo, B. D., Gessaroli, M. E., & Boss, M. W. (Eds.) *Modern theories of measurement: Problems and issues* (pp. 213-248). Ottawa, Canada: Edumetrics Research Group, University of Ottawa.

Engel, U. Gattig, A., & Simonson, J. (2007). Longitudinal multilevel modeling: a comparison of growth curve models and structural equation modeling using panel data from Germany. In K. van Montfort, J. Oud, and A. Satorra (Eds.). *Longitudinal Models in the Behavioral and Related Sciences* (pp. 295-314). NJ: Lawrence Erlbaum Associates, Inc.

Everitt, B. S. (2005). Longitudinal data analysis. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1098-1101), London: John Wiley & Sons.

Fienberg, S. E., & Rinaldo, A. (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*, 137 (11), 3430-3445.

Fischer, G. H., Seliger, E. (1997). Multidimensional linear logisitc models for change. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 323-346), NY: Springer.

Fox, J-P. (2007). Multilevel IRT modeling in practice with the package mlirt. *Journal of Statistical Software*, 20 (5), 1-16.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., & Smith, A. F. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association,* 85 (412), 972-985.

Gelman, A., Carlin, J. B., Stern, H.S., & Rubin, D. B. (2003). *Bayesian data analysis*. London (2nd Ed.): Chapman & Hall.

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7,* 457-472.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science 7*, 473-483.

Gibbons, R. D., & Hedeker, D. (1997). Random-effects probit and logistic regression models for three-level data. *Biometrics*, 73, 43-56.

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. FL:

Chapman & Hall.

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Type of change generated by OD designs. *Journal of Applied Behavioral Science*, 12, 133-157.

Hedeker, D. (2005). Generalized linear mixed models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 729-738), London: John Wiley & Sons.

Holland, P. W., & Wainer, H. (1993). Differential item functioning (1st Ed). NJ: Lawrence Erlbaum Associates, Inc.

Hox, J., & Stoel, R. D. (2005). Multilevel and SEM approaches to growth curve modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1296-1305), London: John Wiley & Sons.

Hsieh, C., & Maier, K.S. (2009). A preliminary Bayesian analysis of incomplete longitudinal data from a small sample: Methodological advances in an international comparative study of educational inequality, International Journal of Research and Method in Education, 32(1), 103-125.

Jackman, S. (2000). Estimation and Inference via Bayesian simulation: An introduction to Markov chain Monte Carlo. *American Journal of Political Science*, 44 (2), 375-404.

Johnson, M. S., Sinharay, S., & Bradlow, E. T. (2007). Hierarchical item response theory models. In C. R. Rao & S. Sinharay (Eds.). *Handbook of Statistics, 26: Psychometrics* (pp. 587-606). Boston: Elsevier North-Holland.

Jöreskog, K. G. (2002). Structural equation modeling with ordinal variables using LISREL. Retrieved 25 December, 2008, from http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf.

Kim, J. S., & Bolt, D.M. (2007). Markov chain Monte Carlo estimation of item response models. *Educational Measurement: Issues and Practice*, 26, 38-51.

Knott, M., Albanese M. T., & Galbraith, J. (1990). Scoring attitudes to abortion. *The Statistician*, 40, 217-223.

Lee, M. D., & Wagenmakers, E. (2005). Bayesian statistical inference in psychology: Comment on Trafimow (2003). *Psychological review*, 112, 662-668.

Li, Y., Bolt, D.M., & Fu, J. (2006). A comparison of alternative models for testlets. *Applied Psychological Measurement*, 30 (1), 3-21.

Liao, T. F. (1994). *Interpreting probability models: logit, probit, and other generalized linear models*. Sage publications, Inc.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association,* 83 (404*),* 1198-1202.

Lord, F. M. (1980). *Application of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Welsley Publishing Company.

Lynch, S. M., & Western, B. (2004). Bayesian posterior predictive checks for complex models. *Sociological Methods and Research*, 32 (3), 301-335.

Maier, K. S. (2001). A Rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics,* 26 (3), 307-330.

May, H. (2006). A multilevel Bayesian item response theory method for scaling socioeconomic status in international studies of education. *Journal of Educational and Behavioral Statistics*, 31 (1), 63-79.

McGrath, K., & Waterton, J. (1986). *British Social Attitudes, 1983-1986, Panel Survey: Technical Report* (London, Social and Community Planning Research).

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.). *New methods for the analysis of change* (pp. 203-240). Washington, DC.: American Psychological Association.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical care*, 44 (11) Suppl 3, S69-S77.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107-122.

Muthén, B. O. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, 22, 43-65.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115-132.

Muthén, B. O. (1996). Growth modeling with binary responses. In A. von Eye & C. Clogg (Eds.), *Categorical variables in developmental research: Methods of analysis* (pp. 37-54). San Diego: Academic Press.

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

Muthén, L. K., & Muthén, B.O. (2006). *Mplus user's guide*. Los Angeles: Muthén & Muthén. Retrieved 28 June, 2006, from http://www.statmodel.com.

Patz, R.J., & Junker, B. W. (1999a). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, 24 (2), 146-178.

Patz, R.J., & Junker, B. W. (1999b). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics*, 24 (4), 342-366.

Patz, R.J., & Yao, L. (2007a). Vertical scaling: Statistical models for measuring growth and achievement. In C. R. Rao and S. Sinharay (Eds.), *Handbook of statistics, Vol. 26, Psychometrics* (pp. 955-975). Amsterdam: Elsevier.

Patz, R.J., & Yao, L. (2007b). Methods and models for vertical scaling. In N. J. Doran, M. Pommerich, and P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 253-272), New York: Springer.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. LA: Sage Publications, Inc

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata* (2^{nd} Ed.). TX: Stata press publication.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

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

Raudenbush, S. W., Johnson, C., & Sampson, R. J. (2003). A multivariate multilevel Rasch model with application to self-reported criminal behavior. *Sociological Methodology*, 33, 169-212.

Raykov, T. (2007). Longitudinal analysis with regressions among random effects: A latent variable modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 14 (1), 146-169.

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rice, J. A. (1995). *Mathematical statistics and data analysis*. CA: Duxbury Press.

Roberts, J. S., & Ma, Q. (2006). IRT models for the assessment of change across repeated measurements. In R. Lissitz (Ed.), *Longitudinal and value added modeling of student performance* (pp. 100-127). Maple Grove, MN: JAM Press.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Rupp, A. A., Dey, D. K., & Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 11 (3), 424-451.

Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.). *New methods for the analysis of change* (pp. 179-200). Washington, DC: American Psychological Association.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7 (2), 147-177.

Seltzer, M. H., Wong, W. H., & Bryk, A. S. (1996). Bayesian analysis in applications of hierarchical models: Issues and methods. *Journal of Educational and Behavioral Statistics*, 21 (2), 131-167.

Singer, J. D. (1998). Using SAS PROC MIXED to fix multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24 (4), 323-355.

Singer, J. D., & Willett, J. B. (2005). Growth curve modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 772-779), London: John Wiley & Sons.

Sinharay, S., Johnson, M. S., & Stern, H. S. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurements*, 30 (4), 298-321.

Sinharay, S., & Stern, H. S. (2003). Posterior predictive model checking in hierarchical models. *Statistical Planning and Inference*, 111, 209-221.

Skrondal, A., & Rabe-Hesketh, S. (2003). Some applications of generalized linear latent and mixed models in epidemiology: Repeated measures, measurement error and multilevel modeling. *Norsk Epidemiologi*, 13 (2), 265-278.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. Boca Raton: Chapman & Hall/CRC.

Skrondal, A., & Rabe-Hesketh, S. (2008). Multilevel and related models for longitudinal data. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 275-299), New York: Springer.

Social and Community Planning Research, *British Social Attitudes Survey Panel Study, 1983-1986* [computer file]. Colchester, Essex: UK. Data Archive [distributor], 15 July 1987. SN: 2197. Retrieved October 5, 2008, from https://www.esds.ac.uk/newRegistration/newLogin.asp.

Spiegelhalter, D. J., Best, N.G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society*, Series B, 64 (4), 583-616.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2003). *WinBUGS user manual*.

Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health. Retrieved 24 October, 2006, from http://www.mrc-bsu.cam.ac.uk/bugs.

Steele, F., & Goldstein, H. (2007). Multilevel models in psychometrics. In C. R. Rao & S. Sinharay (Eds.). *Handbook of Statistics, 26: Psychometrics* (pp. 401-420). Boston: Elsevier North-Holland.

Stefanescu, C., Berger, V. W., & Hershberger, S. L. (2005). Probits. In B. S. Everitt & D. C. Howell (Eds.). *Encyclopedia of Statistics in Behavioral Science* (pp. 1608-1610), vol. 3. London: John Wiley & Sons.

Te Marvelde, J. M., Glas, C. A. W., Van Landeghem, G., & Van Damme J. (2006). Application of multidimensional item response theory models to longitudinal data. *Educational and Psychological Measurement*, 66 (1), 5-34.

Thompson, J., Palmer, T., & Moreno, S. (2006). Bayesian analysis in Stata with WinBUGS. *The Stata Journal*, 6(4), 530-549.

Vermunt, J. (2007). Growth models for categorical response variables: Standard, latent-class, and hybrid approaches. In K. van Montfort, J. Oud, and A. Satorra (Eds.). *Longitudinal Models in the Behavioral and Related Sciences* (pp. 139-158). NJ: Lawrence Erlbaum Associates, Inc.

Western, B. (1999). Bayesian analysis for sociologists: An introduction. *Sociological Methods and Research*, 28 (1), 7-34.

Wiggins, R. D., Ashworth, K., & O'Muircheartaigh, C. A. (1990). Multilevel analysis of attitudes to abortion. *The Statistician*, 40, 225-234.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, 116 (2), 363-381.

Zhang, Z., Hamagami, F., Wang, L., Grimm, K. J., & Nesselroade, J. R. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, 31 (4), 374-383.

Zheng, X., & Rabe-Hesketh, S. (2007). Estimating parameters of dichotomous and ordinal item response models with gllamm. *The Stata Journal*, 7 (3), 313-333.

APPENDIX:

```
# WinBUGS codes for IRM-LGC model with varying residuals (conditional)


#model specification:


model
{
            for (j in 1:n.ind)
            {
                                    for (i in 1:n.item)
                                    {


                                     p1[j,i]<-max(.0001, p1.temp[j,i])
                                     p2[j,i]<-max(.0001, p2.temp[j,i])
                                     p3[j,i]<-max(.0001, p3.temp[j,i])
                                     p4[j,i]<-max(.0001, p4.temp[j,i])
                                     p1.temp[j, i]<-phi(alpha[i]*(theta[j, 1]-beta[i]))
                                    p2.temp[j, i]<-phi(alpha[i]*(theta[j, 2]-beta[i]))
                                    p3.temp[j, i]<-phi(alpha[i]*(theta[j, 3]-beta[i]))
                                    p4.temp[j, i]<-phi(alpha[i]*(theta[j, 4]-beta[i])) #measurement model: the normal ogive

                                    rp1[j,i]<-max(.0001, rp1.temp[j,i])
                                    rp2[j,i]<-max(.0001, rp2.temp[j,i])
                                    rp3[j,i]<-max(.0001, rp3.temp[j,i])
                                    rp4[j,i]<-max(.0001, rp4.temp[j,i])
                                    rp1.temp[j, i]<-phi(alpha[i]*(theta[j, 1]-beta[i]))
                                    rp2.temp[j, i]<-phi(alpha[i]*(theta[j, 2]-beta[i]))
                                    rp3.temp[j, i]<-phi(alpha[i]*(theta[j, 3]-beta[i]))
                                    rp4.temp[j, i]<-phi(alpha[i]*(theta[j, 4]-beta[i]))

                                    ry1[j, i]~dbern(rp1[j, i])#generate the replicated data
                                    ry2[j, i]~dbern(rp2[j, i])
                                    ry3[j, i]~dbern(rp3[j, i])
                                    ry4[j, i]~dbern(rp4[j, i])
                                    chi.y1.rep[j,i]<-pow(ry1[j,i]-p1[j,i], 2)/p1[j,i]*(1-p1[j,i])
                                    chi.y2.rep[j,i]<-pow(ry2[j,i]-p2[j,i], 2)/p2[j,i]*(1-p2[j,i])
                                    chi.y3.rep[j,i]<-pow(ry3[j,i]-p3[j,i], 2)/p3[j,i]*(1-p3[j,i])
                                    chi.y4.rep[j,i]<-pow(ry4[j,i]-p4[j,i], 2)/p4[j,i]*(1-p4[j,i])

                                    y1[j, i]~dbern(p1[j, i])
                                    y2[j, i]~dbern(p2[j, i])
                                   y3[j, i]~dbern(p3[j, i])
                                    y4[j, i]~dbern(p4[j, i])
                                    chi.y1.obv[j,i]<-pow(y1[j,i]-p1[j,i], 2)/p1[j,i]*(1-p1[j,i])
                                    chi.y2.obv[j,i]<-pow(y2[j,i]-p2[j,i], 2)/p2[j,i]*(1-p2[j,i])
                                    chi.y3.obv[j,i]<-pow(y3[j,i]-p3[j,i], 2)/p3[j,i]*(1-p3[j,i])
                                    chi.y4.obv[j,i]<-pow(y4[j,i]-p4[j,i], 2)/p4[j,i]*(1-p4[j,i])
                                    }
             }

chi.y.obv<-sum(chi.y1.obv[,])+sum(chi.y2.obv[,])+sum(chi.y3.obv[,])+sum(chi.y4.obv[,])
chi.y.rep<-sum(chi.y1.rep[,])+sum(chi.y2.rep[,])+sum(chi.y3.rep[,])+sum(chi.y4.rep[,])
chi<-step(chi.y.rep-chi.y.obv) # calculate the posterior predictive p-value
```

```
#structural model: the latent growth curve analysis

        for (j in 1:n.ind)
        {
               LS[j,1:2]~dmnorm(Mu[j,1:2], Inv_cov[1:2,1:2])
               Mu[j,1]<-b1.int[1]+(b1.age[1]*age[j])+(b1.relig[1]*relig[j])+(b1.age.rel[1]*age[j]*relig[j])
               Mu[j,2]<-b2.int[1]


                       for (t in 1:n.occ)
                       {
                       theta[j, t]~dnorm(MuY[j,t], Inv_sig_e[t])
                       MuY[j, t]<-LS[j,1]+LS[j,2]*A[t]
                       }

        }

#Prior distribution

              beta[1]<-0#model identification
              alpha[1]<-1#model identification
              for (i in 2:n.item)
              {
              beta[i]~dnorm(0,1)
              alpha[i] ~ dnorm(0, 1.0E-2)I(0,)
              }

              for (t in 1:n.occ)
               {
               Inv_sig_e[t]<-pow(Sig_e[t], -1)
               Sig_e[t]~dunif(0, 1.0E04)
              }

               b1.int[1]~dnorm(0, 1.0E-4)
               b1.age[1]~dnorm(0, 1.0E-04)
               b1.relig[1]~dnorm(0, 1.0E-04)
               b1.age.rel[1]~dnorm(0, 1.0E-04)
               b2.int[1]~dnorm(0, 1.0E-4)

              A[1]<-0
             for (t in 2:3)
             {
              A[t]~dnorm(0,1.0E-4)
              }
              A[4]<-1

               Inv_cov[1:2,1:2]~dwish(R[1:2,1:2], 3)
               R[1,1]<-1
               R[2,2]<-1
               R[2,1]<-R[1,2]
               R[1,2]<-0

            #Transform the parameters
             Cov[1:2,1:2]<-inverse(Inv_cov[1:2,1:2])
             Sig_L<-Cov[1,1]
            Sig_S<-Cov[2,2]
              rho<-Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
```

```
            # all parameter are put into para(s)
                 for (i in 1:3)
                 {
                   Para[i]<-beta[i+1]
                 }
                 for (i in 2:4)
                  {
                  Para[2+i]<-alpha[i]
                  }


             for (t in 1:4)
             {
                 Para[6+t]<-Sig_e[t]
                 }
                 Para[11]<-Sig_L
                Para[12]<-Sig_S
                Para[13]<-rho
                Para[14]<-b1.int[1]
                 Para[15]<-b1.age[1]
                 Para[16]<-b1.relig[1]
                 Para[17]<-b1.age.rel[1]
                 Para[18]<-b2.int[1]
                 Para[19]<-A[2]
                 Para[20]<-A[3]
                 Para[21]<-chi


 }

# End of model
```