This is to certify that the
dissertation entitled

A UNIFIED MODEL FOR THE ANALYSIS OF
INDIVIDUAL LATNET TRAJECTORIES

presented by

CHUEH-AN HSIEH

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Measurement and Quantitative Methods

Kimberly S. Maier
Major Professor's Signature

5 April 2010
Date

A UNIFIED MODEL FOR THE ANALYSIS OF INDIVIDUAL LATENT TRAJECTORIES

By

Chueh-An Hsieh

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILSOPHY

Measurement and Quantitative Methods

2010

ABSTRACT

A UNIFIED MODEL FOR THE ANALYSIS OF INDIVIDUAL LATENT TRAJECTORIES

By

Chueh-An Hsieh

The application of item response theory models to repeated observations has demonstrated great promise in developmental research. It allows researchers to take into consideration the characteristics of both item response and measurement error in longitudinal trajectory analysis, which improves the reliability and validity of the latent growth curve (LGC) model. This thesis demonstrates the potential of Bayesian methods and proposes a comprehensive modeling framework, combining a measurement model with a structural model. That is, through the incorporation of a commonly used link function and Bayesian estimation, an item response theory model (IRT) can be naturally introduced into a latent variable model (LVM).

All proposed analyses are implemented in WinBUGS 1.4.3 (Spiegelhalter, Thomas, Best, and Lunn, 2003), which allows researchers to use Markov chain Monte Carlo (MCMC) simulation methods to fit complex statistical models and circumvent intractable analytic or numerical integrations. The utility of this IRT-LVM modeling framework was investigated with both simulated and empirical data, and promising results were obtained. As the results indicate, the IRT-LVM utilized information from individual items of the scales at each point in time, allowing the employment of item response characteristics from distinct psychometric models, permitting the separation of

time-specific error and measurement error, and giving researchers a way to evaluate the factorial invariance of latent constructs across different assessment occasions.

ACKNOWLEDGEMENTS

After a few years of work and fun, this is it: my thesis, about models for individual latent trajectories, on which I worked at the Measurement and Quantitative Methods (MQM) program of Michigan State University. I would like to thank everyone at MQM for the good times I had studying there, and for a grant received from the American Educational Research Association (AERA) Grants Program under the National Science Foundation (NSF) Grant #DRL-0941014. There are some people I would like to thank in particular:

Dr. Matthew A. Diemer, for being a good model when I worked with him as a research assistant; from him I can see the importance of maintaining a diligent and persistent attitude toward research; Dr. Richard T. Hoang, for his friendly and unique interpretation of educational statistics, which opens a real world in front of me; Dr. Kimberly S. Maier, for the freedom and patience she always grants me, which have become incredible assets that will greatly enhance my future; Dr. Mark D. Reckase, for his great knowledge base in measurement, which intrigues me, as a way to see things differently; Dr. Alexander A. von Eye, for his fruitful advice and pleasant cooperation, which has continuously helped me to consider how to conduct rigorous research. Also, to my friends in Japan, Taiwan, the United States, and the United Kingdom–without the invaluable friendship of you all, this learning journey toward a Ph.D. could not have been as smooth as it was.

Finally, I would like to dedicate this thesis to my family: through your unconditional support and love, I learned how to push the boundary a bit further!

TABLE OF CONTENTS

CHAPTER 5

APPENDICES

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Longitudinal Data Analysis

The use of growth models in social, behavioral, and educational research has increased rapidly, because it answers important research questions such as concern the nature of psychological and 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; Hsieh and Maier, 2009; 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 among the observed variables, providing estimates of the individual growth parameters and inter-individual differences in change across all members of the population; hence its name, latent growth curve (LGC) analysis.

It is this mean and covariance structure (MCS) 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. Within the HLM framework, time is an independent variable at the lowest level and the individual is defined at the second level, in which time-varying and time-invariant explanatory variables can be incorporated into existing level-1 and level-2 models. Additionally, the intercept and slope describe the mean change status and the change rate, and inter-individual differences in the change profile can be modeled as random effects for either the intercept or the slope of the time variable, or both (Raudenbush and Bryk, 2002). Likewise, within the LGC, the time variable is incorporated as a series of constrained values for the factor loadings of the latent variable representing the shape of the growth curve, along with all the factor loadings of the latent variable constrained to the value of one and representing the initial level. Thus, the latent variable means for the initial level and shape factors depict the mean growth status and the growth rate, and inter-individual differences in the change can be modeled as the covariance of the level and shape factors (Meredith and Tisak, 1990).

While several key differences remain between these two models, at the time of writing this dissertation, the discrepancies have rapidly been disappearing (Curran, Obeidat, Losardo, in press; Preacher, Wichman, MacCallum and Briggs, 2008; Raykov, 2007). One primary difference is that in the HLM, time is treated as a fixed explanatory variable, whereas time is introduced in the LGC model via the factor loadings, which makes HLM the best approach if there are a great many varying occasions across individuals (Snijders 1996; Willett and Sayer, 1994), and the LGC is considered best suited for time-structured data or a fixed occasion design (e.g., Byrne and Crombie, 2003; Skrondal & Rabe-Hesketh, 2008). 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; Preacher et al., 2008; Raudenbush and Bryk, 2002; Willett and Sayer, 1994).

Research Motivation

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; Johnson and Raudenbush, 2006), 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) indicate, 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.

Moreover, 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 (Moustaki and Knott, 2000; Vermunt, 2007). In addition, the calculation of standard errors is challenging when the expectation-maximization (EM) algorithm is used to compute the maximum likelihood estimates (Jamshidian and Jennrich, 2000). Thus, in order to accommodate these, we bridge the gap by resorting to an integrative modeling framework: using the derivative of the generalized linear latent and mixed model[1] (GLLAMM; Skrondal and Rabe-Hesketh, 2004), strengthened by the attributes of the item response theory model (IRT) (e.g., Lord and Novick, 1968), the latent variable model (LVM) (e.g., Muthén, 2002), and the Bayesian estimation approach. An overall "true score" can be generated from a second-order latent growth curve analysis, in which each item provides some sources of information, reduces our uncertainty about the examinees, and reflects respondents' positions on the underlying dimension (e.g., Bollen, 1989; Curran et al., 2007; Fox, 2007; Preacher et al., 2008; Sayer and Cumsille, 2001; Wiggins, Ashworth, O'Muircheartaigh, and Galbraith, 1990).

---

[1] Analogous to the different treatment of the time variable in the HLM and LGC, time is treated as a fixed explanatory variable in the growth model embedded in the GLLAMM, but is introduced via the factor loadings in the present study.

Objectives of the Present Work

The application of item response theory models to repeated observations has demonstrated great promise in developmental research. It allows one to take into consideration the characteristics of both item response and measurement error in longitudinal trajectory analysis, which improves the reliability and validity of the latent growth curve model (e.g., Bollen, 1989; Curran et al., 2007; Fox, 2007; Hsieh and von Eye, in press; Preacher et al., 2008; Sayer and Cumsille, 2001; Wiggins et al., 1990). Within this modeling framework, different types of item response model and latent growth curve analysis can be combined to address various research questions. In addition, different data structures can be accommodated, such as unidimensional vs. multidimensional item response theory models, dichotomous vs. polytomous items, linear vs. nonlinear change trajectories, single vs. multiple domain(s) latent growth curve analyses, etc. In longitudinal studies, although the development of a single behavior is often of interest, it is worthwhile to extend this to multiple domains and simultaneously model their interrelationship across the entire study period (e.g., Cheong, MacKinnon, and Khoo, 2003; Preacher et al., 2008; Raykov, 2007).

In the present study, the hierarchical nature of latent variable problems suggests a Bayesian approach to estimation. In estimating complex statistical models, the capacity of Bayesian methods is undeniable. Bayesian data analysis is seen as having a range of advantages, such as an intuitive probabilistic interpretation of the parameters of interest, the efficient incorporation of prior information to empirical data analysis, the ability to take account of model uncertainty among different models and to draw combined inferences when there is no single pre-eminent model, and so on (Best, Spiegelhalter, Thomas, and Brayne, 1996; Maier, 2001; Rupp, Dey, and Zumbo, 2004; Western, 1999). Additionally, unlike the maximum likelihood estimation (MLE), which requires large samples to approximate the sampling distribution for sample statistics, Bayesian inference can be seen a plausible way to deal with small sample size studies (Congdon, 2005; Lee and Wagenmakers, 2005; Zhang, Hamagami, Wang, Grimm, and Nesselroade, 2007). Beyond its value for this purpose, the Bayesian method also has a unique strength, the systematic incorporation of prior information from previous studies (Scheines, Hoijtink and Boomsma, 1999; Rupp et al., 2004; Zhang et al., 2007). Bayesian methods and Bayes' theorem permit the incorporation of previous findings as supplementary and influential information, whereas traditional likelihood methods cannot do this (Western and Jackman, 1994). By not undertaking statistical analysis in isolation, Bayesian learning draws on existing knowledge in the prior framing of the model and allows the combination of existing evidence with the actual study data at hand during the estimation process (Congdon, 2005). Besides, the interval estimation is a direct product via a Bayesian estimation routine: inference on functions of parameters can easily be obtained, since the full posterior distribution of the parameters is available.

Thus, in order to differentially weigh individual items and examine developmental stability and change over time, this thesis seeks to demonstrate the potential of Bayesian methods and propose a comprehensive modeling framework combining both a measurement model and a structural model. That is, through the incorporation of a commonly used link function and Bayesian estimation, the item response theory (IRT) model can be naturally introduced into the latent variable model (LVM). Despite a large number of components requiring attention, this thesis restricts its focus to the following issues: (1) model formulation: how Bayesians explicitly incorporate (multivariate) multiple repeated measures of discrete scale into a latent growth curve model, in which the unidimensional Rasch (1960) and linear latent growth curve model (RASCH-LLGC), the unidimensional two-parameter normal ogive (e.g., Birnbaum, 1968) and nonlinear latent growth curve model (e.g., Meredith and Tisak, 1990) (2PNO-LGC), and the multidimensional graded response (e.g., De Ayala, 1994) and associative latent growth curve model (e.g., McArdle, 1988) (MGRM-ALGC) are presented; (2) the evaluation of the model parameter estimate performance: as the sample size needed for a particular longitudinal study depends on many factors, an "adequate" sample size is hard to determine unambiguously. As a simplified illustration, we demonstrate how to evaluate the performance of parameter estimates through conducting a Monte Carlo study. For instance, to evaluate the numerical behavior of the average growth trajectory in Bayesian analysis, we launch a small-scale simulation study using a 2×3×2 design with 12 conditions. Given the constant number of repeated assessments and the growth curve reliability (GCR), we assume that the performance of a particular parameter estimate, the stability and variability of the average growth trajectory in the RASCH-LLGC model, is a function of the sample size, the number of items being administrated at each point in time, and the standardized effect size of the average growth trajectory; (3) model application: the capacity of this IRT-LVM comprehensive framework was investigated with two empirical data sets, in which one data set, drawn from part of the British Social Attitudes Panel Survey (1983-1986), revealed the attitude toward abortion of a representative sample of adults aged 18 or older living in Great Britain (McGrath and Waterton, 1986), and the other data set, subsampled from the

National Youth Survey (NYS; Elliott, 1976-1987), depicted the dynamic relations between two interrelated dimensions (namely, social isolation and exposure extent to delinquent peers of adolescents who were aged from 11 to 17 in the year 1976) across five consecutive years (1976-1980).

Since missing data are unavoidable in almost all serious statistical analyses, as an alternative estimation method, the Bayesian inference explicitly models missing outcomes and handles them as extra parameters to estimate (Gelman and Hill, 2007; Jackman, 2000; Patz and Junker, 1999b; Spiegelhalter et al., 2003). Therefore, it becomes straightforward to use this method to effectively estimate any missing values at each iteration. Although the way in which the Bayesian estimation compensates for missing data is similar to the multiple imputation (MI) technique described by Rubin (1987), it extends the MI method by jointly simulating the distributions of variables with missing data, as well as unknown parameters (Carrigan, Barnett, Dobson and Mishra, 2007). It is expected that through this fully Bayesian (FB) method, the missing values can not only be treated as additional parameters to estimate but these parameter estimates can be marginally integrated from an exact joint posterior distribution for all parameters and latent variables. Thus, in the first empirical data example, we illustrate how to incorporate individual-level auxiliary predictors and effectively estimate missing values in a conditional model via the Bayesian estimation approach.

In the second empirical data example, we make use of the multidimensional graded response model (MGRM; De Ayala, 1994; Reckase, 2009) and associative latent growth curve analysis (ALGC; McArdle, 1988) to model the dynamic relations between two interrelated dimensions across five consecutive years (1976-1980). In order to evaluate the performance of this comprehensive modeling approach, we compare and contrast the corresponding parameter estimates using two distinct analytical approaches with a simulated data set, namely, a two-stage IRT-based score analysis and a single-stage IRT-based score analysis. As opposed to the traditionally adopted method (e.g., an average composite), this approach enables the researcher to make use of individual items of the scales at each point in time, allowing the employment of item response characteristics from distinct psychometric models, permitting the separation of time-specific error and measurement error, and providing a common ground for testing measurement invariance across occasions. As for the substantive merit, the following hypothesized associations can be tested: that is, as adolescents perceive themselves to be more socially isolated, the chance that they are engaged with delinquent peers becomes profoundly larger.

Chapter 1

A UNIFIED MODELING APPROACH

As suggested by McArdle (1988), to provide a more rigorous basis for meaningful scaling, the researcher could consider the incorporation of contemporary IRT models and/or the generalized linear models (GLIMs) into the latent growth curve analysis. This is because using the IRT approach provides several distinct benefits over traditional methods. These benefits include facilitating the identification of items which discriminate among respondents across the range of underlying latent abilities, having the report of item statistics and person abilities on the same scale, being flexible in incorporating various auxiliary information, scale construction and measurement invariance examination, and more (see de la Torre and Patz, 2005; Embretson and Reise, 2000; Hambleton and Jones, 1993). When we incorporate random effects in the underlying continuous latent constructs (i.e., when we 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 becomes the generalized linear latent and mixed model, GLLAMM. As a class of multilevel latent variable models, this GLLAMM encompasses the response model and the structural model (Skrondal and Rabe-Hesketh, 2003; 2004), where the IRT model is the response model, and the LGC analysis is the structural model.

A Unidimensional IRT-LVM: 2PL-LGC/2PNO-LGC

In the scenario of unidimensional item response models, the GLIM formulation is typically used. Through a commonly used link function, either a logit or a 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, θ_n represents the latent ability of person *n*, and the model is parameterized as either

$$logit\left[P\left(Y_{in}=1 \mid \theta_n\right)\right]=\alpha_i\left(\theta_n-\beta_i\right) \text{ or}$$

$$probit\left[P\left(Y_{in}=1 \mid \theta_n\right)\right]=\alpha_i\left(\theta_n-\beta_i\right)$$

($i=1,\ldots,I;n=1,\ldots,N$), corresponding to a unidimensional two-parameter logistic (2PL) item response theory model or a unidimensional two-parameter normal ogive (2PNO) model. Here, the abilities can be interpreted either as logits or probits of the probability of a correct response to a particular item. Item difficulty parameters (β_i) are defined as the location of inflection points in the item characteristic curves (ICCs) along the same scale as the latent ability (θ_n), 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 construct, and help determine how well the item discriminates between subjects with different abilities (e.g., Birnbaum, 1968; Lord and Novick, 1968).

As regards the link function, given the similarities between logit and probit of these two models, either model in most applications will give identical substantive conclusions (Liao, 1994; Stefanescu, Berger, and Hershberger, 2005). Normally, by multiplying by a factor of $\frac{\pi}{\sqrt{3}}$, we can go from one set of estimates to the other[2]. However, when we have heavy tails in the distribution of observations, estimates from logit and probit models can differ substantially (Amemiya, 1981). Thus, researchers could opt to use one or the other link function via model comparison. As one of the Markov chain Monte Carlo (MCMC) sampling algorithms, direct Gibbs sampling (Albert, 1992; Chib and Greenberg, 1995; Gelfand, Hills, Racine-Poon and Smith, 1990; Patz and Junker, 1999a) has been implemented 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). That is, the Gibbs sampler can be used for extracting marginal distributions from the full conditional distributions when the complete conditional distributions are of a known distribution form (Geman and Geman, 1984). Therefore, the probit[3] link is considered the more appropriate function for estimating the two-parameter normal ogive (2PNO) IRT-LGC model.

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). Being a useful version of the random coefficient model, a single-domain latent growth

---

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

[3] 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).

curve analysis was presented, 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 (e.g., Byrne and Crombie, 2003; Skrondal and Rabe-Hesketh, 2008). Specifically, like a bifactor model, the univariate latent growth curve model can be formulated as

$$\theta_{(t)n} = \tau_t + \lambda_{0t}\zeta_{0n} + \lambda_{1t}\zeta_{1n} + \varepsilon_{(t)n}$$

($t = 1,\ldots,T; n = 1,\ldots,N$), where the $\theta_{(t)n}$, depicting the propensity of holding the property of a certain dimension at the t^{th} occasion for participant *n*, are the foci of the study; τ_t is the intercept of the structural model; ζ_{0n} and ζ_{1n} are the true initial level and shape factors; and $\varepsilon_{(t)n}$ represents the level-1 residuals. The data are time-structured and balanced in occasions: all subjects were measured on an identical set of occasions and possessed complete data points, $t = 1,\ldots,T$. In addition, the loadings for the initial level factor ζ_{0n} are fixed at $\lambda_{0t} = 1$ ($\forall t$), and the loadings for the shape factor ζ_{1n} are set equal to λ_{1t}. As the nonlinear latent trajectory is essential for analyzing more complicated situations, it has been found useful in establishing a better model-data goodness of fit. In addition, it is feasible to model a nonlinear change trajectory using a bifactor model with free factor loadings for ζ_{1n} (Meredith and Tisak, 1990). According to Raykov and Marcoulides (2006), this level and shape (LS) model is equally useful regardless of whether the developmental trajectory is linear or nonlinear. Finally, to make the model simplified and identifiable, we remove the intercept (τ_t)

from the structural model, set $\lambda_{11} = 0$ and $\lambda_{1T} = 1$, and estimate the coefficients for intermediate time points.

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

$$logit\left[P\left(Y_{i(t)n} = 1 \mid \theta_{(t)n}\right)\right] = \alpha_{i(t)}\left(\theta_{(t)n} - \beta_{i(t)}\right) \text{ or}$$

$$probit\left[P\left(Y_{i(t)n} = 1 \mid \theta_{(t)n}\right)\right] = \alpha_{i(t)}\left(\theta_{(t)n} - \beta_{i(t)}\right)$$

($i = 1,\ldots,I; t = 1,\ldots,T; n = 1,\ldots,N$), where subscript *t* represents the different occasions. In the present study, when the assumption of strong measurement invariance was adopted (Meredith and Teresi, 2006; Sayer and Cumsille, 2001), we impose equality for each of the item parameters over time[4] (i.e., assuming that neither item difficulties nor item discriminations vary across different points in time), which further reduces $\alpha_{i(t)}$ to α_i and $\beta_{i(t)}$ to β_i from the above mathematical formula. If the invariance of the factor structure fails to hold over time, the difference in means may be partially attributable to differences in the scale of the latent variables (Blozis, 2007). Thus, through the estimated item characteristic curves (ICCs) for a unidimensional two-parameter item response model, this unified model can be specified as

$$P(Y_{i(t)n} = 1 \mid \theta_{(t)n}) = \frac{exp(v_{i(t)n})}{\left(1 + exp(v_{i(t)n})\right)}$$

---

[4] 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).

($i = 1,\ldots,I; t = 1,\ldots,T; n = 1,\ldots,N$), where $v_{i(t)n}$ is the linear predictor (i.e., $\alpha_i\left(\theta_{(t)n} - \beta_i\right)$), and again, $\theta_{(t)n}$ can be replaced by $\zeta_{0n} + \lambda_{1t}\zeta_{1n} + \varepsilon_{t(n)}$. As the model becomes complex, for identification purposes we exclude the intercept from the structural model, fix the first discrimination parameter at one, and set the first item difficulty parameter to equal zero. By doing so, we enforce other individual-level covariates to affect the response via the latent variable only (Skrondal and Rabe-Hesketh, 2004).

In summary, with the imposition of a sampling distribution assumption, this GLLAMM can be categorized into three subcomponents: (1) the level-1 sampling model; (2) the link function; and (3) the structural model (Raudenbush and Bryk, 2002). Alternatively, this unified model can be regarded as encompassing the following two parts: either a two-parameter normal ogive model or a two-parameter logistic item response model for the unidimensional binary data, $P\left(Y_{i(t)n} = 1 \mid \theta_{(t)n}, \alpha_i, \beta_i\right)$, is the measurement model, where $\theta_{(t)n}$ represents the latent ability for the subject *n* at the t^{th} occasion, and β_i and α_i are the item parameters. The structural model, $P\left(\theta_{(t)n} \mid \lambda, \zeta\right)$, serves to link the latent abilities with time-varying and time-invariant covariates. Specifically, for instance, the first component, $P\left(Y_{i(t)n} = 1 \mid \theta_{(t)n}, \alpha_i, \beta_i\right)$, the probability that the subject *n* has the ability

$\theta_{(t)n}$ to endorse an item at the t^{th} occasion, is given by the normal ogive item response theory model.

$$P\left(Y_{i(t)n}=1 \mid \theta_{(t)n},\alpha_i,\beta_i\right)$$
$$=\Phi\left(\alpha_i\left(\theta_{(t)n}-\beta_i\right)\right)=\int_{-\infty}^{\alpha_i\left(\theta_{(t)n}-\beta_i\right)}\frac{1}{\sqrt{2\pi}}e^{\frac{-1}{2}z^2}dz$$

($i=1,\ldots,I;t=1,\ldots,T;n=1,\ldots,N$), where $\Phi(\cdot)$ represents the standard normal cumulative distribution function (CDF); and β_i and α_i are the item difficulty and item discrimination parameters for a dichotomously-scaled item *i*. Here, for a given item *i*, we denote its corresponding parameter as ξ_i, that is, $\xi_i=\left(\beta_i,\alpha_i\right)$.

As the second component of the unified model, the underlying latent ability serves as the outcome variable in the structural model, $P\left(\theta_{(t)n} \mid \lambda,\zeta\right)$, which establishes the relation between latent abilities and time-varying and invariant covariates. The time-varying and invariant variables are conceptualized as explanatory covariates for the latent variables. Thus, the corresponding level-1 and level-2 structural model can be specified as

$$\theta_{(t)n}=\zeta_{0n}+\lambda_{1t}\zeta_{1n}+\varepsilon_{(t)n} \text{ and}$$

$$\zeta_{0n}=\nu_{00}+\gamma_{01}W_1+\ldots+\gamma_{0q}W_q+\upsilon_{0n}$$

$$\zeta_{1n}=\nu_{10}+\gamma_{11}W_1+\ldots+\gamma_{1q}W_q+\upsilon_{1n}$$

($t = 1,\ldots,T; n = 1,\ldots,N; q = 1,\ldots,Q$), where the λ_{0t}, λ_{1t}, ζ_{0n}, ζ_{1n} are level-1 factor loadings and latent growth parameters for the initial level and shape factors, and $\varepsilon_{(t)n}$ are independent and identically distributed as $N\left(0,\sigma^2\right)$. With regard to γ_{0q}, γ_{1q}, and W_q, they are level-2 partial regression coefficients and predictors (individual characteristics) of each latent growth parameter, that is, the latent initial status and the change rate, and υ_{0n} and υ_{1n} are followed a bivariate normal distribution with a mean vector of zero and a variance-covariance matrix T, $N\left(0,T\right)$. In this structural model, the growth factors are latent variables with random effects: the level-1 and level-2 models define a population with *N* level-2 units (each individual as the primary sampling unit) and there are *t* ($t = 1,\ldots,T$) level-1 units within each level-2 unit ($n = 1,\ldots,N$). This model assumes that each person was randomly sampled from a larger population and each of them has his/her own latent trajectory.

$$T = var\begin{bmatrix} \upsilon_{0n} \\ \upsilon_{1n} \end{bmatrix} = \begin{bmatrix} \sigma^2_{\upsilon 0} & \sigma_{\upsilon 01} \\ \sigma_{\upsilon 10} & \sigma^2_{\upsilon 1} \end{bmatrix}$$

As with any item response theory model, this IRT-LGC model is over-parameterized and needs to be identified. The indeterminacy is caused by the fact that the item parameters associated with ordered categorical variables and the distribution of underlying continuous variables, $N\left(\mu,\sigma^2\right)$, are not identified. Usually, the identification problem is tackled by fixing $\left(\mu,\sigma^2\right)$ at some pre-assigned values.

Depending upon the specific research question, however, it is better not to impose restrictions on person parameters when these parameters are of primary interest (Lee, 2007). Thus, we consider imposing the identification conditions on the observed categorical variables, the less interesting nuisance parameters. Generally, there are no necessary and sufficient conditions for identifiability: the problem needs to be addressed on a case-by-case basis. In the existing literature, different ways are found for model identification: (1) fixing the first item discrimination parameter at the value of one ($\alpha_1 = 1$), and the first item difficulty parameter at the value of zero ($\beta_1 = 0$) (for binary items) or fixing the first item discrimination parameter at the value of one ($\alpha_1 = 1$), and the first item's first threshold parameter at the value of zero ($\beta_{11} = 0$) (for polytomous items); (2) fixing the first item discrimination parameter ($\alpha_1 = 0$) at the value of one, and the mean of the latent growth intercept at the value of zero ($\zeta_{0n} = 0$); and (3) fixing the product of discrimination parameters at the value of one ($\prod_i \alpha_i = 1$) and the sum of difficulty parameters at the value of zero ($\sum_i \beta_i = 0$) (for binary items) or fixing the product of discrimination parameters at the value of one ($\prod_i \alpha_i = 1$) and the first item's first threshold parameter at the value of zero ($\beta_{11} = 0$) (for polytomous items) (Fox, 2007; Muthén and Muthén, 1998-2007). In this study, either the first or the second scaling option was adopted.

As regards the general assumptions for the IRT-LGC model, taking the two-parameter normal ogive model as an example: given the subject latent ability ($\theta_{(t)n}$) and item parameters ($\xi_i = \left(\beta_i, \alpha_i\right)$), the probability of the subject *n* endorsing a particular item *i* at the t^{th} occasion is defined as

$p_{i(t)n} = P\left(Y_{i(t)n} = 1 \mid \theta_{(t)n}, \beta_i, \alpha_i\right) = P\left(Y_{i(t)n} = 1 \mid \theta_{(t)n}, \xi_i\right)$. It is assumed that each observed outcome variable $Y_{i(t)n}$ follows a Bernoulli distribution with the expectation value of $p_{i(t)n}$,

$$Y_{i(t)n} \mid p_{i(t)n} \sim Bernoulli(\, p_{i(t)n})$$

($i = 1,\ldots,I; t = 1,\ldots,T; n = 1,\ldots,N$). The latent continuous measurement underlying the dichotomous outcomes on the item level is assumed to follow a standard normal distribution. In the structural model, the level-1 residual variance (σ^2) and level-2 variance-covariance matrix (T) are identically and independently distributed as an inverse gamma and inverse Wishart distributions, respectively. Additionally, the level-1 residual variance can be assumed as either homogeneous or heterogeneous across different assessment occasions within individuals, and the level-2 variance component follows a bivariate normal distribution with a mean vector of zero and covariance matrix of T. This variance-covariance matrix T is assumed to be constant for all level-2 clusters. As for the statistical interpretation of random effects, for instance, the second level random intercept, υ_{0n}, accounts for the variation of the initial status (ζ_{0n}) around the fixed population intercept (ν_{00}) not explained by the covariates, W_q. The same interpretation applies to the random shape factor. Finally, the assumptions associated with each level residual can be summarized as follows:

$$E\left(\varepsilon_{(t)n}\right) = 0, \quad E(\upsilon_{0n}) = E(\upsilon_{1n}) = 0,$$

$$var(\zeta_{0n}) = var(\upsilon_{0n}) = \sigma^2_{\upsilon_0}, \quad var(\zeta_{1n}) = var(\upsilon_{1n}) = \sigma^2_{\upsilon_1},$$

$$cov(\zeta_{0n}, \zeta_{1n}) = cov(\upsilon_{0n}, \upsilon_{1n}) = \sigma_{\upsilon 01},$$

$$cov\left(\upsilon_{0n}, \varepsilon_{(t)n}\right) = cov\left(\upsilon_{1n}, \varepsilon_{(t)n}\right) = 0.$$

A Multidimensional IRT-LVM: MGRM-ALGC

Analogously, strengthened by the attributes of the MIRT model and the LVM, a multivariate multilevel polytomous item response theory model embedded in an associative growth curve analysis is proposed. Through the cumulative logit transformation, the logit of responding in category *j* and a higher versus a lower category than *j* (i.e., $log\left(\frac{P^{*}(Y_{in} \geq j)}{P^{*}(Y_{in} < j)}\right)$), the conditional probability of endorsing a particular response alternative, given the latent trait, can easily be specified (e.g., Tuerlinckx and Wang, 2004). Based on a multidimensional item response theory model with simple structure, in which each item measures only one particular latent construct and there is no item in common across different constructs (e.g., Adams, Wilson, and Wang, 1997; McDonald, 1999), we proposed a unified modeling approach using a parallel generalized linear latent and mixed model (*p*GLLAMM) to simultaneously estimate the latent growth trajectories for a dual-domain propensity level. As expected, this approach can be further developed in a straightforward manner, accommodating a more complex structure (e.g., tests with within-item multidimensionality) and a richer set of auxiliary information (e.g., having additional levels above persons). Applications of these models can be found in the literature on educational testing and psychometrics. Corresponding to a logistic multidimensional graded response model, subscript *i* represents an item in a test and the response is scored using *j* for a polytomous item. In this setting, Θ_{dn} represents a trait vector for respondent *n* on dimension *d*, and the model is parameterized as

$$logit\left[P_{ij}^{*}(\Theta_{dn})\right] = log\left\{\frac{P_{ij}^{*}(\Theta_{dn})}{1-P_{ij}^{*}(\Theta_{dn})}\right\} = \alpha_{id}\Theta_{dn} - \beta_{ij}$$

$(i = 1,\ldots,I; j = 1,\ldots,m_i; d = 1,\ldots,D; n = 1,\ldots,N)$. Here, the abilities can be interpreted as logits of the cumulative probability that respondent *n* will endorse a particular item response category *j* and higher, at a given Θ_d level. The elements in the α_{id}-vector stand for the multidimensional discrimination parameters for item *i* on dimension *d*, giving the weight of each dimension *d* on item *i*; the multidimensional item difficulty parameters, β_{ij}, are scalar parameters and can be defined as the location in the latent trait space where the category response surface achieves its maximum slope and, thus, where the item is most informative (Reckase, 1985).

As the multivariate version of the random coefficient model, the multiple-domain latent growth curve analysis allows one to model the situation that individuals differ not only in their intercepts but in other aspects of their trajectory over time with respect to multidimensional latent variables (McArdle, 1988). Again, since the chronological ordering of responses and the clustering of responses within individuals are two important features of longitudinal data, in order to accommodate random effect regressions among the growth factors for two dimensions[5], a longitudinal model with a parallel process of change was proposed to allow for dependence among responses on the same subject (e.g., Cheong et al., 2003; Preacher et al., 2008; Raykov, 2007).

---

[5] This analysis is one kind of structural equation model with regressions among latent variables which represent aspects of distinct individual growth curves, with each of these being modeled along a particular dimension.

Specifically, for each dimension, like a bifactor model, the latent growth curve model can be formulated as

$$\Theta_{d(t)n} = \tau_{dt} + \lambda_{d0t}\zeta_{d0n} + \lambda_{d1t}\zeta_{d1n} + \varepsilon_{d(t)n}$$

($d = 1,\ldots,D; t = 1,\ldots,T; n = 1,\ldots,N$), where the $\Theta_{d(t)n}$ are the foci of the study, which depict the propensity of holding the property of certain dimension ($d = 1$ or $d = 2$) at the t^{th} occasion for participant *n*. The intercept term (τ_{dt}) is typically constrained to zero, yielding a simplified model structure; ζ_{d0n} and ζ_{d1n} are the true initial status and shape factors associated with each dimension *d*; and the $\varepsilon_{d(t)n}$ represent the level-1 residuals for dimension *d* in the structural model. The data under study are time-structured and balanced in occasions: all subjects are measured on an identical set of occasions and provide complete data points, $t = 1,\ldots,T$. In addition, the loadings for the intercept factor ζ_{d0n} are fixed at $\lambda_{d0t} = 1$ ($\forall t$), and the loadings for the shape factor ζ_{d1n} are set equal to λ_{d1t}. To make the model identifiable, we set $\lambda_{d11} = 0$ and $\lambda_{d1T} = 1$ for each dimension, and estimate the coefficients for intermediate time points. In terms of substantial interpretation, fixing $\lambda_{d11} = 0$ indicates that time was centered on the first wave of data collection, which allows the researcher to interpret participant *n*'s initial status from the very beginning of the study (Singer and Willett, 2005). Alternatively, the research could consider adopting the orthogonal design matrix, such as imposing the value of zero for the factor loading associated with the mid assessment occasion, which alleviates the problem of multicollinearity in the latent growth curve model with higher-order polynomial coefficients (e.g., quadratic, or cubic).

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

$$logit\left[P_{ij}^{*}(\Theta_{d(t)n})\right]=log\left\{\frac{P_{ij}^{*}(\Theta_{d(t)n})}{1-P_{ij}^{*}(\Theta_{d(t)n})}\right\}=\alpha_{tid}\Theta_{d(t)n}-\beta_{tij}$$

($i=1,\ldots,I; j=1,\ldots,m_i; d=1,\ldots,D; t=1,\ldots,T; n=1,\ldots,N$), where the additional subscript t represents the different occasions. In the present work, assuming that the assumption of strict measurement invariance holds (Meredith and Teresi, 2006; Sayer and Cumsille, 2001), the residual variances can be constrained to a constant value, and each of the item parameters is identical over time, which further reduces α_{tid} to α_{id} and β_{tij} to β_{ij}. Thus, through the estimated item characteristic curves (ICCs) for a particular unidimensional graded response model, this unified model can be specified as

$$\begin{aligned}
&P\left(Y_{i(t)n}=j \middle| \Theta_{d(t)n}, \beta_{ij}, \alpha_{id}\right) = P_{ij}\left(\Theta_{d(t)n}\right) \\
&= P^{*}_{i(j-1)}\left(\Theta_{d(t)n}\right) - P^{*}_{i(j)}\left(\Theta_{d(t)n}\right) \\
&= P^{*}\left(Y_{in} \geq (j\text{-}1) \middle| \Theta_{d(t)n}, \beta_{ij}, \alpha_{id}\right) - P^{*}\left(Y_{in} \geq j \middle| \Theta_{d(t)n}, \beta_{ij}, \alpha_{id}\right) \\
&= \frac{exp\left(\alpha_{id}\left(\Theta_{d(t)n} - \beta_{i(j-1)}\right)\right)}{1+exp\left(\alpha_{id}\left(\Theta_{d(t)n} - \beta_{i(j-1)}\right)\right)} - \frac{exp\left(\alpha_{id}\left(\Theta_{d(t)n} - \beta_{i(j)}\right)\right)}{1+exp\left(\alpha_{id}\left(\Theta_{d(t)n} - \beta_{i(j)}\right)\right)} \\
&= \frac{exp\left(v_{i(j-1)d(t)n}\right)}{1+exp\left(v_{i(j-1)d(t)n}\right)} - \frac{exp\left(v_{i(j)d(t)n}\right)}{1+exp\left(v_{i(j)d(t)n}\right)}
\end{aligned}$$

$(i=1,\ldots,I; j=1,\ldots,m_i; d=1,\ldots,D; t=1,\ldots,T; n=1,\ldots,N)$, where $P^{*}_{i,m_i}\left(\Theta_{d(t)n}\right)=0$, $P^{*}_{i,0}\left(\Theta_{d(t)n}\right)=1$, and $v_{i(j-1)d(t)n}$ and $v_{i(j)d(t)n}$ are linear predictors, $\left(\alpha_{id}\left(\Theta_{d(t)n}\text{-}\beta_{i(j\text{-}1)}\right)\right)$ and $\left(\alpha_{id}\left(\Theta_{d(t)n}\text{-}\beta_{i(j)}\right)\right)$, respectively. Again, $\Theta_{d(t)n}$ can be replaced by $\zeta_{d0n}+ \lambda_{d1t}\zeta_{d1n}+\varepsilon_{d(t)n}$, for each dimension *d*. As with other estimation approaches, various identification constraints are needed when complex models are encountered. In order to address rotational indeterminacy, in this MGRM-ALGC model we assume a multidimensional model with simple structure, fix the first discrimination parameter associated with each construct to the value of one, with zero loadings otherwise, and constrain the first threshold associated with the first item in each dimension to the value of zero. Moreover, in order to resolve the metric indeterminacy, we try two different scaling options: (1) instead of imposing the constraints on the item threshold parameters, we fix the initial level growth factor associated with each dimension to the value of zero; (2) fixing the level-1 residual variances for each construct to a constant value, either the value of one or not.

Chapter 2

MODEL FORMULATION

In order to differentially weigh individual items and examine developmental stability and change over time, for illustrative purposes, the formulation for a specific model, an MGRM-ALGC, is presented in what follows, since a simpler model formulation can easily be derived. As a class of multilevel latent variable models, this derivative GLLAMM (i.e., *p*GLLAMM) encompasses the response model and the structural model (Skrondal and Rabe-Hesketh, 2003; 2004), where the multidimensional graded response model is the response model, and the associative latent growth curve analysis is the structural model. Thus, with longitudinal designs, the data are multivariate multilevel in nature with a set of ordinal categorical responses nested within each person on each dimension and measurement occasion, with the response model, the structural model, and five-level indices ($i = 1,\ldots,I; j = 1,\ldots,m_i; d = 1,\ldots,D; t = 1,\ldots,T;$ $n = 1,\ldots,N$) as the required elements.

The Measurement 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, 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 with discrete scales of measurement into the model, a second-order factor structure is used to investigate the developmental trajectory over time, which allows the researcher to evaluate the factorial invariance of the latent constructs across measurement waves, and permits the separation of time-specific error and measurement error (Blozis, 2007; Sayer and Cumsille, 2001).

Unidimensional graded response models: GRMs. In the graded response model (GRM; Samejima, 1969, 1997), the probability associated with the observed score equal to and above the threshold category *j* is defined as

$$P^*\left(Y_{in} \geq j \middle| \theta_n, \alpha_i, \beta_{ij}\right) = P_{ij}^*\left(\theta_n\right) = \frac{exp\left(\alpha_i\left(\theta_n - \beta_{ij}\right)\right)}{1 + exp\left(\alpha_i\left(\theta_n - \beta_{ij}\right)\right)}$$

($i = 1,\ldots,I; j = 1,\ldots,m_i; n = 1,\ldots,N$), where Y_{in} denotes the response matrix given by respondent *n* to item *i*, and α_i and β_{ij} represent item discrimination and threshold parameters. Within each item, there are m_i observed response categories and

$\left(m_i\text{-}1\right)$ thresholds. For each alternative ($j = 1, 2, \ldots, m_i$), there exists an ordering relation such that $\beta_{i,1} < \beta_{i,2} < \ldots < \beta_{i,m_i-1}$, in which the corresponding threshold parameters indicate the propensity for a respondent to change from one response category to another. With respect to the discrimination parameter, the model characterizes how well an item discriminates among people with different abilities. Usually, a good item comes with a large discrimination parameter, and with threshold parameters which span a wide range on the trait scale. Unlike the partial credit model, which treats distinct thresholds within each item independently (Masters, 1982), the GRM considers the endorsement of a particular response alternative as requiring the successful accomplishment of all previous steps (e.g., Reckase, 2009). Thus, calculating the probability of endorsing a specific response category can be achieved by

$P\left(Y_{in} = j \mid \theta_n, \beta_{ij}, \alpha_i\right) = P_{ij}\left(\theta_n\right) = P^*_{i(j-1)}\left(\theta_n\right) - P^*_{i(j)}\left(\theta_n\right)$, with

$P^*_{i,m_i}\left(\theta_n\right) = 0$ and $P^*_{i,0}\left(\theta_n\right) = 1$.

Multidimensional graded response models: MGRMs. In practical applications, however, items do not necessarily measure a single unified component; therefore, a more general, multidimensional model should be considered. That is, when an instrument consists of several subscales, the researcher needs to adopt an IRT model of multidimensionality for calculating a respondent's conditional probability of correctly responding to an item. Although psychological processes have constantly been found to be more complex and several subscales on an instrument may tap distinct latent abilities, the abilities are not necessarily independent. With respect to this, the MIRT has shown promise when dealing with situations commonly encountered in educational and psychological testing, such as multiple traits being required for endorsing an item, tests containing mutually exclusive subsets of items, the underlying dimensions being correlated, etc. (e.g., Adams et al., 1997; Reckase, 1997). As an extension of the unidimensional graded response model, a multivariate version for polytomously scored items (De Ayala, 1994; Reckase, 2009) can be expressed as follows:

$$P^*\left(Y_{in} \geq j \middle| \Theta_{dn}, \alpha_{id}, \beta_{ij}\right) = P_{ij}^*\left(\Theta_{dn}\right) = \frac{exp\left(\sum_d \alpha_{id}\Theta_{dn} - \beta_{ij}\right)}{1 + exp\left(\sum_d \alpha_{id}\Theta_{dn} - \beta_{ij}\right)}$$

($i = 1,\ldots,I; j = 1,\ldots,m_i; d = 1,\ldots,D; n = 1,\ldots,N$),where vector Θ_{dn} represents the trait level for subject *n* on dimension *d* (i.e., a person's position in the *d*-dimensional latent space was represented by the vector $\Theta = \left(\theta_1, \theta_2, \ldots, \theta_d\right)$); vector α_{id} stands for the multidimensional discrimination parameters for item *i* on dimension *d*, giving the weight of each dimension *d* on item *i*, and β_{ij}, a scalar parameter, is the multidimensional item difficulty parameter. Like its unidimensional counterpart, in the MGRM, subject responses to item *i* are categorized into m_i ordered categories with $\left(m_i\text{-}1\right)$ category thresholds, and higher category options indicate greater Θ level, in which the Θ level could be any one or any combination of the abilities required for solving an item. Thus, $P_{ij}^*\left(\Theta_n\right)$ can be interpreted as the conditional probability of a randomly selected respondent *n* with latent traits Θ responding in category *j* or higher for item *i*. Because $P_{ij}^*\left(\Theta_n\right)$ is the cumulative probability of responding in category *j* or higher on item *i*, the probability of responding in a particular category, $P_{ij}\left(\Theta_n\right)$, equals the difference between the cumulative probabilities for adjacent categories (i.e., $P_{ij}\left(\Theta_n\right) = P^{*}_{i(j-1)}\left(\Theta_n\right) - P^{*}_{i(j)}\left(\Theta_n\right)$). Moreover, an item's multidimensional discrimination parameters can be interpreted in a similar manner as factor loadings in factor analysis[6]. Thus, based on the scale structure, the relationships among latent traits can be customized accordingly: for instance, traits could have a complex or simple structure, where the complex structure implies that there are one or more items measuring all *d* dimensions, and the former indicates that each item measures exactly what it is supposed to measure (e.g., Bolt and Lall, 2003; Skrondal and Rabe-Hesketh, 2004). As regards the β_{ij}-parameter, it determines the location in the latent trait space at which the category response surface achieves its maximum slope and, thus, where the item is most informative (Reckase, 1985).

---

[6] Even though the statistical formulation and procedure of factor analysis (FA) and MIRT are virtually identical, the research focus and major application for each approach are quite different. Interested readers may refer to Reckase (1997) for more details.

The Structural Model

Compared to traditional longitudinal models, the growth curve model is considered a highly flexible approach, because of its capacity to handle a variety of complexities, such as missing data, unequally spaced time points, non-normally distributed or discretely-scaled repeated measures, non-linear trajectories, and a multivariate growth process (Curran et al., in press). Perhaps the most intuitively appealing way of specifying a latent 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 (shape) across the entire study period, each arising from a specific level in a natural hierarchy, called a two-stage model formulation (Rabe-Hesketh and Skrondal, 2008; Singer and Willett, 2005).

Univariate latent growth curve analysis: LGC. Taking the perspective of latent response formulation, change can be modeled in repeated latent constructs, making it possible for the error in the measurement model to be decomposed into individual time-specific deviation (i.e., $\varepsilon_{d(t)n}$) and measurement error. As Blozis (2007) puts it, being the subject of analysis, the latent variable encompasses time-specific error without the confounding influence of measurement error. This is because at each point in time, a common factor is assumed to account for the dependencies among a set of categorically scored items and allow for the separation of the error variances not attributable to growth.

The level-1 structural model Using LISREL notation, the univariate level-1 structural model can be expressed as follows:

$$\begin{bmatrix} \theta_{d(1)n} \\ \theta_{d(2)n} \\ \theta_{d(3)n} \\ \cdots\cdots \\ \theta_{d(T)n} \end{bmatrix} = \begin{bmatrix} \tau_{d1} \\ \tau_{d2} \\ \tau_{d3} \\ \cdots \\ \tau_{dT} \end{bmatrix} + \begin{bmatrix} \lambda_{d01} & \lambda_{d11} \\ \lambda_{d02} & \lambda_{d12} \\ \lambda_{d03} & \lambda_{d13} \\ \cdots & \cdots \\ \lambda_{d0T} & \lambda_{d1T} \end{bmatrix} \begin{bmatrix} \zeta_{d0n} \\ \zeta_{d1n} \end{bmatrix} + \begin{bmatrix} \varepsilon_{d1n} \\ \varepsilon_{d2n} \\ \varepsilon_{d3n} \\ \cdots\cdots \\ \varepsilon_{d(T)n} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 1 & \lambda_{d12} \\ 1 & \lambda_{d13} \\ \cdots & \cdots \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \zeta_{d0n} \\ \zeta_{d1n} \end{bmatrix} + \begin{bmatrix} \varepsilon_{d1n} \\ \varepsilon_{d2n} \\ \varepsilon_{d3n} \\ \cdots\cdots \\ \varepsilon_{d(T)n} \end{bmatrix}$$

($d = 1,\ldots,D; t = 1,\ldots,T; n = 1,\ldots,N; \tau_{dt} = 0; \lambda_{d11} = 0; \lambda_{d1T} = 1$). Since the repeated measures ($\theta_{d(t)n}$) have been extracted from the unidimensional graded response model through the cumulative logit link function, relating the expected response to the linear predictor, a linear combination of person-specific random effect ($\theta_{d(t)n}$) and item-by-logit indicators, equation above is the structural model. As mentioned earlier, the term $\theta_{d(t)n}$ refers to the propensity of an individual *n* at time *t* on the particular dimension *d,* and is a function of latent variables (representing the underlying initial status (ζ_{d0n}) and the relative growth or decline trajectory (ζ_{d1n})), and time-specific disturbance residuals ($\varepsilon_{d(t)n}$). Additionally, if there is a significant amount of variation to be explained, analysis can proceed in a stepwise manner by adding time varying covariates (TVCs), as time-specific predictors of the repeated measures[7]. To model a nonlinear growth/decline trajectory, we adopt the suggestion of Meredith and Tisak (1990): fixing all λ_{d0t} equal to one and setting λ_{d11} and λ_{d1T} to be zero and one for model identification purposes. By doing so, we let the model freely estimate the intermediate time coefficients. Adopting the assumption typically made in structural equation models (that $\varepsilon_{d(t)n}$ are identically and independently normally distributed with mean (zero) and variance (ψ)), we fix disturbance residuals at the level-1 structural model to be time-homoskedastic, which can be equal to a constant value and makes these time-specific error variances identically distributed over time within each person. Because the random-effect ($\theta_{d(t)n}$) can be further represented by the variances of ζ_{d0n} and ζ_{d1n} at the second level of structural model, the ULGC represents one kind of random-effect model.

$$\Theta_{\varepsilon_d} = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix}_{T \times T}$$

---

[7] As illustrated by Bauer (2009), when fitting models for categorical data, model comparisons are impeded because of the implicit rescaling of the model estimates which take place with the inclusion of new covariates. Thus, in order to have the estimates on a common scale and facilitate the model comparisons, a scaling factor is needed to apply to each component of the random effect model, with the exception that the successive models differ only in the inclusion/exclusion of cluster-level covariates. Since we included no time-varying covariate throughout the present work, a rescaling method would not be applicable here.

The level-2 structural model. The level-2 structural model allows us to distinguish the change trajectories between 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 explained by Singer and Willett (2005), an appropriate level-2 model has the following four 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 level-2 formula specifies a relationship between the individual growth parameter and the time-invariant characteristics of individuals, and (4) each level-2 formula must contain the stochastic component, because those individuals who share a common predictor could vary in their specific change trajectories, hence the name random coefficient models. Thus, an unconditional level-2 LGC model can be expressed as

$$\begin{bmatrix} \zeta_{d0n} \\ \zeta_{d1n} \end{bmatrix} = \begin{bmatrix} \nu_{d00} \\ \nu_{d10} \end{bmatrix} + \begin{bmatrix} \upsilon_{d0n} \\ \upsilon_{d1n} \end{bmatrix}$$

($d = 1,\ldots,D; n = 1,\ldots,N$), where equation above represent regression equations among the latent variables, one for each level-1 growth parameter. In an unconditional model, the ζ_{d0n} and ζ_{d1n} factors have ν_{d00} and ν_{d10} as corresponding intercepts and the residuals are υ_{d0n} and υ_{d1n}. As one of the advantages of casting IRT models in a multilevel structure, the researcher is thereby enabled to incorporate different contextual variables as auxiliary information while estimating the models, which not only improves the estimation of person abilities, but the calibration of item parameters (Mislevy, 1987). Besides, unlike the conventional two-stage procedure, the simultaneous estimation of a multivariate multilevel IRT model avoids the problems of attenuation bias when the study focus is to regress the latent trait variables on other explanatory covariates (Bolt and Kim, 2005). Thus, when time-invariant covariate(s) (TIVs) are introduced into the model, other things being equal, at the individual-level the between-person variability associated with each growth factor can be augmented as

$$\begin{bmatrix} \zeta_{d0n} \\ \zeta_{d1n} \end{bmatrix} = \begin{bmatrix} v_{d00} \\ v_{d10} \end{bmatrix} + \begin{bmatrix} \gamma_{d01} \\ \gamma_{d11} \end{bmatrix} \left[TIV_{1n} \right] + \begin{bmatrix} \upsilon_{d0n} \\ \upsilon_{d1n} \end{bmatrix}$$

Usually, υ_{d0n} and υ_{d1n} are assumed to have a bivariate normal distribution with zero mean and unstructured covariance matrix in both unconditional and conditional situations. Therefore, the distribution associated with residual variances and covariance of the true initial level and shape factors can be expressed as following equation, which permits the level-1 growth parameters to differ across individuals.

$$\psi = COV(\upsilon) = \begin{bmatrix} \sigma^2_{\upsilon_{d0}} & \sigma_{\upsilon_{d01}} \\ \sigma_{\upsilon_{d10}} & \sigma^2_{\upsilon_{d1}} \end{bmatrix}$$

Multivariate latent growth curve analysis: the associative LGC. Despite the fact that developing behaviors are typically intercorrelated, many studies examining the covariance matrix among these behaviors have been static, primarily based on cross-sectional measures taken at one point in time (Duncan, Duncan, and Strycker, 2001). However, with increased interest in the development of interrelated behaviors, the focus of the research has switched from static models to the development of dynamic models, in which the latter incorporate both the time dimension and the intra- and inter-individual variability of behavior trajectories. As originally conceptualized by

Tucker (1966), the multivariate latent growth curve model has been considered a more general and dynamic view of the correlates of change, making it possible for the researcher to obtain both the common and the specific effects of predictors, and examine the associative relationship among several key developmental variables at the same time (Duncan, Duncan, Strycker, Li, and Alpert, 1999; Duncan, Duncan, and Strycker, 2000).

Being extracted from the multidimensional graded response model, each developmental variable of interest is an unobservable propensity level. In order to validate the rationale in conducting an associative LGC, analytically the researcher needs to ensure that there is sufficient interindividual variation in the initial status and the growth rate for each univariate dimension. Once each univariate construct can be successfully modeled, the researcher can model all the developmental latent variables simultaneously. The associative latent growth curve model depicted in Figure 2.1 describes the form of growth and the pattern of associations among growth factors for each dimension of interest. In addition, in order to capture the nonlinear trajectory embedded in each developmental variable, the shape factor loadings are constrained to zero at the first assessment occasion and one at the last assessment occasion, and the coefficients for intermediate time points are freely estimated.

This bivariate latent growth curve model can be expressed as

$$\Theta_{d(t)n} = \zeta_{d0n} + \lambda_{d1t}\zeta_{d1n} + \varepsilon_{d(t)n}$$

($d = 1,\ldots,D; t = 1,\ldots,T; n = 1,\ldots,N$), where we include a level and shape factor (e.g., ζ_{10} and ζ_{11}, and ζ_{20} and ζ_{21} for dimensions θ_1 and θ_2, respectively) and the corresponding deviations. This model allows the identification of growth in each dimension as well as the covariation among them. However, as Ferrer and

McArdle indicate (2003), the relation expressed by the covariance of both slopes is not time dependent, for it overlooks possible interrelations between the dimensions over time[8].

As its univariate counterpart, being capable of allowing the straightforward examination of intraindividual change as well as interindividual variability, the associative LGC makes available a variety of analyses of growth and developmental processes to a wide audience of researchers. For instance, apart from the capabilities leading to greater understanding of multiple developmental trajectories, this associative LGC is also appealing as a way of examining the antecedents, processes, and consequences of change (e.g., Willett and Sayer, 1994). Although many other techniques have been developed to capitalize on the special features of longitudinal research, the class of statistical methods contained in the latent growth curve is highly flexible in model articulation, providing enhanced statistical power for testing hypotheses, and demonstrating greater correspondence between the statistical model and traditional theory used to explain developmental trajectories (Curran et al., in press; Preacher et al., 2008).

---

[8] Unlike the bivariate latent difference scores model (BLDS), this bivariate latent growth model cannot capture the feature of time-lagged sequences between dimensions (Ferrer and McArdle, 2003).

Chapter 3

BAYESIAN INFERENCE

As two major components of the unified model, the existence of a small number of latent factors under multivariate discrete data and the combination of the measurement and structural model for hierarchically nested data structures, a general two-level latent variable model with ordered categorical variables is adopted to account for the individual latent trajectories. Motivated by its various advantages, Bayesian estimation was used for analyzing the current proposed model. Recent MCMC methods in statistical computing for posterior simulation greatly enhance the applicability of the Bayesian inference. Through the application of MCMC to simulate observations from the posterior distribution, one basic strategy is to augment the observed data with the hypothetical latent data which come from latent measurements and latent variables. Thus, in this study, using the Gibbs sampler algorithm coupled with the Metropolis-Hastings algorithm, the MCMC is constructed to circumvent the intractable numerical integration.

Estimating Statistical Complex Models Using the Markov Chain Monte Carlo (MCMC)

Being the centerpiece of Bayesian inference, Bayes' theorem can be expressed as $f(\Omega \mid Y) \propto f(Y \mid \Omega) f(\Omega)$, indicating that the joint posterior density is proportional to the product of the likelihood function and the prior density for the model parameters, where Ω represents the unknown parameters and latent variables, Y denotes the observed response data, and $f(\Omega \mid Y)$ is the posterior probability density function. The "$f(\quad)$" can be replaced by "$p(\quad)$" 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 $f(Y \mid \Omega)$ denotes the likelihood function of the model parameters, given the response data (Y), and $f(\Omega)$ is the prior density for the model parameters, representing the relative likelihoods of particular parameter values before accessing the data.

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 the 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 $f(\Omega \mid Y)$. 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 can 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.

Sampling Procedures

The mechanism by which sampling is conducted varies depending on the known features of the posterior distribution, $f\left(\Omega \mid Y\right)$. In general, various types of sampling algorithms are considered within MCMC, two of which are the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller, 1953) and the Gibbs sampling (Gelfand et al., 1990; Geman and Geman, 1984). Also known as rejection sampling, the key to the former is trying to find a suitable candidate-generating density for suggesting a new value, given the current value in the chain. 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) suggest the use of Metropolis-Hastings within Gibbs (MHwG) for the two- and three-parameter logistic model (e.g., Birnbaum, 1968; Lord and Novick, 1968) as well as the generalized partial credit model (Muraki, 1992).

As a special case of the Metropolis-Hastings, the Gibbs sampling involves cycling through smaller subsets of parameters 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). The subset parameter may be univariate or multivariate, such as sampling from the full conditional posterior distributions of each unknown or blocks of unknowns. As stated by Dunson et al. (2005) and Fox (2007), such techniques as parameter expansion, updating parameters in blocks instead of one by one, have a dramatic impact on computational efficiency and help improve the mixing rate of Markov chains. 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 make it easier to simulate draws (Casella and George, 1992; Jackman, 2000; Patz and Junker, 1999a, 1999b).

Specification of Priors

As mentioned 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 and latent variables plays an important part in the Bayesian approach. Unlike those of the frequentists, Bayesian methods provide a clear channel 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). 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). Moreover, as sample sizes increase, priors are generally asymptotically irrelevant, and the estimates obtained from the Bayesian and frequentist methods should approach identical values (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). 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 the sampling in MCMC computationally efficient (Johnson et al., 2007; Kim and Bolt, 2007; Rupp et al., 2004). In other words, by assigning noninformative priors to the model parameters of interest, the researcher allows the data to provide as much information as possible by themselves. However, in order to facilitate model identification, the researcher may consider using a prior density with high precision. For instance, throughout the present study a normal prior with high precision was utilized for item difficulty parameters, and a truncated normal prior was adopted for item discrimination parameters. The complete specification of different priors can be found in the appropriate sections of the practical illustration chapter.

Monitoring the Markov Chain(s) and Evaluating the Model Goodness of Fit

For the model estimated via the Bayesian Markov chain Monte Carlo (MCMC), as implemented in WinBUGS 1.4.3 (Spiegelhalter et al., 2003), the 'burn-in' period for the MCMC chains was determined using the method proposed by Gelman and Rubin (1992). Although Geyer (1992) suggests that generating one single long chain is more efficient in using the simulation output, it leads to more complex Monte Carlo standard error expressions[9]. As opposed to running a single long sequence, Gelman and Rubin (1992) argue that, to monitor the model convergence, it is important to run multiple chains using a range of different starting values (Seltzer, Wong, and Bryk, 1996). Thus, in the present work we perform Bayesian analysis using multiple independent chains with over-dispersed starting values.

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 whenever possible from existing maximum likelihood-based estimation programs. 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. In order to ensure that each chain has converged to its stationary distribution and stable parameter

---

[9] 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).

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) propose 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, mathematically, the DIC is defined as $DIC = \overline{D(\Omega)} + p_D$, where $\overline{D(\Omega)}$ is a measure of lack of fit, representing an estimated average discrepancy between model and data, and 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, estimate of the DIC index can be requested from the WinBUGS program, in which a smaller DIC represents a better fit of the model and 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 another criterion used for assessing the model goodness of fit (Gelman, Carlin, Stern, and Rubin, 2003). Mathematically, the posterior predictive distribution can be written as: $P\left(Y^{rep} \middle| Y\right) = \int P\left(Y^{rep} \middle| \Omega\right) P\left(\Omega \middle| Y\right) d\Omega$, where Y^{rep} denotes replicated values of Y, and Ω represents all model parameters and latent variables (Sinharay and Stern, 2003). The integral defining the posterior predictive distribution consists of two parts: the sampling distribution ($P\left(Y^{rep}\middle|Y\right)$) and the posterior distribution for model parameters and latent variables ($P\left(\Omega\middle|Y\right)$). That 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; an overall discrepancy statistics utilized in one of the present study is the Bayesian chi-square: the sum of squares of the outfit measures[10].

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 $T\left(Y^{rep}\right)$, and defined as $p = P\left(T\left(Y^{rep}\right) \geq T(Y)\right)$, where this tail-area probability (or *p*-value) is estimated from the simulation as the proportion of the *N* replications for which $T\left(Y^{rep}\right) \geq T(Y)$, 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

---

[10] Even though it has advantages over standard applications of fit statistics, this chi-square-type measure

the model, and suggests that the presumed model does not 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).

---

should be interpreted with great caution. According to Sinharay et al. (2006), in IRT model checking it is not a suitable discrepancy measure and fails to detect the problems with inadequate psychometrics models.

Chapter 4

PRACTICAL ILLUSTRATION

The ease of implementing Markov chain Monte Carlo (MCMC) simulation methods demonstrates much potential for statistically complex models in which they can find future application. In this section, the utility of this IRT-LVM comprehensive framework was investigated with examples using both simulated and empirical data, in which three models were presented in turn, namely, the unidimensional Rasch (1960) and linear latent growth curve model (RASCH-LLGC); the unidimensional two-parameter normal ogive (e.g., Birnbaum, 1968) and nonlinear latent growth curve model (e.g., Meredith and Tisak, 1990) (2PNO-LGC), and the multidimensional graded response (e.g., De Ayala, 1994) and associative latent growth curve model (e.g., McArdle, 1988) (MGRM-ALGC).

Using the RASCH-LLGC to Evaluate the Model Parameter Estimate Performance

Unlike the two-parameter IRT model, the Rasch model assumes an identical discrimination parameter for each item, implying that the relative severity of the items is indistinguishable for all subjects (Rasch, 1960). Other key assumptions include (1) local independence and (2) additivity, in which the former represents a set of items measuring a single underlying latent variable; the latter implies that there is a readily interpretable ordering of items and persons, since item differences and person differences contribute additivity to the same scale, the log-odds of an affirmative response (Johnson and Raudenbush, 2006; Raudenbush, Johnson, and Sampson, 2003). As for the structural component, expanding on traditional repeated-measures analysis, the linear latent growth curve model allows one to simultaneously model within-person change patterns, and between-person differences in the characteristics of latent trajectories (Curran, et al., in press).

Monte Carlo simulation study. Under the framework of the IRT-LGC, we demonstrate how to evaluate the performance of parameter estimates through conducting a Monte Carlo study. As the sample size needed for a particular longitudinal study depends on many factors, such as the complexity of the model, the number of assessment occasions, the standardized effect size associated with the polynomial coefficient of interest (ex., linear, quadratic, or cubic), the variation between and within participants, the amount of missing data, etc (Curran et al., in press; Hertzog, von Oertzen, Ghisletta, and Lindenberger, 2008; Muthén and Muthén, 2002; Raudenbush and Liu, 2001), an "adequate" sample size is hard to unambiguously determine. As a simplified illustration, a specific IRT-LGC model is investigated, in which the Rasch model for dichotomous items is the measurement model and a linear latent growth curve model (LLGC) with four equidistant time points is the structural model. Given the constant number of repeated assessments and the growth curve reliability (GCR), we assume that the performance of a particular parameter estimate, the stability and variability of the average growth trajectory, is a function of sample size, the number of items being administrated at each point in time, and the standardized effect size of the average growth trajectory.

Based on a Monte Carlo sample size study, Muthén and Muthén (2002) suggest that for a linear growth curve model without a covariate (i.e., a unconditional model), the following specification of the covariance matrix reflects a commonly seen scenario, showing that the variation of the intercepts is generally larger than that of the linear growth rate in longitudinal studies, and the covariance between them is set to zero.

$$T = \begin{bmatrix} .5 & 0 \\ 0 & .1 \end{bmatrix}$$

In addition, according to Hertzog et al. (2008), they indicate that the GCR would have an impact on the power of detecting individual differences associated with the change profile, that is, the variance of the slope factor. Having two components, the GCR can be defined as the variance determined by the latent growth curve at each point in time, divided by the total variance of repeated measures. In this study, to partial out the influence of this confounding factor, we assume that residual variances are homogeneous across different points in time and fixed at the value of one, which is the general practice for conducting power analyses in the multilevel model framework (Snijders and Bosker, 1993). In order to have acceptable GCR values across the entire study period, we follow Muthén and Muthén's (2002) observation and rescale the elements in the covariance matrix by a factor of 2, which results in a modified covariance matrix and the respective GCR values of .50, .55, .64, and .74[11].

$$T = \begin{bmatrix} 1 & 0 \\ 0 & .2 \end{bmatrix}$$

Adopting Cohen's definition of the magnitude of effect sizes (Cohen, 1988), we specify two different standardized effect sizes for the mean of the linear growth trajectory: that is, the small effect size (.14) and the medium effect size (.28). These values are calculated as follows:

$$\delta = \frac{\nu_{10}}{\sqrt{\sigma^2_{\upsilon 1}}}$$

, where δ is the magnitude of the standardized effect size, and ν_{10} and $\sigma^2_{\upsilon 1}$ represent the overall linear time effect and the corresponding variance associated with this linear slope factor. Using the values of .316 and .632, we obtain the corresponding small and medium effect sizes for the linear growth trajectory (ν_{10}); that is, .14 ($.316\times\sqrt{.2}$) for the small effect size and .28 ($.632\times\sqrt{.2}$) for the medium effect size.

---

[11] The formula for calculating the GCR can be expressed as,

$$R^2(\theta_t) = \frac{\left(\sigma^2_{\upsilon 0} + t^2\sigma^2_{\upsilon 1} + 2t\sigma_{\upsilon 01}\right)}{\left(\sigma^2_{\upsilon 0} + t^2\sigma^2_{\upsilon 1} + 2t\sigma_{\upsilon 01} + \sigma^2_{\varepsilon_t}\right)}$$

, where $\sigma^2_{\upsilon 0}$, $\sigma^2_{\upsilon 1}$ and $\sigma^2_{\upsilon 01}$ are the variances and covariance associated with the intercept and slope factors; $\sigma^2_{\varepsilon_t}$ is the residual variance for the underlying latent variables at time *t*, and *t* is the time coefficient (i.e., 0, 1, 2, and 3) in a linear growth trajectory model (Muthén and Muthén, 2002).

As regards the number of items being administered at each time in the same point, we chose 5, 10, and 15 items to represent three different lengths of the scale. Using the unidimensional Rasch model, item difficulty parameters were selected from the range of [-2, 2] with equal intervals. For instance, for a scale of 5 items, the item difficulty parameters are pre-specified as β=[-2,-1, 0, 1, 2]. For a 10-item test, the item difficulty parameters are β=[-2, -1.556, -1.111, -.667, -.222, .222, .667, 1.111, 1.556, 2]. Analogously, for a test of 15 items, the item difficulty parameters are β=[-2, -1.714, -1.429, -1.143, -.857, -.571, -.286, 0, .286, .571, .857, 1.143, 1.429, 1.714, 2]. The observed dichotomous outcome variables from this RASCH-LLGC model were generated by comparing the probability of the correct response with a random number generated from a standard uniform distribution, U[0, 1].

As Curran et al. note (in press), in order to have reliable estimates from the growth curve models, sample sizes approaching at least 100 are often preferred. However, achieving accurate estimates in LGC models with discretely scaled variables requires relatively large sample sizes. Generally speaking, Lee (2007) suggests that, when analyzing dichotomous data, researchers need at least "30a" sample sizes in order to achieve reasonably accurate results, where "a" is the number of unknown parameters. Therefore, as the unknown parameters in this RASCH-LLGC model with three different lengths of scale are 8, 13 and 18, we select sample sizes of 125 and 250 as the two investigating levels[12].

---

[12] Even though the sample sizes for these two investigating-level seem small in the typical IRT model estimation, Muthén and Curran (1997) argues that in growth models it is the total number of person-by-time observations that plays an important role in model estimation and statistical power.

In summary, to evaluate the numerical behavior of the average growth trajectory (i.e., the stability and variability of latent mean associated with the slope factor in an RASCH-LLGC model), the simulation used a 2×3×2 design with 12 conditions, in each of which a total of 100 replications were generated using the free software R (R Development Core Team, 2009) and the models were implemented and estimated using WinBUGS 1.4.3 (Spiegelhalter et al., 2003). Specifically, we generate data sets which represent the alternative hypothesis (i.e., the mean of the slope factor is statistically significant different from the specified values, .14 and .28). However, in Bayesian analysis, using the percentage of replications where the null hypothesis was rejected as a proxy estimate for power determination should proceed with caution. As indicated by Lee (2007), the standard error estimates are usually overestimated in Bayesian SEM analysis. Thus, he suggests that the hypothesis testing should be approached by means of model comparisons through the Bayes factor (BF) or DIC, in particular for models with dichotomous variables. Also, as the information carried by the dichotomous data is relatively rough, it is important to monitor the model convergence with great care, for it requires more iterations for the MCMC algorithm to converge. Therefore, for each replication, we execute the algorithm by means of running three independent chains with over-dispersed initial values and take the first 25,000 iterations as the burn-in period for each chain. That is, a total of an additional 15,003 (5,001*3) iterations for three chains was carried out to define the sampling distribution of each parameter in the model. In addition, a common method used for assessing convergence is to compute the Gelman-Rubin statistic, the potential scale reduction factor (PSRF), which compares within-chain variability to the variability among chains (Gelman and Rubin, 1992). When for each parameter of interest the PSRF approaches one, it suggests that the model reaches convergence. Finally, the summary of population values used in this RASCH-LLGC model can be found in Table 4.1.2.

The following criteria are used for evaluating the model parameter performance, such as the bias (BIAS), the root mean squares (RMS) between the true values and the corresponding estimates, and the ratio of the standard errors estimates to the sample standard deviations, $SE(\hat{v}_{10})/SD(\hat{v}_{10})$, in which the bias of the estimates and the root mean squares between the true values and the corresponding estimates are computed as follows:

$$BIAS \text{ of } \hat{v}_{10} = E\left[\hat{v}_{10}^{r} - v_{10}^{0}\right]$$

$$RMS \text{ of } \hat{v}_{10} = \left\{\frac{1}{100}\sum_{r=1}^{100}\left[\hat{v}_{10}^{r} - v_{10}^{0}\right]^{2}\right\}^{1/2}$$

, where $\hat{v}_{10}^{r}$ and v_{10}^{0} are the r^{th} estimate of v_{10} and its true value, respectively. In order to study the behavior of the numerical standard error estimates, let $SD(\hat{v}_{10})$ be the sample standard deviation obtained from $\left\{\hat{v}_{10}^{r} : r = 1,\ldots,100\right\}$, and $SE(\hat{v}_{10})$ be the mean of the numerical standard errors estimates of $\hat{v}_{10}$ obtained via, $E\left[\left(T^{*}-1\right)^{-1}\sum_{t=1}^{T^{*}}\left(v_{10}^{(t)} - \hat{v}_{10}\right)\left(v_{10}^{(t)} - \hat{v}_{10}\right)^{T}\right]$, where T^{*} is the total number of simulates obtained from the posterior distribution, and

$\hat{v}_{10} = T^{*-1} \sum_{t=1}^{T^*} v_{10}^{(t)}$. When the standard errors estimates are close to the sample standard deviations, $SE(\hat{v}_{10})$ should be close to $SD(\hat{v}_{10})$, and the ratio of $SE(\hat{v}_{10}) / SD(\hat{v}_{10})$ should be close to one, in which the ratio can be used for assessing the behavior of the numerical standard error estimates. Thus, based on the definitions of $SE(\hat{v}_{10})$ and $SD(\hat{v}_{10})$, it is found that the sample standard deviation of $\left\{\hat{v}_{10}^{r} : r = 1,\ldots,100\right\}$ is smaller than the mean of the numerical standard error estimates, indicating that the variability of the Bayesian estimates, the average change rate, is relatively small, which may be regarded as an advantage of the Bayesian estimates. However, it also indicates that the numerical standard error estimates of the Bayesian approach ($SE(\hat{v}_{10})$) are overestimated, which is in line with our expectations, as a converged MCMC chain will have explored all of the parameter space and provided a full picture of the posterior distribution. Finally, it is found that in most cases the design factors investigated in the present study, such as the sample size, the standardized effect size, and the number of items, all execute positive influences with respect to the stability and variability of the parameter estimate of interest (see Table 4.1.3). That is, by increasing the sample size, the magnitude of the standardized effect size, and the number of administered items, the promise of reducing bias and increasing precision for the average growth trajectory in the RASCH-LLGC model can be validated.

Prior knowledge incorporation. In this section, we demonstrate how the use of prior information affects the parameter estimates and standard deviations from a small data set. In the previous simulation study, baseline priors and conjugate priors are used in all Bayesian analyses. Specifically, the mean of the shape factor is estimated using a normal distribution prior. As regards the covariance matrix of the random effect parameters, the conjugate prior, the inverse Wishart distribution, is used. As for the item difficulty parameters, in order to facilitate model identification, we adopt a normal prior density with tight precision and treat them as the baseline priors. The complete specifications of the least-informative, half-informative and full-informative priors are displayed in Table 4.1.4.

Using the least-informative, half-informative and full-informative priors, the results of parameter estimates and associated standard deviations from the simulated data set, one with a small standardized effect size of the average growth trajectory (.14), a sample size of 125, and ten dichotomous items (SE125I10), are given in Table 4.1.5. The results appear to show that the standard deviations when adopting vague priors were relatively large. When analyzing the data again with half- and full-informative priors, the corresponding standard deviations were reduced: obviously, with more information on priors, the standard deviations became smaller through comparing their counterparts which had been obtained using half- and full-informative priors. This illustrates the way in which the use of informative priors can increase the statistical power and reduce parameter uncertainty, implying that informative priors can be viewed as additional or extra data points (Gelman and Hill, 2007; Zhang et al., 2007). Thus, through Bayes' law, we demonstrate how posterior probabilities are revised in the light of new information and bridge individual expressions of uncertainty to contact with real-world data generating mechanism.

Fit of the 2PNO-LGC to the Abortion Data

Despite the large number of components requiring attention when selecting an appropriate statistical model, this section restricts its focus to the following issues: (1) model formulation: how Bayesians explicitly incorporate multiple dichotomous repeated measures into a latent growth curve analysis. In order to differentially weigh individual items, and examine developmental stability and change over time, one specific model, an 2PNO-LGC, is presented, in which the model combines the two-parameter normal ogive item response theory model (e.g., Lord and Novick, 1968) and latent growth curve analysis (e.g., Meredith and Tisak, 1990); (2) model equivalence: 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, such as the HLM and LGC models. To assess the advantages and disadvantages of these two distinct modeling frameworks, we illustrate their different characteristics and use in applications with simulated data; (3) missing data compensation: as 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, when the missing data generation mechanism, missing at random (MAR; Rubin, 1987), is sustainable, 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 et al., 2007; Gelman and Hill, 2007).

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 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 4.2.1[13]. However, when we perform a confirmatory factor analysis (CFA) to examine the underlying construct using the software of Mplus (Muthén and Muthén, 1998-2007), we find these seven items seem not to 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 4.2.1) and remove the extreme circumstance factor from subsequent analyses. By doing so, the gamma change[14] can be ruled out through conducting a CFA on the scale at four time periods. 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.

The breakdown of analyses and response patterns for complete cases and available cases can be found from Table 4.2.2 to Table 4.2.5. 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 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

---

[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 be inferred only 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.

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 kept in mind constantly (Knott, Albanese, and Galbraith, 1990; Fienberg and Rinaldo, 2007).

The sampling method is 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 is to look at trends and changes in attitudes over time, a longitudinal rather than a repeated cross-sectional design is adopted here (McGrath and Waterton, 1986; Wiggins et al., 1990). In this 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 used 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. Thus, a conceptual modeling framework is depicted in Figure 4.2.1.

Unconditional models. In subsequent analyses, baseline priors and conjugate priors are used for the measurement model parameters and structural model parameters. Specifically, the means of initial level and shape are estimated using normal distribution priors, and two kinds of non-informative prior 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 4.2.6. 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 are performed. It is found that the results from these two approaches are close to each other within at least one decimal place: in the situation of running three independent chains, the first 20,000 iterations are 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 used to define the posterior distribution of each parameter. Similarly, for a single long chain, we use a burn-in period of 19,998, with parameter estimates based on the 50,000 subsequent iterations (see Figures 4.2.2-4.2.3). The output is 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. When 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 Figure 4.2.4). 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 4.2.7, 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 4.2.8 (the right panel), where we see that the estimated discrimination parameters for item 2 and item 3 are both greater than one and larger than for the other two items, indicating that item 2 and item 3 better discriminate the underlying propensity level than do item 1 and item 4. 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 associated with item 4 is the largest among the four, indicating that "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 to abortion over time (their estimated values are 1.077, .581, 1.095, and .391, respectively, being statistically significant at the first, and third points of time), suggest that the existence of additional outcome variation at level-1 of the structural model 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 to abortion.

Model equivalence. 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, such as the HLM and LGC models. To assess the advantages and disadvantages of these two distinct modeling frameworks, we illustrate their respective characteristics and application use with a simulated data set, in which the population values were adopted from a previously modified analysis result, the one with the probit link and constant level-1 residual variance. The simulated data are generated using the free software of R (R Development Core Team, 2009), and the models are implemented and estimated using WinBUGS 1.4.3 (Spiegelhalter et al., 2003). As indicated before, in the structural model, ‘time’ in the HLM and LGC model has specific consequences for the analysis results:

$$\theta_{(t)n} = \lambda_{0t}\zeta_{0n} + \lambda_{1t}\zeta_{1n} + \varepsilon_{t(n)} \text{ and}$$

$$\zeta_{0n} = \nu_{0n} + \upsilon_{0n}$$

$$\zeta_{1n} = \nu_{1n} + \upsilon_{1n}$$

($t = 1,\ldots,T; n = 1,\ldots,N$). In the HLM, ζ_{0n} and ζ_{1n} are random parameters and λ_{1t} is an observed variable representing time or a time-varying covariate, which makes HLM the best approach if there are a great many variations of occasion between individuals (Snijders, 1996; Willett and Sayer, 1994). However, in the LGC, ζ_{0n} and ζ_{1n} are the latent variables and λ_{0t} and λ_{1t} are factor loadings. Because λ_{1t} cannot vary across subjects, LGC is considered best suited for time-structured data or a fixed occasion design (e.g., Byrne and Crombie, 2003). Although LGC modeling can be used for designs with varying occasions by modeling all existing occasions and viewing the varying occasions as problems of missing data, this approach is difficult to manage when the number of varying occasions is excessive (Bauer, 2003; Curran, 2003; Hox and Stoel, 2005).

As can be seen in Table 4.2.9, the parameter estimates are rather similar and both approaches lead to identical substantive conclusions. However, there is a caution: to facilitate the comparison between these two approaches, in the HLM we manually fix the estimates of the time variable to be the same as the true values, since time coefficients in the HLM are fixed explanatory variables (i.e., we fix the population parameters $s2$ equal to -1.741, and $s3$ equal to .064), which makes the number of estimated parameters in the HLM two fewer than their counterparts in the LGC model. In addition, according to the overall goodness of fit provided via the deviance information criterion (DIC) (Spiegelhalter et al., 2002), we conclude that these two models fit the data equally well.

Generally, latent growth curve analysis is preferred in many situations because of its greater flexibility. For instance, standard SEM software supplies more options, such as providing omnibus goodness-of-fit indices for a model (i.e., allowing for a saturated model with which any fitted model can be compared) and being more flexible in modeling and hypothesis testing (i.e., testing complex mediational mechanisms through the decomposition of effects and investigating moderational mechanisms through multiple group analysis, to name only a few) (Bauer, 2003; Chou, Benter and Pentz, 1998; Curran, 2003; Hox and Stoel, 2005; MacCallum et al., 1997; Willett and Sayer, 1994). Still, the HLM is preferable whenever the growth model must be embedded in a larger number of hierarchical data levels (Snijders, 1996). Adding additional layers to the model is relatively difficult if the SEM framework is used. While several key differences remain between these two models, at the time of writing, the discrepancies are rapidly disappearing (Preacher et al., 2008; Raykov, 2007).

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 complete 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 the results from the full dataset (the one containing missing outcomes) do not differ systematically from the complete cases in unconditional models, the unprovable missing data generation mechanism, missing completely at random (MCAR; Rubin, 1987), seems sustainable. Moreover, a hypothesis regarding the missing data mechanism is tested: the corresponding significance value associated with Little's MCAR test (Little, 1988) is .222, indicating that the data are missing completely at random. As mentioned earlier, because Bayesian treats missing values as additional parameters which need to be estimated, for those respondents with incomplete survey responses, handling missing data this way helps 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 4.2.10), 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 auxiliary predictors of gender, age, and religious status, to deal with the multivariate missing categorical data at each occasion. Based on the result shown in Table 4.2.11, both data sets provide estimates with identical substantial interpretation and 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.

Taken together, the application of IRTs to responses gathered from repeated assessments allows us to take into consideration the characteristics of both item responses and measurement error in the analysis of individual developmental trajectories. As a simplified demonstration, in the present study we consider the modeling of a unidimensional latent construct only. 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, and simultaneously make use of their interrelationship when we have multiple interrelated dimensions across the entire study period.

Using the MGRM-ALGC to Study the Parallel Process of Change

As a simplified demonstration, the goal of the following analyses is to illustrate how this comprehensive hybrid model, the MGRM-ALGC, allows one to depict relations among respective growth factors using data from the National Youth Survey (NYS; Elliott, 1976-1987).

Participants. Based on a multistage cluster-sampling design, the NYS employed a probability sample of households in the continental United States. The sample covers urban, suburban, and rural geographic areas. To be assessed for five consecutive years, the panel sample comprised 1,725 adolescents ranging from 11 to 17 years of age (*M*=13.87, *SD*=1.945) at Year 1, 1976. Of these 1,725 randomly selected participants, 838 completed all 13 outcome measures across five occasions (i.e., after listwise deletion of all missing values, the number of complete cases is 838, implying that attrition and other form of missingness approximated half the size of the sample). The participants described themselves as Caucasian (*n*=690), African American (*n*=99), Mexican American (*n*=35), Native American (*n*=4), Asian (*n*=8), and others (*n*=2). Among them, 82.6% percent were from two-parent families. The questionnaire covered a wide array of measures to assess participants' social isolation status and their exposure extent to delinquent peers. Adolescents with complete demographic data[15] (*n*=802) reported a slightly higher level than their counterparts with incomplete responses (*n*=36), except for the second and third assessment occasions; similarly, adolescents with complete demographic data (*n*=802) reported a somewhat greater extent of exposure to delinquent peers than their counterparts with incomplete cases (*n*=36), except for the third and fifth assessment occasions. However, no statistically significant difference was detected in the two situations. Descriptive statistics for each dimension's IRT scale scores are presented in Tables 4.3.1a and 4.3.1b.

Measures. Few studies consider the dynamic relations between adolescents' mental health and other problem behaviors, although there has been substantial evidence of their relations in both cross-sectional and longitudinal samples (e.g., Cohen, Reinherz, and Frost, 1994; Swahn and Dovonan, 2003). Thus, in the present study we decide to examine the associations between adolescents' social isolation and engagement with delinquent peers through the observation of dynamic trajectories between these two dimensions. The selection of these two constructs was based on the extant literature, suggesting a link between the way in which adolescents perceived their emotional status and the likelihood that they were associated with delinquent peers. Based on this conceptual framework, we are interested in examining the corresponding dynamics underlying this bivariate system as it evolved over time. A total of 13 polytomous items were selected as outcome measures on each occasion, each of which is a five-point Likert-type scale with higher scores reflecting severe status. Among them, the first six variables measure the construct of social isolation and the remaining seven describe the extent of adolescents' exposure to delinquent peers (see Table 4.3.2).

Dimensionality assessment. As part of the investigation of the NYS, the data represent the responses to 13 items regarding adolescents' social isolation status and the extent of their exposure to delinquent peers by a selected panel of 838 from the years 1976 to 1980. A confirmatory factor analysis (CFA) with categorical indicators was

---

[15] Demographic variables include the marital status of their parents, family income, gender, ethnicity, and

performed to examine the dimensionality using Mplus (Muthén and Muthén, 1998-2007). The response frequencies for these 13 items are listed in Table 4.3.2. As observed in the frequency table, it was found that response alternatives equal to or greater than three tend to have small frequencies, implying that the data were rather sparse and asymptotic normality of the maximum likelihood estimator may not apply. The CFA results suggested that these 13 items measured two latent constructs for each of the five years. The fit of the five models was respectable, with Comparative Fit Indices (CFI) between .965 and .982, Tucker-Lewis Fit Indices (TLI) between .973 and .985, and Root Mean Square Error of Approximation (RMSEA) between .043 and .071.

Scores from perceived social isolation and exposure extent to delinquent peers are plotted in Figures 4.3.1a and 4.3.1b. Each of the plots contains data from a random subsample of 44 adolescents, in which each line represents an individual's IRT scale scores followed through five occasions. These plots illustrate some important features of the data. Generally, intra-individual variability over time is evident. This observation applies for both dimensions. Also, there is great inter-individual variability within groups, indicating great change heterogeneity.

Identification constraints and prior distribution specification. As with other estimation approaches, various identification constraints are needed when complex models are encountered. In the present study, for the MGRM-ALGC model, in order to address rotational indeterminacy, we assume a multidimensional model with simple structure (i.e., each item measures one dimension of ability and there is no cross-loading of items), fix the first discrimination parameter associated with each construct to one and zero loadings otherwise (i.e., alpha[1,1]<-1, alpha[1,7:13]<-0; alpha[2,1:6]<-0,

---

age.

alpha[2,7]<-1), and constrain the first threshold associated with the first item's multidimensional item difficulty parameter in each dimension to zero (i.e., d[1,1]<-0; d[2,7]<-0). Moreover, in order to resolve the metric indeterminacy, we compare and contrast two different scaling options: either constraining the initial latent growth factor from each dimension to the value of zero or fixing level-1 residual variances for each construct to a constant value (i.e., set variances for both θ_1 and θ_2 equal to particular constants). As regards model convergence checking and subsequent statistical inference, we adopt Gelman and Rubin's (1992) suggestion of running three independent chains with over-dispersed starting values. Because WinBUGS treats an initial 4,000 iterations as the default adaptive phase under the general normal- proposal Metropolis algorithm, we take these 4,000 iterations as the burn-in period and sample an additional 4,000 iterations from each independent chain (Spiegelhalter et al., 2003). Thus, the point estimate of the model parameter and corresponding standard error were computed from the mean and standard deviation of the remaining 12,000 observations (i.e., 12,000=4,000*3) sampled from each parameter's marginal posterior distribution. For instance, the mean estimate of an overall time effect associated with a particular dimension ($\hat{v}_{d10}$) can be calculated as $\hat{v}_{d10} = \left(T^*\right)^{-1} \sum_{t=1}^{T^*} v_{d10}^{(t)}$, where T^* is the total number of simulates obtained from the posterior distribution. Since we have large sample of v_{d10} from its posterior distribution, an estimate of $SE\left(\hat{v}_{d10}\right)$ can be directly obtained from the sample covariance matrix,

$$E\left[\left(T^{*}-1\right)^{-1}\sum_{t=1}^{T^{*}}\left(v_{d10}^{(t)}-\hat{v}_{d10}\right)\left(v_{d10}^{(t)}-\hat{v}_{d10}\right)^{T}\right]^{\frac{1}{2}}$$

. As T^{*} becomes infinity, these Bayesian estimates tend to approach to their corresponding posterior means in probability.

As regards the prior density specification, in subsequent analyses baseline priors and conjugate priors are used for the measurement model parameters and structural model parameters. That is, order to facilitate model identification, a normal prior with tight precision, *N*(0, .5), was utilized for item difficulty parameters, and a truncated normal prior, *N*(0, 1.0E-02)I(0,) was adopted for item discrimination parameters. In addition, the level-1 residual variance (σ^{2}) is identically and independently distributed as an inverse gamma distribution with shape and scale parameters being set to the value of one. Specifically, in the unidimensional GRM-LGC model, the means of initial level and shape factors are estimated using multivariate normal distribution priors. In regard to the covariance matrix of the random effect parameters, the conjugate prior, the inverse Wishart distribution is adopted. As for the MGRM-ALGC model, the Θ-vector is next decomposed into two sets of latent growth factors and assumed to be distributed as a multivariate normal distribution. For both dimensions, the means of initial level and shape factors are estimated using multivariate normal distribution priors, and the inverse Wishart distribution is adopted for the covariance matrix of the random effect parameters from each dimension. The complete specification of different priors can be found in Table 4.3.3.

Empirical results. Extracted from the multidimensional graded response model, each developmental variable of interest is an unobservable propensity level. In order to validate the rationale in conducting an associative LGC, analytically the researcher needs to ensure that there is sufficient interindividual variation in the initial status and growth rate for each univariate dimension. Once each univariate construct can be successfully modeled, the researcher can model all the developmental latent variables simultaneously. The associative latent growth curve model used in the present study describes the form of growth and the pattern of associations among growth factors for each of the following dimensions, namely, the degree of adolescents' social isolation and the extent of exposure to delinquent peers. In addition, in order to capture the nonlinear trajectory embedded in each developmental variable, the shape factor loadings are constrained to zero and one at the first and last assessment occasions, and the coefficients for intermediate time points are freely estimated.

Unidimensional model: the GRM-LGC.

Social isolation. The results of parameter estimates and associated standard deviations from the complete data set (*n*=838) are given in Table 4.3.4 (left panel), where we see the estimated discrimination parameters for items 4 and 5 all significantly greater than the value of one, indicating that these items better discriminate the underlying person ability than the other items do. Because greater discrimination indicates a stronger relationship between an item and the underlying latent trait, we may say that the items "nobody at school cares" and "don't belong at school" are more closely related to the construct of feeling socially isolated than other items, such as "teachers don't call on me", "outsiders with family", and ''no project work from teachers". As for the item difficulty parameter estimates, the estimated item threshold parameter associated with the very last response category in item 6, $\hat{\beta}$ [6,4], is the largest, indicating that endorsing in the response category of 4 in the following item, "no project work from teachers", is the hardest alternative for respondents to reach. That is, the endorsement of this item reflects a higher level propensity to feel isolated than do the other items.

As for the substantive interpretation regarding the structural model, the empirical result shows that without controlling any explanatory variable, a mean growth curve emerges with a true initial level of 1.542 ($p<.01$) and a change rate of -.342 ($p<.01$). The significant variation between the respondents around the mean value associated with the initial level ($\hat{\sigma}_L^2$ =1.538) implies that, overall, these subjects initiate their growth process at different phases, which not only reveals systematic differences in the change trajectory among participants but also suggests true variation remaining in one of the growth parameters, indicating the need for additional time-invariant covariates (e.g., Singer and Willett, 2005). The correlation between the initial level and change rate is -.109 ($\hat{\sigma}_{LS}/\left(\hat{\sigma}_L \cdot \hat{\sigma}_S\right)$, *ns*), indicating that the initial level has no predictive power for the change rate. Finally, it was found that there exists a piecewise linear trajectory (i.e., the estimated slopes for five repeated assessments are $s1 = 0$ (fixed), $s2 = .857$ ($p<.01$), $s3 = 1.295$ ($p<.01$), $s4 = 1.230$ (fixed), and $s5 = 1$ (fixed)) in terms of the participants' perceived levels of social isolation.

Exposure to delinquent peers. Similarly, in Table 4.3.4 (right panel), we can see that the estimated discrimination parameter for item 6 is the largest out of seven, indicating that "stole something worth more than $50 dollars" is more closely related to hanging out with delinquent peers than other items. As regards the item difficulty parameter estimates, overall, the estimated threshold parameters associated with item 5 are rather large, implying that selling hard drugs is a hard item to endorse: those adolescents who endorsed higher category alternatives for this item were more likely to be associated with delinquent friends. In addition, without controlling any explanatory variable, we obtain a mean growth curve with a true initial level of -.874 ($p<.01$) and a change rate of -.519 ($p<.01$). The significant variation around the latent means for these two growth factors ($\hat{\sigma}_L^2$ =2.788 and $\hat{\sigma}_S^2$ =2.504) indicates that there remains room for individual-level covariates and contextual variables. In addition, because the initial level has no predictive power for the change rate ($\hat{\rho}_{LS}$ =.002, *ns*), the change rate demonstrates a gradual decline pattern, no matter what the respondent starting level. Likewise, a segmented latent trajectory was found (i.e., the estimated slopes for five repeated assessments are $s1 = 0$ (fixed), $s2 = .203$ ($p<.01$), $s3 = .503$ ($p<.01$), $s4 = .977$ ($p<.01$), and $s5 = 1$ (fixed)) in the dimension of deviant peer affiliation.

Multidimensional model: the MGRM-ALGC.

Unconditional model: A two-level model. The associative latent growth model allows for the assessment of relationships among individual parameters for adolescents' social isolation level and exposure extent to delinquent peers, and for the estimation of means, variances, and covariances associated with the growth factors for each developmental dimension. Gelman and Rubin's (1992) suggestion of running multiple independent chains with over-dispersed starting values for checking model convergence is adopted. The model reaches convergence: in all the Gelman-Rubin statistics, the potential scale reduction factor (PSRF) approaches one for each quantity of interest (see Figure 4.3.2). Parameter estimates indicate a significant rate of change in the development of both adolescents' social isolation and extent of exposure to delinquent peers. Being consistent with other developmental studies, generally, the results suggest a relative downward trend in these two dimensions during adolescence, except for the fourth occasion in the social isolation dimension ($s14 = 1.070$, p<.01). In addition, both variances of level and shape factors associated with each dimension are significant (i.e., 2.470,1.554;3.047,2.664), an indication that significant individual variations remain in these two developmental variables, which further justifies the implementation of a univariate LGC for each dimension, and the application of an associate LGC between two of them.

Table 4.3.5a presents the correlations between the levels and shapes for adolescents' social isolation and extent of exposure to delinquent peers. The levels and shapes associated with each dimension are all significantly correlated, except for the correlation between the change rate of social isolation and initial level of the extent of exposure to delinquent peers (.109, *ns*), and that between initial level and rate of change in the affiliation with delinquent peers (-.006, *ns*). Thus, the hypothesized associations between these two constructs are validated. That is, in terms of substantive interpretation, as adolescents perceived themselves more socially isolated, the chance that they are engaged with delinquent peers becomes profoundly larger (.292 and .523). As shown in Table 4.3.5b, the estimates for the multidimensional item discrimination and difficulty parameters estimated as fixed effects range from .571 to 1.453, and from -1.443 to 8.388, respectively.

As with any item response theory model, this MGRM-ALGC model is over-parameterized and needs to be identified. In the above analysis, the identification problem is tackled by (1) fixing the first discrimination parameter associated with each construct to the value of one, with zero loadings otherwise; (2) constraining the first threshold associated with the first item in each dimension to the value of zero; (3) imposing the level-1 residual variances for each construct to the value of one. As mentioned earlier, there are no necessary and sufficient conditions for identifiability; the problem needs to be addressed on a case-by-case basis. Thus, in what follows the other two scaling options were explored, in which compared to the identification constraints adopted in the previous analysis, in which one removes constraints from the level-1 residual variances and the first item's first threshold associated with each construct but imposes constraints on the initial latent variables (i.e., scaling option 1), while the other removes constraints from the level-1 residual variances without any concomitant changes (i.e., scaling option 2). The results were compared and contrasted with those of the previous analysis (i.e., the original scaling). As the results indicate (see Table 4.3.6), each scaling option provides convergent substantive interpretation and is equally effective in resolving the indeterminacy.

Comparison of two analytical approaches.

Additionally, in terms of the fixed and random effects, and the intermediate time coefficients from the structural model (i.e., the associative latent growth curve model, ALGC), we compare and contrast the corresponding parameter estimates using two distinct analytical approaches with a simulated data set, namely, a two-stage IRT based score analysis and a single-stage IRT based score analysis. The population values of the simulated data are adopted from the results of previous empirical data analysis, the unconditional model with the level-1 residual variances from each dimension being fixed at the value of one. The simulated data were generated using the free software of R (R Development Core Team, 2009), and the models were implemented and estimated using WinBUGS 1.4.3 (Spiegelhalter et al., 2003).

As expected, the pattern of significance from two IRT-based approaches is quite similar, except that the two-stage estimation approach fails to take into account enough uncertainty. Furthermore, the results confirm that the proposed unified model is relevant to applications such as multilevel analysis and meta-analysis, for they favor random effects models in which 'pooling strength' acts to provide more reliable inferences about individual cases (Congdon, 2005, 2006; Gelman and Hill, 2007; Luke, 2004; Raudenbush and Bryk, 2002). Unlike the conventional two-stage procedure, the simultaneous estimation of a multivariate multilevel IRT model avoids problems of attenuation bias when the study focus is to regress the latent trait variables on other explanatory covariates (e.g., Bolt and Kim, 2005).

The MIRT model used for the simultaneous estimation of multiple-domain latent growth trajectories can be viewed as a general framework for obtaining the dynamic interrelationship among multiple behavioral dimensions across the entire study span. As Adams et al. (1997) and de la Torre and Patz (2005) suggest, when dimensions are related but supposedly distinct, taking the correlation into account can lead to noticeable improvements in parameter estimates and individual measurements, in particular when there are several short subscales and the underlying dimensions are correlated. As the empirical results above indicate, employing a simultaneous estimation of multiple-domain subscales not only provides direct estimates of the relations between the latent dimensions but helps reduce the standard error of the parameter estimates of interest, in particular for parameters which present difficulties in reaching convergence in the unidimensional scenario (cf. Table 4.3.4 vs. Table 4.3.5b).

Conditional model: A Two-level model.

One of the advantages of casting IRT models in a hierarchical structure is that it enables the researcher to incorporate different contextual variables as auxiliary information while estimating the models, which not only improves the estimation of person abilities but the calibration of item parameters (Mislevy, 1987). As mentioned above, unlike the conventional two-stage procedure, the simultaneous estimation of a multivariate multilevel IRT model avoids problems of attenuation bias when the study focus is to regress the latent trait variables on other explanatory covariates (e.g., Bolt and Kim, 2005). In order to illustrate the capacity of this comprehensive modeling framework, we expand the model by adding person-level covariates. That is, building upon the previous unconditional model, we include participants' gender (0=FEMALE and 1= MALE) as the person-level predictor.

Generally, we interpret the parameters within each level in a similar way to the coefficients in regular regression. Thus, in this example, the two respective level-2 slope parameters capturing the effect of gender address the following research question: in terms of social isolation status and delinquent peer affiliation: what is the difference in the average trajectory of true change associated with participants' biological gender? Here, the final result from a parsimonious model was presented: as shown in Table 4.3.8b (right panel), the fixed effect estimates associated with the initial level of delinquent peer affiliation in the level-2 model are statistically significant (.267, $p<.05$), implying that, on average, boys have a higher initial exposure extent than their counterparts (FEMALE=0).

However, there is no gender difference associated with other latent growth parameters. In addition, the level-2 residuals, υ_{d0nk} and υ_{d1nk}, represent the portions of the individual growth parameters unexplained by the covariate of change, GENDER, for each dimension, indicating that there still remains significant between-person variability among adolescents after accounting for the effect of gender. These results again suggest the need for additional time-invariant predictors for each dimension. According to the overall goodness of fit provided via DIC, in this particular example we could not reach the conclusion that the effect of biological gender improves interpretation (76,453.4<76,462.9). That is, even though a smaller DIC represents a better fit of the model, a difference of less than ten units between models does not provide sufficient evidence for favoring one model over another (Spiegelhalter et al., 2003). Hence, these two models are considered to fit the data equally well. Recall that the multidimensional item parameters are estimated as fixed effects in the model. As shown in Table 4.3.8b, the multidimensional item difficulty estimates ranged from -1.444 to 8.435, and multidimensional item discrimination estimates ranged from .570 to 1.476.

In order to model the parallel 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. Thus, we illustrate how this unified approach allows the depiction of relations among respective growth factors, represented in both the initial level and the change rate for each of two interrelated dimensions. However, there are several ways of further extending the analyses reported here. First, the autocorrelation between identical measures across different occasions can be studied. Second, we might consider incorporating other social contextual risk and protective factors on adolescents' problem-related behaviors. From a substantive point of view, it would be beneficial to understand what factors influence specific problem behaviors and problem behaviors in general. As mentioned earlier, such information may better represent the traditional theory underpinning developmental trajectories and be useful in guiding effective intervention and prevention programs for young people. Finally, because both empirical and substantive differences may be critical for the correct interpretation of the dynamics and influences of change, as McArdle (1988) and Duncan et al. (2001) suggest, studies with a broad selection of different multivariate approaches, such as the range of models and the corresponding statistical power for detecting meaningful differences, all deserve continuous effort and exploration.

Chapter 5

DISCUSSION AND CONCLUSION

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 accordingly be 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 the estimate of each model parameter and latent variable (e.g., Curran et al., 2007; Dunson et al., 2005; Embretson, 1994; Rasch, 1960; Roberts and Ma, 2006).

In addition, as longitudinal data analysis has played a significant role in empirical research within developmental science, the researcher should bear in mind that the decision regarding the longitudinal research design can be made in an a priori manner based on a Monte Carlo study. Alternatively, the research could also consider performing a post hoc power analysis before reaching the conclusion that there is no statistical significance in a given context. Finally, when change is studied, it is common to ask whether change occurs as a result of treatment interventions or different group memberships, that is, whether the change component, such as the differences in average intercept, slope, and/or other polynomial coefficients, can be discerned and predicted by other contextual variables. Thus, researchers are encouraged to design and conduct a Monte Carlo study tailored to their specific research questions while determining the sample size at a reasonable level of power and validating their statistical inference conclusions.

In estimating complex statistical models, the capacity of Bayesian methods is undeniable, for they allow an intuitive probabilistic interpretation of the parameters of interest and the efficient incorporation of prior information to empirical data analysis (Rupp et al., 2004). Advantaged as they are by modern simulation and sampling methods, such as the Markov chain Monte Carlo (MCMC) algorithm, Bayesians allow for the representation of parameter densities which may be far from normal, whereas traditional maximum likelihood estimation relies on asymptotic normality approximations (Best et al., 1996; Maier, 2001). Unlike classical inference, the Bayesian methods treat unknown parameters as random variables and interpret traditional statistics in a more intuitive way. The consequences of taking a Bayesian point of view reflect the probability values in hypotheses and confidence intervals on parameters, both of which are more concordant with commonsense interpretations (Keller, 2005; Rice, 1995). That is, in the Bayesian paradigm, the interpretation of a Bayesian $100(1-\alpha)\%$ credible set is more straightforward than that made by the frequentists. In classical inference, the confidence interval is a probability statement about the interval, while in the Bayesian approach, the credible interval is a statement about the unknown parameter (Phillips, 2005; Rice, 1995; Wasserman, 2003).

As mentioned, MCMC sample-based estimation methods overcome numerical integration problems and allow the handling of high-dimensional problems and the exploration of the distribution of parameters, regardless of the forms of distributions of likelihood and parameters (Jackman, 2000; Keller, 2005). In addition to this advantage and that of straightforward interpretation, Bayesian methods also provide a clear approach for incorporating prior information, which increases the statistical power of the analysis and contributes to the accumulation of scientific findings. As Congdon (2005) suggests, informative subjective priors allow researchers to build on previous research and can be justified on the basis of archival materials and the weight of established evidence and opinion elicited from scientific specialists. As illustrated in one of practical illustrations, we demonstrate how informative priors affect the parameter estimates and standard deviations from a small data set and how they can be treated as extra data information while conducting an analogy analysis.

Significance of the Present Work

The ease of implementing MCMC demonstrates much potential for statistically complex models in which they can find future application. Specifically, one of the IRT-LGC derivatives, the MGRM-ALGC model presented here, provides an integrated approach to modeling development in a consecutive and simultaneous manner which includes multivariate multiple ordered categorical measures as outcomes. The MIRT model used for the simultaneous estimation of multiple-domain latent growth trajectories can be viewed as a general framework for obtaining the dynamic interrelationship among multiple behavioral dimensions across the entire study span. As Adams et al. (1997) and de la Torre and Patz (2005) suggest, when dimensions are related but supposedly distinct, taking the correlation into account can lead to noticeable improvements in parameter estimates and individual measurements, in particular when there are several short subscales and the underlying dimensions are correlated. As the empirical results above indicate, employing a simultaneous estimation of multiple-domain subscales not only provides direct estimates of the relations between the latent dimensions but helps reduce the standard error of the parameter estimates of interest, in particular for parameters which present difficulties in reaching convergence in the unidimensional scenario.

Being a flexible multivariate multilevel model, this MGRM-ALGC model produces parameter estimates which are readily estimable and interpretable. For instance, in addition to the parameter estimates for the latent trajectory of each individual, it also generates the interpretation of the items as descriptive measures for portraying the interaction between persons and items (e.g., Reckase, 1997). Substantively, this associative model helps establish the interrelationship among subjects' multiple behaviors over time and estimates the corresponding covariation in the developmental dimensions. In practice, this extension allows the researcher to evaluate the dynamic structure of both intra- and inter-individual change, rendering a rational sequence in testing the adequacy of latent growth curve representations of behavioral dynamics (Duncan et al., 1999, 2004). Methodologically, as the fusion of a number of approaches, embedding the multidimensional item response theory model into multivariate latent growth curve analysis allows one to extend the model to a multivariate second-order analysis, gives one a way to evaluate the factorial invariance of latent constructs across different assessment occasions, and permits one to separate time-specific error and measurement error (Blozis, 2007; Sayer and Cumsille, 2001).

Future Research

In the present work, the utility of this IRT-LVM comprehensive framework was investigated with two real data examples and a simulated study. Promising results were obtained, in which one data drawn from part of the British Social Attitudes Panel Survey 1983-1986 revealed the attitude to abortion of a representative sample of adults aged 18 or older living in Great Britain (see McGrath and Waterton, 1986). As a simplified illustration, we first investigated the dimensionality of the scale using confirmatory factor analysis, and assumed that there was no differential item functioning (DIF) to remove the corresponding gamma and beta changes. However, as Lord (1980) points out, because the latent ability obtained from IRT models are invariant across measures of the same construct but with different psychometric properties, the generalizability of this unified model to designs with different item samples administered on different occasions opens a promising avenue for future research. For instance, the inclusion of a set of shared anchor items over time and subsets of items altered on the basis of developmental relevance across the entire study span, namely, incomplete designs or planned missingness (e.g., Schafer and Graham, 2002), is a direction worth pursuing, for it not only expands the possibilities for linking and vertical scaling across studies and over time, but results in powerful and efficient experimental designs for the analysis of individual developmental trajectories (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).

Although assessments which measure growth over large grade spans on a common scale predate modern advances in latent trait models, as a fundamental task, it is important to conduct an up-to-date literature review and study on the classification of the different latent variable models used for examining general issues in growth modeling and vertical scaling. The taxonomy could be based on selection criteria such as model parameters and the latent variable of interest, the types of information provided via these scales, separate versus concurrent calibration, appropriate conditions for model application, etc. It is hoped that, through a systematically sound categorization, a conceptual framework can be sketched, which enables educational researchers and psychometricians to delineate the relations between different models and help them find their own models tailored to the substantive domain knowledge and available data at hands. These models include: Anderson's longitudinal model with a latent correlation (1985), Embretson's multidimensional Rasch model for learning and change (MRMLC) (1991), Adams, Wilson, and Wang's multidimensional random coefficients multinomial logit model (MRCMLM) (1997), Fischer and Seliger's multidimensional linear logistic model (1997), and Patz and Yao's multidimensional multigroup item response model for vertical scaling (2007a, 2007b), to name a few.

Moreover, it is expected that this modeling framework can be applied to large-scale assessments and facilitate the investigation of a promising practice area: analyzing students' annual growth and change across a range of grades, for example. In practice, 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, researchers could consider comparing and contrasting other estimation approaches to implementing the analysis, such as the adaptive Gauss-Hermite quadrature procedure with different options controlling the number of quadrature points used for each dimension of the integration[16], and releasing such strict assumptions as the stability of the item parameters over time and among different subpopulations, together with the assumption of local independence. For instance, in addition to the indirect effects via the latent variable, researchers could investigate whether the individual-level covariates on the responses have direct effects. That is, presuming that the scales are psychometrically sound, the phenomena of differential item functioning (DIF) can be examined, in which the DIF represents the fact that the probability of endorsing an item differs among 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 might not be fairly assessed by the instrument.

Likewise, the random effect IRT models, defining an additional random effect for each testlet and/or item bundle, can be adopted to account for dependencies between like items across different points in time (e.g., De Boeck, 2008; Li et al., 2006; Rijmen, Tuerlinckx, De Boeck and Kuppens, 2003). Additionally, in both empirical data analyses, we employed the usual single-group analysis, including subjects' demographic characteristics, such as the gender of the participants, as the time-invariant covariate (TIC). However, it is important to know that when all other parameters remain the same across different subpopulations, having TICs only introduces differences in conditional means for the growth factors. As a further point noted by Fischer and Seliger (1997), it is unrealistic to guarantee that a sufficiently unidimensional scale is applicable to all respondents: because the factor structure in different groups, such as males and females,

---

[16] Te Marvelde et al. (2006) argued that for more scales and time points, the adaptive Gauss-Hermite

black and white, etc. will generally differ. Putting this recommendation into practice implies that research should be based on multiple-group invariance analysis (Meredith and Horn, 2001). Researchers could consider the application of multiple-group growth models, such as the latent class growth models and growth mixture models, to identify homogeneous subgroups within the larger heterogeneous population (Curran et al., in press). Finally, as latent variables play an important part in this generalized linear latent and mixed modeling framework, it is desirable to develop the semiparametric Bayesian method (Lee, 2007) and other approaches (e.g., van den Oord, 2005) to relax its regular multivariate normality assumption.

---

quadrature method may become unfeasible, but this requires further investigation.

APPENDICES

APPENDIX A

Table 4.1.1
The Simulation Design Layout

Design factor	No. of participants	No. of items	Standardized effect size of the average growth trajectory
Investigating levels	125, 250	5, 10, 15	Small (.14), Medium (.28)

Table 4.1.2
The Population Values used in the RASCH-LLGC Model

Measurement model

Item difficulty parameters:
A. 5 items (-2,-1, 0, 1, 2)
B. 10 items (-2, -1.556, -1.111, -.667, -.222, .222, .667, 1.111, 1.556, 2)
C. 15 items (-2, -1.714, -1.429, -1.143, -.857, -.571, -.286, 0, .286, .571, .857, 1.143, 1.429, 1.714, 2)

Structural model

Intercept mean: 0.00

Slope mean: .14 vs. .28

Intercept variance: 1.00

Slope variance: .20

Correlation between intercept and slope: 0.00

residual variance(s): 1.00

Occasions of measurement: 0, 1, 2, 3

GCR/R-square values: .50, .55, .64, .74.

Table 4.1.3

Performance of the Estimated Average Latent Trajectory in the RASCH-LLCG Model

	True value	Estimates Average	*Range*	*BIAS*	*RMS*	*SE*	*SD*	*SE/SD*	power
SE125I05	.140	.155	[.062, .285]	.015	.054	.068	.052	1.308	.64
SE125I10	.140	.148	[.042, .260]	.008	.046	.061	.046	1.326	.72
SE125I15	.140	.141	[.092, .239]	.001	.034	.060	.032	1.875	.81
ME125I05	.280	.293	[.148, .398]	.013	.058	.069	.056	1.232	1.00
ME125I10	.280	.288	[.190, .383]	.008	.041	.062	.040	1.550	1.00
ME125I15	.280	.280	[.207, .346]	.000	.034	.060	.034	1.765	1.00
SE250I05	.140	.158	[.111, .217]	.018	.032	.047	.027	1.741	1.00
SE250I10	.140	.147	[.100, .180]	.007	.020	.043	.019	2.263	1.00
SE250I15	.140	.142	[.107, .182]	.002	.016	.042	.016	2.625	1.00
ME250I05	.280	.293	[.230, .342]	.012	.030	.048	.027	1.778	1.00
ME250I10	.280	.276	[.247, .316]	-.004	.020	.044	.019	2.316	1.00
ME250I15	.280	.279	[.228, .320]	-.001	.016	.043	.016	2.688	1.00

Note. For instance, SE250I05 stands for the condition with small standardized effect size of the average growth trajectory (.14), the sample size of 250, and five dichotomous items.

Table 4.1.4

Different Types of Prior Used for the Simulated Data Set (SE125I10)

Parameter	True value	Least informative priors	Half informative priors	Full informative priors
β_1	-2.000	N(0,.25)[a]	N(-2, 22.735)	N(-2, 45.469)
β_2	-1.556	N(0,.25)	N(-1.556, 24.902)	N(-1.556, 49.804)
β_3	-1.111	N(0,.25)	N(-1.111, 27.887)	N(-1.111, 55.775)
β_4	-.667	N(0,.25)	N(-.667, 29.495)	N(-.667, 58.990)
β_5	-.222	N(0,.25)	N(-.222, 29.450)	N(-.222, 58.899)
β_6	.222	N(0,.25)	N(.222, 29.815)	N(.222, 59.629)
β_7	.667	N(0,.25)	N(.667, 29.136)	N(.667, 58.272)
β_8	1.111	N(0,.25)	N(1.111, 28.097)	N(1.111, 56.194)
β_9	1.556	N(0,.25)	N(1.556, 23.716)	N(1.556, 47.431)
β_{10}	2.000	N(0,.25)	N(2, 21.471)	N(2, 42.943)
μ_L	0	---	---	---
μ_S	.14	N(0, .25)	N(.14, 127.836)	N(.14, 255.673)
$\begin{pmatrix}\sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_L^2\end{pmatrix}^{-1}$	$\begin{pmatrix}1 & 0 \\ 0 & .2\end{pmatrix}^{-1}$	Wishart $\left[\begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix},3\right]$	Wishart $\left[\begin{pmatrix}3.5 & 0 \\ 0 & .7\end{pmatrix},5\right]$	Wishart $\left[\begin{pmatrix}7 & 0 \\ 0 & 1.4\end{pmatrix},10\right]^{b}$
σ_ε^2	1	---	---	---

Note. a. Inside the parenthesis, the second quantity stands for the precision of the parameter.

b. First of all, let $\begin{bmatrix}1 & 0 \\ 0 & .2\end{bmatrix}$ equal the prior guess for the mean of the 2 × 2 variance/covariance matrix Σ . Second, choose the degrees-of-freedom parameter, $\nu=10$, that roughly represents an equivalent prior sample size. Third, define a matrix $S=(\nu-2-1)\times\begin{bmatrix}1 & 0 \\ 0 & .2\end{bmatrix}=\begin{bmatrix}7 & 0 \\ 0 & 1.4\end{bmatrix}$.

Table 4.1.5

Parameter Estimates with Different Priors for the Simulated Data

	True value	Simulated data set: SE125I10					
		Least informative priors		Half informative priors		Full informative priors	
		Estimate (*EAP*)	*SD*	Estimate (*EAP*)	*SD*	Estimate (*EAP*)	*SD*
β_1	-2.000	-1.87*	.155	-1.952*	.106	-1.963*	.093
β_2	-1.556	-1.547*	.148	-1.608*	.097	-1.604*	.085
β_3	-1.111	-1.000*	.140	-1.077*	.090	-1.084*	.076
β_4	-.667	-.535*	.140	-.610*	.086	-.623*	.074
β_5	-.222	-.161	.136	-.229*	.087	-.230*	.073
β_6	.222	.397*	.137	.315*	.085	.301*	.072
β_7	.667	.806*	.138	.728*	.086	.721*	.074
β_8	1.111	1.131*	.138	1.073*	.089	1.076*	.076
β_9	1.556	1.564*	.146	1.509*	.095	1.512*	.083
β_{10}	2.000	2.060*	.152	2.000*	.105	1.997*	.092
μ_L	.000	.000	---	.000	---	.000	---
μ_S	.140	.164*	.060	.151*	.047	.148*	.041
σ_L^2	1.000	.993*	.251	1.046*	.237	1.068*	.225
σ_S^2	.200	.191*	.051	.159*	.047	.164*	.044
σ_{LS}	.000	.450*	.171	.471*	.163	.390*	.150
σ_ε^2	1.000	1.000	---	1.000	---	1.000	---
DIC		3,464.580		3,462.600		3,459.440	

Note. a. *p<.05 (1.96); b. The convergence is assessed via three independent chains with 30,000 iterations each, where the first 25,000 was discarded as burn-in.

Table 4.2.1

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.

Table 4.2.2

Breakdown Table for the Restricted Data/Complete Cases

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. a. 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 4.2.3

Breakdown Table for the Full Data/Available Cases

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. a. 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 4.2.4

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 4.2.5

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		

Note. Response pattern 9 stands for the missing value.

Table 4.2.6

Different Types of Prior Used in the Present Study

Measurement model	
Parameter	Baseline priors
$\beta 2$	
$\beta 3$	N(0, 1)[a]
$\beta 4$	
$\alpha 2$	
$\alpha 3$	N(0, 1.0E-02)I(0,)
$\alpha 4$	
Structure model	
Parameter	Non informative priors
$S2$	N(0, 1.0E-4)
$S3$	
μ_L	N(0, 1.0E-4)
μ_S	
$\begin{pmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_L^2 \end{pmatrix}^{-1}$	Wishart$\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, 2\right]$
$1/\sigma_{\varepsilon t}^2$	(1) $1/\sigma_{\varepsilon t}^2$ ~Gamma(.001, .001) (2) $\sigma_{\varepsilon t}^2$ ~Unif(0,1.0E04)

Note. a. Inside the parenthesis, the second quantity stands for the precision of the parameter.

Table 4.2.7
Parameter Estimates of the 2PNO-LGC Model (Restricted Data)

	Priors input $\alpha \sim$ dnorm (0, 1.0E-02)I(0,) and $\beta\sim$ dnorm(0,1)				Priors input $\alpha \sim$ dnorm (0, 1.0E-02)I(0,) and $\beta\sim$ dnorm(0,1)			
	Probit link		Logit link[a]		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*
$\beta 1$	.000	---	.000	---	.000	---	.000	---
$\beta 2$	.201*	.071	.167*	.071	.186*	.066	.185*	.069
$\beta 3$	.223*	.070	.195*	.072	.210*	.068	.210*	.069
$\beta 4$	.636*	.071	.662*	.094	.677*	.088	.699*	.090
$\alpha 1$	1.000	---	1.000	---	1.000	---	1.000	---
$\alpha 2$	1.600*	.182	1.449*	.186	1.441*	.185	1.384*	.197
$\alpha 3$	1.514*	.165	1.319*	.155	1.304*	.161	1.256*	.161
$\alpha 4$	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^2_L	2.159*	.284	2.908*	.483	2.742*	.487	2.953*	.623
σ^2_S	.136*	.049	.144*	.040	.143*	.058	.144*	.061
ρ_{LS}	-.076	.180	-.137	.165	-.017	.191	-.021	.214
$\sigma^2_{\varepsilon 1}$	.856*	.210	1.005*	.243	1.007*	.258	1.077*	.307
$\sigma^2_{\varepsilon 2}$	.157	.206	.086	.197	.183	.287	.581	.387
$\sigma^2_{\varepsilon 3}$	.873*	.192	1.061*	.281	1.057*	.270	1.095*	.304
$\sigma^2_{\varepsilon 4}$	.071	.099	.181	.190	.170	.189	.391	.224
Index	*DIC*=3,329.41; Bayesian p=.552		*DIC*=3,370.06; Bayesian p=.488		*DIC*=3,347.52 ; Bayesian p=.513		*DIC*=3,338.53 ; Bayesian p=.494	

Note. a. Multiplying by a factor of 1.701; b.*p <.05 (1.96).

Table 4.2.8 Sensitivity Analysis: Parameter Estimates of the 2PNO-LGC Model

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*[b]	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^2_L	2.73*	.506	0.029	2.953*	.623	0.030
σ^2_S	.144*	.057	0.003	.144*	.061	0.003
ρ_{LS}	-.019	.204	0.010	-.021	.214	0.010
$\sigma^2_{\varepsilon 1}$	.996*	.265	0.012	1.077*	.307	0.013
$\sigma^2_{\varepsilon 2}$	.546	.348	0.020	.581	.387	0.019
$\sigma^2_{\varepsilon 3}$	1.016*	.275	0.013	1.095*	.304	0.012
$\sigma^2_{\varepsilon 4}$	.364	.203	0.011	.391	.224	0.010
Index	*DIC*=3,340.25; Bayesian *p*-value=.504			*DIC*=3,338.53 ; Bayesian *p*-value=.494		

(Restricted Data)

Note. a. *p <.05 (1.96); b. *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 4.2.9

Bayesian Estimates of the Model Parameters under (1) the HLM and (2) the LGC Model for a Simulated Data Set

Parameter	True value	HLM	LGC
$\beta 1$	.000	---	---
$\beta 2$	.183	.151* (.050)	.152* (.052)
$\beta 3$	.210	.252* (.055)	.254* (.055)
$\beta 4$	.728	.663* (.063)	.663* (.064)
$\alpha 1$	1.000	---	---
$\alpha 2$	1.298	1.316* (.121)	1.319* (.120)
$\alpha 3$	1.181	1.042* (.085)	1.046* (.086)
$\alpha 4$	.934	1.043* (.086)	1.045* (.086)
$S1$	.000	---	---
$S2$	-1.741	---	-1.409* (.371)
$S3$	.064	---	-.050 (.173)
$S4$	1.000	---	---
μ_L	.394	.328* (.094)	.334* (.105)
μ_S	.399	.399* (.042)	.470* (.084)
σ_L^2	3.192	3.111* (.419)	3.088* (.418)
σ_S^2	.132	.143* (.038)	.178* (.065)
ρ_{LS}	.049	.102 (.106)	.208 (.156)
σ_ε^2	1.000	.701* (.106)	.710* (.106)
DIC		5,841.250	5,847.230

Note. a. *p<.05 (1.96); b. Standard deviations are given in parentheses.

Table 4.2.10
Unconditional Models: Parameter Estimates of the 2PNO-LGC Model (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*[b]	Estimate (*EAP*)	*SD*	*mcse*
$\beta 1$	.000	---	---	.000	---	---
$\beta 2$	.185*	.069	0.002	.189*	.066	0.002
$\beta 3$	.210*	.069	0.002	.205*	.067	0.002
$\beta 4$	.699*	.090	0.004	.724*	.082	0.003
$\alpha 1$	1.000	---	---	1.000	---	---
$\alpha 2$	1.384*	.197	0.008	1.382*	.171	0.007
$\alpha 3$	1.256*	.161	0.006	1.291*	.156	0.006
$\alpha 4$	.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^2_L	2.953*	.623	0.030	2.957*	.505	0.023
σ^2_S	.144*	.061	0.003	.148*	.059	0.003
ρ_{LS}	-.021	.214	0.010	.029	.202	0.009
$\sigma^2_{\varepsilon 1}$	1.077*	.307	0.013	1.019*	.269	0.011
$\sigma^2_{\varepsilon 2}$	.581	.387	0.019	.536	.330	0.016
$\sigma^2_{\varepsilon 3}$	1.095*	.304	0.012	1.023*	.271	0.010
$\sigma^2_{\varepsilon 4}$	.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. a. *p<.05 (1.96); b. *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 4.2.11

Conditional Models: Parameter Estimates of the 2PNO-LGC Model

Parameter	Restricted data (*n*=284)			Full data (*n*=323)		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Measurement model						
$\beta 1$	.000	.000	.000	.000	.000	.000
$\beta 2$	.182* (.071)	.180* (.068)	.173* (.068)	.196* (.069)	.191* (.067)	.193* (.066)
$\beta 3$	.209* (.074)	.205* (.068)	.197* (.070)	.214* (.070)	.207* (.068)	.209* (.068)
$\beta 4$	.734* (.092)	.688* (.086)	.675* (.089)	.779* (.092)	.730* (.089)	.737* (.089)
$\alpha 1$	1.000	1.000	1.000	1.000	1.000	1.000
$\alpha 2$	1.309* (.174)	1.403* (.181)	1.417* (.188)	1.282* (.154)	1.363* (.168)	1.354* (.159)
$\alpha 3$	1.173* (.146)	1.271* (.149)	1.285* (.157)	1.2* (.144)	1.298* (.158)	1.278* (.158)
$\alpha 4$	.918* (.109)	1.005* (.113)	1.017* (.120)	.916* (.098)	1.0* (.116)	.985* (.107)
Structural model						
S1	.000	.000	.000	.000	.000	.000
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	1.000	1.000	1.000	1.000	1.000
$\beta 1.int$	-.366 (.243)	-.197 (.188)	-.180 (.180)	-.378 (.232)	-.252 (.181)	-.231 (.182)
$\beta 1.gender$	.219 (.384)			.147 (.382)		
$\beta 1.age$	.606 (.370)	.550* (.273)	.555* (.264)	.520 (.355)	.475 (.259)	.481 (.263)
$\beta 1.relig$	2.468* (.675)	1.62* (.382)	1.613* (.374)	1.882* (.606)	1.469* (.367)	1.507* (.367)
$\beta 1.gen.age$	-.0167 (.609)			-.036 (.583)		
$\beta 1.gen.rel$	-1.12 (.827)			-.488 (.773)		
$\beta 1.age.rel$	-2.122* (.797)	-1.252* (.481)	-1.253* (.473)	-1.38 (.727)	-.990* (.453)	-1.026* (.463)
$\beta 1.gen.age.rel$	1.169 (1.063)			.485 (.981)		

(continued on next page)

Table 4.2.11 (cont'd)

Parameter	Restricted data (*n*=284)			Full data (*n*=323)		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
$\beta 2.int$	.388* (.143)	.314* (.091)	.344* (.083)	.387* (.133)	.336* (.081)	.369* (.083)
$\beta 2.gender$	.514 (.267)	.181 (.126)		.394 (.233)	.159 (.113)	
$\beta 2.age$	.073 (.219)			.0818 (.194)		
$\beta 2.relig$	-.047 (.446)			.222 (.391)		
$\beta 2.gen.age$	-.346 (.377)			-.230 (.315)		
$\beta 2.gen.rel$	-.171 (.544)			-.322 (.469)		
$\beta 2.age.rel$	-.174 (.508)			-.443 (.470)		
$\beta 2.gen.age.rel$	.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 (.270)	0.107 (.232)	.037 (.213)	.209 (.236)	.126 (.228)	.067 (.210)
$\sigma_{\varepsilon 1}^2$	1.167* (.327)	0.996* (.276)	.977* (.266)	1.159* (.306)	.999* (.274)	1.014* (.258)
$\sigma_{\varepsilon 2}^2$	1.198* (.442)	.749 (.404)	.614 (.370)	1.002* (.419)	.645 (.356)	.607 (.337)
$\sigma_{\varepsilon 3}^2$	1.158* (.318)	1.047* (.293)	1.024* (.290)	1.155* (.304)	1.035* (.287)	1.056* (.276)
$\sigma_{\varepsilon 4}^2$	.419 (.258)	.407 (.225)	.4101 (.227)	.409 (.261)	.359 (.203)	.388 (.233)
Goodness of fit indices	*DIC*=3,337; Bayesian *p*=.478	*DIC*=3,340; Bayesian *p*=.489	*DIC*=3,342; Bayesian *p*=.488	*DIC*=3,638; Bayesian *p*=.48	*DIC*=3,639; Bayesian *p*=.495	*DIC*=3,639; Bayesian *p*=.494

Note. a. Each number inside the parenthesis stands for the standard deviation of the estimate.
b. *p<.05 (1.96).

Table 4.3.1a
Summary Statistics for Longitudinal NYS Data: Social Isolation

A. Summary statistics for NYS IRT scale scores over five assessment occasions

	NYS-1976	NYS-1977	NYS-1978	NYS-1979	NYS-1980
Mean	1.555	1.238	1.063	1.154	1.212
SD	1.506	1.550	1.611	1.568	1.504
Skewness	-.091	-.264	-.456	-.418	-.623
Kurtosis	.356	.048	.209	.162	.377

B. Correlation matrix for NYS IRT scale scores for five assessment occasions

	NYS-1976	NYS-1977	NYS-1978	NYS-1979	NYS-1980
NYS-1976	1				
NYS-1977	.660*	1			
NYS-1978	.608*	.740*	1		
NYS-1979	.527*	.682*	.782*	1	
NYS-1980	.533*	.692*	.730*	.780*	1

Note. a. Based on the sample of 838 participants; b. * *p*<.05 (1.96).

Table 4.3.1b
Summary Statistics for Longitudinal NYS data: Deviant Peers Affiliation

A. Summary statistics for NYS IRT scale scores over five assessment occasions

	NYS-1976	NYS-1977	NYS-1978	NYS-1979	NYS-1980
Mean	-.862	-1.007	-1.079	-1.412	-1.377
SD	1.811	1.853	1.986	2.390	2.386
Skewness	.178	.125	.068	.104	.083
Kurtosis	-.084	-.289	-.195	-.417	-.500

B. Correlation matrix for NYS IRT scale scores for five assessment occasions

	NYS-1976	NYS-1977	NYS-1978	NYS-1979	NYS-1980
NYS-1976	1				
NYS-1977	.793*	1			
NYS-1978	.763*	.818*	1		
NYS-1979	.633*	.721*	.838*	1	
NYS-1980	.641*	.724*	.834*	.906*	1

Note. a. Based on the sample of 838 participants; b. **p*<.05 (1.96).

Table 4.3.2
Response Frequencies to 13 Outcome Measures

NYS-1976: Social Isolation
(Please tell me how much you agree or disagree with these statements about you...)

	Strongly disagree	Disagree	Neither	Agree	Strongly agree
1. Don't fit in with friends	175	528	56	58	21
2. Teachers don't call on me	145	501	92	81	19
3. Outsiders with family	315	447	33	33	10
4. Nobody at school cares	210	493	64	62	9
5. Don't belong at school	205	526	53	39	15
6. No project work from teachers	126	520	90	86	16

NYS-1976: Exposure to Delinquent Peers
(Think of the people you listed as your close friends. During the last year how many of them have...)

	None	Very few	Some of them	Most of them	All of them
7. Destroyed property	522	229	68	15	4
8. Stole something worth $5 dollars or less	460	237	89	40	12
9. Hit someone	367	288	126	34	23
10. Broke into vehicle	763	56	17	1	1
11. Sold hard drugs	804	22	12	0	0
12. Stole something worth $50 dollars or more	777	43	13	1	4
13. Suggested you break the law	615	133	62	11	17

NYS-1977: Social Isolation
(Please tell me how much you agree or disagree with these statements about you...)

	Strongly disagree	Disagree	Neither	Agree	Strongly agree
1. Don't fit in with friends	214	529	46	42	7
2. Teachers don't call on me	181	500	99	52	6
3. Outsiders with family	351	420	34	26	7
4. Nobody at school cares	245	484	67	34	8
5. Don't belong at school	249	500	49	32	8
6. No project work from teachers	115	541	110	66	6

NYS-1977: Exposure to Delinquent Peers
(Think of the people you listed as your close friends. During the last year how many of them have...)

	None	Very few	Some of them	Most of them	All of them
7. Destroyed property	526	232	65	11	4
8. Stole something worth $5 dollars or less	462	235	88	40	13
9. Hit someone	434	267	100	25	12
10. Broke into vehicle	764	60	9	3	2
11. Sold hard drugs	797	30	9	0	2
12. Stole something worth $50 dollars or more	791	39	8	0	0
13. Suggested you break the law	610	141	58	20	9

(*continued on next page*)

Table 4.3.2 (cont'd)

NYS-1978: Social Isolation
(Please tell me how much you agree or disagree with these statements about you...)

	Strongly disagree	Disagree	Neither	Agree	Strongly agree
1. Don't fit in with friends	275	502	33	25	3
2. Teachers don't call on me	197	537	74	28	2
3. Outsiders with family	358	412	41	18	9
4. Nobody at school cares	263	471	66	36	2
5. Don't belong at school	247	499	54	33	5
6. No project work from teachers	116	513	133	73	3

NYS-1978: Exposure to Delinquent Peers
(Think of the people you listed as your close friends. During the last year how many of them have...)

	None	Very few	Some of them	Most of them	All of them
7. Destroyed property	528	230	61	14	5
8. Stole something worth $5 dollars or less	455	238	109	26	10
9. Hit someone	484	233	99	17	5
10. Broke into vehicle	752	70	11	4	1
11. Sold hard drugs	779	39	13	5	2
12. Stole something worth $50 dollars or more	779	43	12	2	2
13. Suggested you break the law	605	135	63	20	15

NYS-1979: Social Isolation
(Please tell me how much you agree or disagree with these statements about you...)

	Strongly disagree	Disagree	Neither	Agree	Strongly agree
1. Don't fit in with friends	259	526	30	21	2
2. Teachers don't call on me	166	590	55	25	2
3. Outsiders with family	353	422	31	22	10
4. Nobody at school cares	201	520	78	35	4
5. Don't belong at school	236	522	46	30	4
6. No project work from teachers	100	471	176	86	5

NYS-1979: Exposure to Delinquent Peers
(Think of the people you listed as your close friends. During the last year how many of them have...)

	None	Very few	Some of them	Most of them	All of them
7. Destroyed property	559	209	62	4	4
8. Stole something worth $5 dollars or less	477	228	104	18	11
9. Hit someone	527	221	67	16	7
10. Broke into vehicle	744	68	19	3	4
11. Sold hard drugs	761	50	24	3	0
12. Stole something worth $50 dollars or more	764	50	19	2	3
13. Suggested you break the law	599	139	72	13	15

(continued on next page)

Table 4.3.2 (cont'd)

NYS-1980: Social Isolation
(Please tell me how much you agree or disagree with these statements about you...)

	Strongly disagree	Disagree	Neither	Agree	Strongly agree
1. Don't fit in with friends	243	549	32	13	1
2. Teachers don't call on me	147	605	67	17	2
3. Outsiders with family	323	442	51	16	6
4. Nobody at school cares	199	541	77	18	3
5. Don't belong at school	198	545	57	34	4
6. No project work from teachers	100	477	194	63	4

NYS-1980: Exposure to Delinquent Peers
(Think of the people you listed as your close friends. During the last year how many of them have...)

	None	Very few	Some of them	Most of them	All of them
7. Destroyed property	584	185	55	7	7
8. Stole something worth $5 dollars or less	490	212	103	24	9
9. Hit someone	546	213	67	12	0
10. Broke into vehicle	742	76	18	1	1
11. Sold hard drugs	735	73	24	2	4
12. Stole something worth $50 dollars or more	747	66	21	4	0
13. Suggested you break the law	591	143	68	23	13

Note. Frequency response calculation was based on the sample of 838 participants.

Table 4.3.3

Different Types of Prior Used in the Present Study

Measurement model	
Parameter	Baseline priors
βeta	N(0, .5)[a]
alpha	N(0, 1.0E-02)I(0,)
Structure model	
Parameter	Least-informative priors
S2 *S3* *S4*	N(0, 1.0E-2)
μ_L μ_S	N(0, 1.0E-02)
Level-1 residual variances for each dimension	
$\left(\sigma^2_{\varepsilon_d}\right)^{-1}$	Gamma(1,1)

Random effect component: Unidimensional GRM-LGC

$\begin{pmatrix} \sigma^2_L & \sigma_{LS} \\ \sigma_{LS} & \sigma^2_L \end{pmatrix}^{-1}$	$\text{Wishart}\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, 3\right]$

Random effect component: Multidimensional MGRM-ALGC:

$\begin{pmatrix} \sigma^2_{1L} & \sigma_{1LS} & \sigma_{1L}\sigma_{2L} & \sigma_{1L}\sigma_{2S} \\ \sigma_{1LS} & \sigma^2_{1S} & \sigma_{1S}\sigma_{2L} & \sigma_{1S}\sigma_{2S} \\ \sigma_{1L}\sigma_{2L} & \sigma_{1S}\sigma_{2L} & \sigma^2_{2L} & \sigma_{2LS} \\ \sigma_{1L}\sigma_{2S} & \sigma_{1S}\sigma_{2S} & \sigma_{2LS} & \sigma^2_{2S} \end{pmatrix}^{-1}$	$\text{Wishart}\left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, 10\right]$

Note. a. Inside the parenthesis, the second quantity stands for the precision of the parameter.

Table 4.3.4

Unconditional Models: Parameter Estimates of the GRM-LGC Model for Each Dimension

Three independent chains (8,000 iterations, 4,000 burn-in)

	Social Isolation (*n*=838)			Deviant Peer Affiliation (*n*=838)		
	Estimate (*EAP*)	*SD*	*mcse*[b]	Estimate (*EAP*)	*SD*	*mcse*
$\beta[1,1]$	.000	---	---	.000	---	---
$\beta[1,2]$	4.358*	.083	.004	2.602*	.074	.003
$\beta[1,3]$	5.233*	.102	.005	4.753*	.138	.004
$\beta[1,4]$	7.114*	.188	.006	6.009*	.219	.005
$\beta[2,1]$	-.915*	.087	.004	-.694*	.057	.003
$\beta[2,2]$	3.943*	.115	.006	1.485*	.083	.004
$\beta[2,3]$	5.378*	.167	.009	3.297*	.138	.007
$\beta[2,4]$	7.865*	.304	.013	4.785*	.209	.010
$\beta[3,1]$	.7608*	.057	.003	-.583*	.076	.003
$\beta[3,2]$	4.503*	.137	.008	2.624*	.139	.006
$\beta[3,3]$	5.485*	.176	.009	5.488*	.253	.011
$\beta[3,4]$	6.930*	.255	.012	7.555*	.369	.015
$\beta[4,1]$	-.113	.064	.003	1.944*	.090	.005
$\beta[4,2]$	3.737*	.103	.006	3.650*	.151	.008
$\beta[4,3]$	4.952*	.146	.008	4.987*	.228	.011
$\beta[4,4]$	7.113*	.261	.011	5.667*	.298	.012
$\beta[5,1]$	.018	.060	.003	2.885*	.144	.008
$\beta[5,2]$	3.813*	.107	.006	4.407*	.218	.011
$\beta[5,3]$	4.711*	.139	.008	6.331*	.348	.016
$\beta[5,4]$	6.333*	.223	.011	7.220*	.448	.018
$\beta[6,1]$	-2.051*	.136	.007	2.177*	.097	.006
$\beta[6,2]$	3.126*	.099	.005	3.549*	.148	.008
$\beta[6,3]$	5.085*	.169	.009	4.932*	.230	.011
$\beta[6,4]$	8.855*	.364	.016	5.580*	.301	.012
$\beta[7,1]$				.721*	.078	.004
$\beta[7,2]$				2.656*	.132	.007
$\beta[7,3]$				4.485*	.207	.010
$\beta[7,4]$				5.756*	.277	.013
$a1$	1.000	---	---	1.000	---	---
$a2$	.841*	.034	.002	1.125*	.048	.002
$a3$	.995*	.043	.002	.598*	.025	.001
$a4$	1.074*	.044	.002	1.629*	.096	.004
$a5$	1.270*	.055	.003	.980*	.058	.003
$a6$	.681*	.029	.001	1.899*	.129	.006
$a7$				.793*	.035	.002

(*continued on next page*)

Table 4.3.4 (cont'd)

Three independent chains (8,000 iterations, 4,000 burn-in)						
	Social Isolation (*n*=838)			Deviant Peer Affiliation (*n*=838)		
	Estimate (*EAP*)	*SD*	*mcse*	Estimate (*EAP*)	*SD*	*mcse*
S1	.000	---	---	.000	---	---
S2	.857*	.238	.016	.203*	.060	.003
S3	1.295*	.319	.022	.503*	.063	.003
S4	1.230*	.179	.011	.977*	.077	.004
S5	1.000	---	---	1.000	---	---
μ_L	1.542*	.074	.003	-.874*	.083	.003
μ_S	-.342*	.069	.003	-.519*	.095	.004
σ_L^2	1.538*	.320	.021	2.788*	.281	.014
σ_S^2	.619	.370	.026	2.504*	.397	.021
ρ_{LS}	-.109	.252	.017	.002	.078	.003

Note. a. *p<.05 (1.96); b. Being one kind sampling error, the Monte Carlo standard error (MCSE) can always be reduced by lengthening the chain (Kim and Bolt, 2007).

Table 4.3.5a

Correlations among Adolescents' Social Isolation and Extent of Exposure to Delinquent Peers

	Social isolation		Exposure extent to delinquent peers	
	Level	*Shape*	*Level*	*Shape*
Social isolation				
Level	1			
Shape	-.387*	1		
Exposure extent to delinquent peers				
Level	.292*	.109	1	
Shape	-.203*	.523*	-.006	1

Note. a. *$p<.05$ (1.96).

Table 4.3.5b

Unconditional Models: Parameter Estimates of the MGRM-ALGC Model for Both Dimensions

	Three independent chains (8,000 iterations, 4,000 burn-in)					
	Social Isolation (*n*=838)			Deviant Peer Affiliation (*n*=838)		
	Estimate (*EAP*)	*SD*	*mcse*[b]	Estimate (*EAP*)	*SD*	*mcse*
β[1,1]	.000	---	---	.000	---	---
β[1,2]	4.603*	.085	.004	2.624*	.078	.003
β[1,3]	5.511*	.103	.005	4.787*	.139	.005
β[1,4]	7.467*	.192	.006	6.013*	.212	.006
β[2,1]	-.835*	.057	.002	-.809*	.070	.003
β[2,2]	3.162*	.088	.003	1.609*	.074	.003
β[2,3]	4.362*	.108	.004	3.636*	.110	.004
β[2,4]	6.479*	.197	.004	5.291*	.170	.004
β[3,1]	.640*	.064	.003	-.366*	.047	.002
β[3,2]	4.289*	.113	.005	1.550*	.054	.002
β[3,3]	5.256*	.134	.005	3.266*	.091	.002
β[3,4]	6.682*	.189	.006	4.509*	.147	.003
β[4,1]	-.212*	.061	.003	2.813*	.120	.006
β[4,2]	3.828*	.102	.004	5.298*	.194	.007
β[4,3]	5.113*	.126	.005	7.109*	.279	.009
β[4,4]	7.402*	.224	.006	7.979*	.356	.010
β[5,1]	-.092	.071	.003	2.651*	.090	.003
β[5,2]	4.560*	.132	.006	4.068*	.131	.004
β[5,3]	5.667*	.153	.006	5.788*	.226	.005
β[5,4]	7.660*	.226	.007	6.563*	.320	.007
β[6,1]	-1.443*	.057	.002	3.436*	.145	.007
β[6,2]	2.017*	.065	.002	5.610*	.219	.009
β[6,3]	3.345*	.082	.003	7.586*	.316	.011
β[6,4]	5.977*	.177	.003	8.388*	.394	.012
β[7,1]	---	---	---	.539*	.053	.003
β[7,2]	---	---	---	2.069*	.068	.003
β[7,3]	---	---	---	3.516*	.099	.003
β[7,4]	---	---	---	4.518*	.140	.003
a1	1.000	---	---	1.000	---	---
a2	.709*	.032	.001	1.063*	.045	.002
a3	.845*	.037	.002	.571*	.025	.001
a4	.917*	.037	.002	1.355*	.072	.003
a5	1.070*	.047	.002	.840*	.050	.002
a6	.571*	.027	.001	1.453*	.079	.004
a7	---	---	---	.752*	.033	.001

(*continued on next page*)

Table 4.3.5b (cont'd)

Three independent chains (8,000 iterations, 4,000 burn-in)

	Social Isolation (*n*=838)			Deviant Peer Affiliation (*n*=838)		
	Estimate (*EAP*)	*SD*	*mcse*	Estimate (*EAP*)	*SD*	*mcse*
S1	.000	---	---	.000	---	---
S2	.580*	.074	.003	.217*	.056	.003
S3	.925*	.088	.004	.504*	.060	.003
S4	1.070*	.077	.003	.990*	.076	.004
S5	1.000	---	---	1.000	---	---
μ_L	1.629*	.087	.004	-.932*	.086	.003
μ_S	-.416*	.074	.002	-.548*	.099	.004
σ_L^2	2.470*	.289	.015	3.047*	.297	.015
σ_S^2	1.554*	.328	.019	2.664*	.405	.021
ρ_{LS}	-.387*	.070	.003	-.006	.077	.003
Goodness of fit index	*DIC*=76,453.4					

Note. a. *p<.05 (1.96); b. Being one kind sampling error, the Monte Carlo standard error (MCSE) can always be reduced by lengthening the chain (Kim and Bolt, 2007).

Table 4.3.6

Unconditional Models: Parameter Estimates of the MGRM-ALGC Model with Different Scaling Options (Both Dimensions)

Three independent chains (8,000 iterations, 4,000 burn-in)						
	Social Isolation (*n*=838)			Deviant Peer Affiliation (*n*=838)		
	EAP estimate (*SD*)			EAP estimate (*SD*)		
	Original scaling	Scaling option 1	Scaling option 2	Original scaling	Scaling option 1	Scaling option 2
$\beta[1,1]$	.000	-1.764*	.000	.000	.451*	.000
	(fixed)	(.083)	(fixed)	(fixed)	(.078)	(fixed)
$\beta[1,2]$	4.603*	2.688*	4.398*	2.624*	3.137*	2.563*
	(.085)	(.086)	(.100)	(.078)	(.103)	(.082)
$\beta[1,3]$	5.511*	3.575*	5.280*	4.787*	5.310*	4.692*
	(.103)	(.104)	(.119)	(.139)	(.159)	(.147)
$\beta[1,4]$	7.467*	5.499*	7.187*	6.013*	6.532*	5.909*
	(.192)	(.191)	(.201)	(.212)	(.227)	(.219)
$\beta[2,1]$	-.835*	-2.128*	-.780*	-.809*	-.221*	-.832*
	(.057)	(.071)	(.058)	(.070)	(.082)	(.071)
$\beta[2,2]$	3.162*	1.892*	3.223*	1.609*	2.185*	1.585*
	(.088)	(.066)	(.089)	(.074)	(.095)	(.072)
$\beta[2,3]$	4.362*	3.101*	4.425*	3.636*	4.210*	3.610*
	(.108)	(.086)	(.109)	(.110)	(.129)	(.110)
$\beta[2,4]$	6.479*	5.237*	6.540*	5.291*	5.862*	5.260*
	(.197)	(.183)	(.195)	(.170)	(.182)	(.171)
$\beta[3,1]$	.640*	-.895*	.701*	-.366*	-.048	-.372*
	(.064)	(.066)	(.070)	(.047)	(.054)	(.047)
$\beta[3,2]$	4.289*	2.790*	4.353*	1.550*	1.871*	1.539*
	(.113)	(.086)	(.116)	(.054)	(.064)	(.054)
$\beta[3,3]$	5.256*	3.770*	5.322*	3.266*	3.590*	3.254*
	(.134)	(.108)	(.137)	(.091)	(.099)	(.090)
$\beta[3,4]$	6.682*	5.216*	6.752*	4.509*	4.834*	4.499*
	(.189)	(.167)	(.191)	(.147)	(.154)	(.148)
$\beta[4,1]$	-.212*	-1.868*	-.155*	2.813*	3.457*	2.780*
	(.061)	(.082)	(.065)	(.120)	(.149)	(.121)
$\beta[4,2]$	3.828*	2.178*	3.860*	5.298*	5.884*	5.266*
	(.102)	(.080)	(.107)	(.194)	(.213)	(.196)
$\beta[4,3]$	5.113*	3.471*	5.141*	7.109*	7.652*	7.078*
	(.126)	(.101)	(.130)	(.279)	(.292)	(.283)
$\beta[4,4]$	7.402*	5.789*	7.425*	7.979*	8.494*	7.940*
	(.224)	(.209)	(.226)	(.356)	(.362)	(.353)
$\beta[5,1]$	-.092	-2.051*	-.018	2.651*	3.075*	2.641*
	(.071)	(.090)	(.071)	(.090)	(.112)	(.091)
$\beta[5,2]$	4.560*	2.662*	4.631*	4.068*	4.481*	4.061*
	(.132)	(.098)	(.128)	(.131)	(.149)	(.132)
$\beta[5,3]$	5.667*	3.786*	5.739*	5.788*	6.183*	5.792*
	(.153)	(.118)	(.149)	(.226)	(.239)	(.233)
$\beta[5,4]$	7.660*	5.819*	7.734*	6.563*	6.929*	6.552*
	(.226)	(.200)	(.221)	(.320)	(.325)	(.321)
$\beta[6,1]$	-1.443*	-2.479*	-1.400*	3.436*	4.100*	3.422*
	(.057)	(.070)	(.057)	(.145)	(.191)	(.150)
$\beta[6,2]$	2.017*	.988*	2.063*	5.610*	6.217*	5.610*
	(.065)	(.051)	(.066)	(.219)	(.255)	(.224)

(*continued on next page*)

Table 4.3.6 (cont'd)

	Social Isolation (*n*=838)			Deviant Peer Affiliation (*n*=838)		
	EAP estimate (*SD*)			EAP estimate (*SD*)		
	Original scaling	Scaling option 1	Scaling option 2	Original scaling	Scaling option 1	Scaling option 2
$\beta[6,3]$	3.345*	2.320*	3.392*	7.586*	8.131*	7.593*
	(.082)	(.066)	(.082)	(.316)	(.336)	(.324)
$\beta[6,4]$	5.977*	4.971*	6.028*	8.388*	8.901*	8.394*
	(.177)	(.170)	(.182)	(.394)	(.405)	(.402)
$\beta[7,1]$				.539*	.952*	.522*
				(.053)	(.067)	(.052)
$\beta[7,2]$				2.069*	2.477*	2.051*
				(.068)	(.083)	(.066)
$\beta[7,3]$				3.516*	3.921*	3.498*
				(.099)	(.112)	(.098)
$\beta[7,4]$				4.518*	4.921*	4.501*
				(.140)	(.150)	(.139)
$a1$	1.000	1.000	1.000	1.000	1.000	1.000
	(fixed)	(fixed)	(fixed)	(fixed)	(fixed)	(fixed)
$a2$	.709*	.803*	.787*	1.063*	1.061*	1.106*
	(.032)	(.044)	(.043)	(.045)	(.047)	(.053)
$a3$	.845*	.961*	.933*	.571*	.577*	.590*
	(.037)	(.052)	(.050)	(.025)	(.026)	(.029)
$a4$	.917*	1.024*	.999*	1.355*	1.306*	1.408*
	(.037)	(.051)	(.049)	(.072)	(.071)	(.076)
$a5$	1.070*	1.219*	1.179*	.840*	.820*	.881*
	(.047)	(.065)	(.059)	(.050)	(.050)	(.056)
$a6$	.571*	.643*	.634*	1.453*	1.394*	1.524*
	(.027)	(.038)	(.036)	(.079)	(.081)	(.096)
$a7$	---	---	---	.752*	.750*	.785*
				(.033)	(.037)	(.039)
$S1$	.000	.000	.000	.000	.000	.000
	(fixed)	(fixed)	(fixed)	(fixed)	(fixed)	(fixed)
$S2$	.580*	.592*	.556*	.217*	.266*	.218*
	(.074)	(.067)	(.069)	(.056)	(.052)	(.056)
$S3$	.925*	.915*	.881*	.504*	.524*	.511*
	(.088)	(.081)	(.078)	(.060)	(.058)	(.061)
$S4$	1.070*	1.057*	1.044*	.990*	1.002*	1.007*
	(.077)	(.072)	(.071)	(.076)	(.071)	(.073)
$S5$	1.000	1.000	1.000	1.000	1.000	1.000
	(fixed)	(fixed)	(fixed)	(fixed)	(fixed)	(fixed)
μ_L	1.629*	.000	1.523*	-.932*	.000	-.916*
	(.087)	(fixed)	(.085)	(.086)	(fixed)	(.084)
μ_S	-.416*	-.424*	-.363*	-.548*	-.732*	-.522*
	(.074)	(.069)	(.071)	(.099)	(.097)	(.092)
σ_L^2	2.470*	2.152*	2.198*	3.047*	3.015*	2.869*
	(.289)	(.251)	(.247)	(.297)	(.300)	(.294)
σ_S^2	1.554*	1.454*	1.565*	2.664*	2.875*	2.526*
	(.328)	(.275)	(.271)	(.405)	(.435)	(.400)
ρ_{LS}	-.387*	-.429*	-.433*	-.006	-.017	-.027
	(.070)	(.192)	(.189)	(.077)	(.223)	(.218)
$\hat{\sigma}_\varepsilon^2$	1.000	.715*	.711*	1.000	.932*	.867*
	(fixed)	(.076)	(.073)	(fixed)	(.084)	(.089)

Note. a. *p<.05 (1.96).

Table 4.3.7

Results from the ALGC model Using Two Analytical Approaches with a Simulated Data Set

Three independent chains (8,000 iterations, 4,000 burn-in)

	True value	Social Isolation (*n*=838)		True value	Deviant Peer Affiliation (*n*=838)	
		2 stage IRT	1 stage IRT		2 stage IRT	1 stage IRT
		Parameter estimate (*SD*)	Parameter estimate (*SD*)		Parameter estimate (*SD*)	Parameter estimate (*SD*)
S1	.000	.000	.000	.000	.000	.000
S2	.580	.707* (.050)	.668* (.078)	.217	.176* (.040)	.214* (.059)
S3	.925	.956* (.055)	.972* (.086)	.504	.518* (.038)	.490* (.064)
S4	1.070	1.114* (.060)	1.133* (.091)	.990	1.040* (.045)	1.037* (.097)
S5	1.000	1.000	1.000	1.000	1.000	1.000
μ_L	1.629	1.486* (.064)	1.479* (.087)	-.932	-1.072* (.054)	-1.052* (.078)
μ_S	-.416	-.366* (.053)	-.384* (.070)	-.548	-.458* (.056)	-.476* (.093)
σ_L^2	2.470	2.564* (.172)	2.757* (.278)	3.047	1.859* (.121)	2.286* (.230)
σ_S^2	1.554	.987* (.135)	1.429* (.273)	2.664	1.331* (.148)	2.310* (.448)
ρ_{LS}	-.387	-.376* (.050)	-.420* (.060)	-.006	.420* (.062)	.134 (.089)

Note. a. *p<.05 (1.96).

Table 4.3.8a

Correlations among Adolescents' Social Isolation and Extent of Exposure to Delinquent Peers

	Social isolation		Exposure extent to delinquent peers	
	Level	*Shape*	*Level*	*Shape*
Social isolation				
Level	1			
Shape	-.392*	1		
Exposure extent to delinquent peers				
Level	.289*	.105	1	
Shape	-.205*	.522*	-.009	1

Note. a. *p<.05 (1.96).

Table 4.3.8b

Estimates of Fixed and Random Effect Parameters in the MGRM-ALGC Model

Three independent chains (8,000 iterations, 4,000 burn-in)

	Social Isolation (*n*=838)			Deviant Peer Affiliation (*n*=838)		
	Estimate (*EAP*)	*SD*	*mcse*[b]	Estimate (*EAP*)	*SD*	*mcse*
β[1,1]	.000	---	---	.000	---	---
β[1,2]	4.603*	.087	.004	2.612*	.074	.003
β[1,3]	5.511*	.106	.005	4.769*	.136	.004
β[1,4]	7.472*	.194	.006	5.994*	.210	.005
β[2,1]	-.836*	.056	.002	-.819*	.068	.003
β[2,2]	3.162*	.087	.003	1.599*	.072	.003
β[2,3]	4.362*	.107	.004	3.625*	.110	.004
β[2,4]	6.481*	.195	.004	5.276*	.169	.005
β[3,1]	.639*	.067	.003	-.371*	.048	.002
β[3,2]	4.289*	.114	.005	1.545*	.053	.002
β[3,3]	5.256*	.136	.005	3.263*	.091	.002
β[3,4]	6.684*	.191	.005	4.510*	.151	.003
β[4,1]	-.211*	.065	.003	2.806*	.116	.005
β[4,2]	3.829*	.108	.004	5.300*	.190	.007
β[4,3]	5.114*	.130	.005	7.116*	.273	.008
β[4,4]	7.402*	.224	.006	7.983*	.352	.009
β[5,1]	-.095	.071	.004	2.642*	.090	.004
β[5,2]	4.560*	.130	.006	4.061*	.129	.004
β[5,3]	5.667*	.150	.006	5.784*	.113	.009
β[5,4]	7.662*	.222	.007	6.558*	.318	.007
β[6,1]	-1.444*	.057	.002	3.453*	.151	.007
β[6,2]	2.013*	.066	.003	5.642*	.230	.009
β[6,3]	3.339*	.082	.003	7.627*	.331	.012
β[6,4]	5.970*	.179	.004	8.435*	.411	.013
β[7,1]				.532*	.053	.002
β[7,2]				2.062*	.066	.002
β[7,3]				3.510*	.098	.003
β[7,4]				4.512*	.138	.003
a1	1.000	---	---	1.000	---	---
a2	.710*	.031	.001	1.066*	.043	.002
a3	.845*	.036	.002	.572*	.026	.001
a4	.917*	.038	.002	1.365*	.071	.003
a5	1.070*	.044	.002	.845*	.050	.002
a6	.570*	.027	.001	1.476*	.084	.004
a7				.756*	.033	.001

(continued on next page)

Table 4.3.8b (cont'd)

Three independent chains (8,000 iterations, 4,000 burn-in)

	Social Isolation (*n*=838)			Deviant Peer Affiliation (*n*=838)		
	Estimate (*EAP*)	*SD*	*mcse*	Estimate (*EAP*)	*SD*	*mcse*
S1	.000	---	---	.000	---	---
S2	.579*	.078	.004	.214*	.059	.003
S3	.911*	.087	.005	.507*	.062	.003
S4	1.056*	.079	.004	1.000*	.075	.004
S5	1.000	---	---	1.000	---	---
v_{d00}	1.626*	.091	.004	-1.077*	.112	.004
β_{d01}	---	---	---	.267*	.134	.004
v_{d10}	-.417*	.076	.003	-.536*	.098	.004
β_{d11}	---	---	---	---	---	---
σ_L^2	2.477*	.283	.015	3.026*	.306	.016
σ_S^2	1.605*	.351	.021	2.622*	.445	.025
ρ_{LS}	-.392*	.070	.003	-.009	.082	.004
Goodness of fit index	*DIC*=76,462.9					

Note. a. *p<.05 (1.96); b. Being one kind sampling error, the Monte Carlo standard error (MCSE) can always be reduced by lengthening the chain (Kim and Bolt, 2007).

APPENDIX B

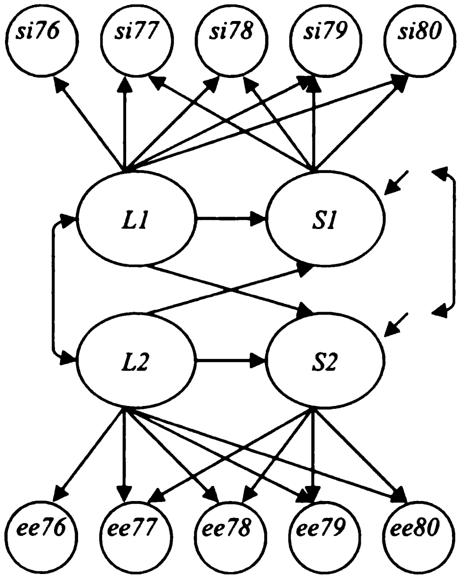

Figure 2.1

Path diagram of a bivariate latent growth model.

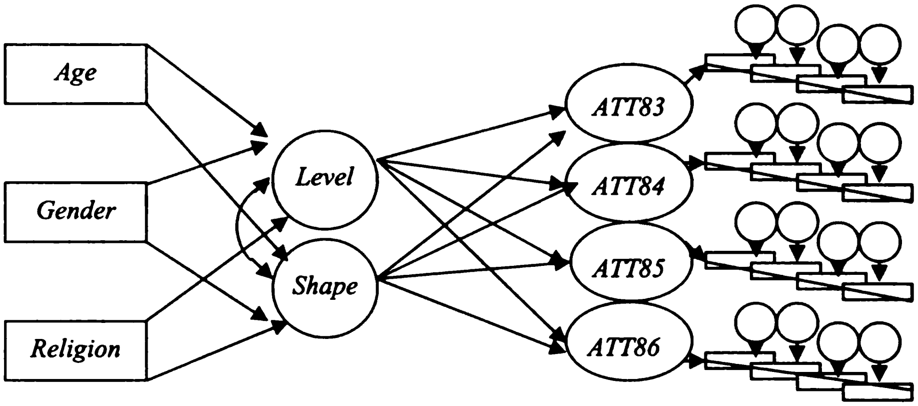

Figure 4.2.1

Path diagram of a four-wave 2PNO-LGC model.

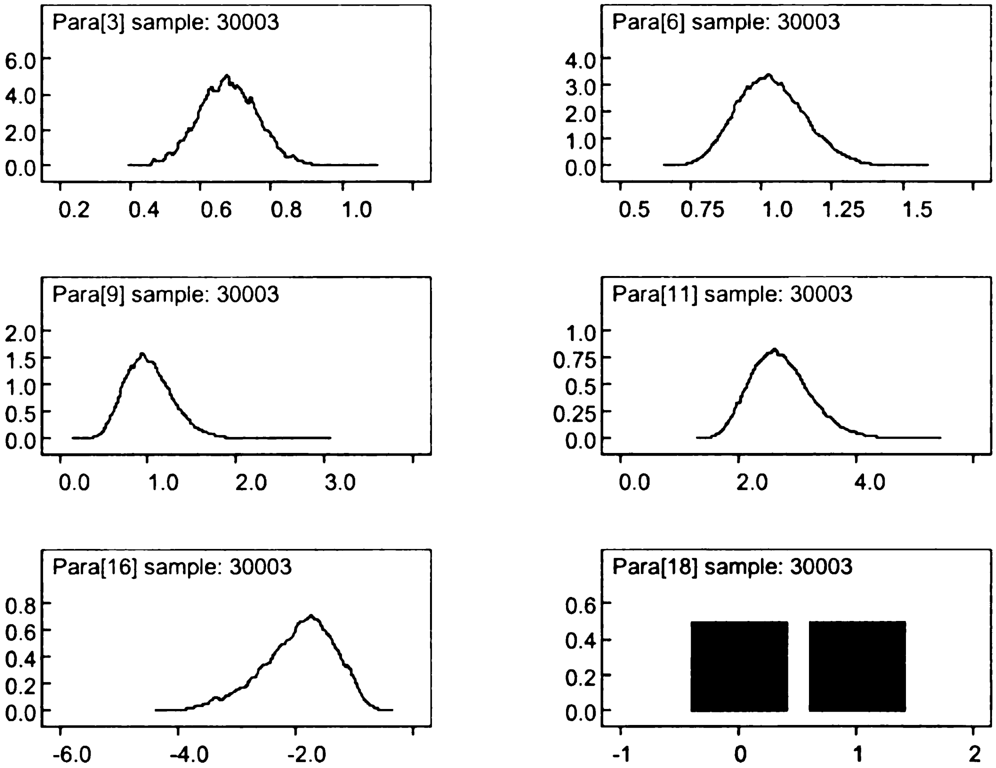

Figure 4.2.2

Kernel density for the restricted data: One single long chain (excerpted).

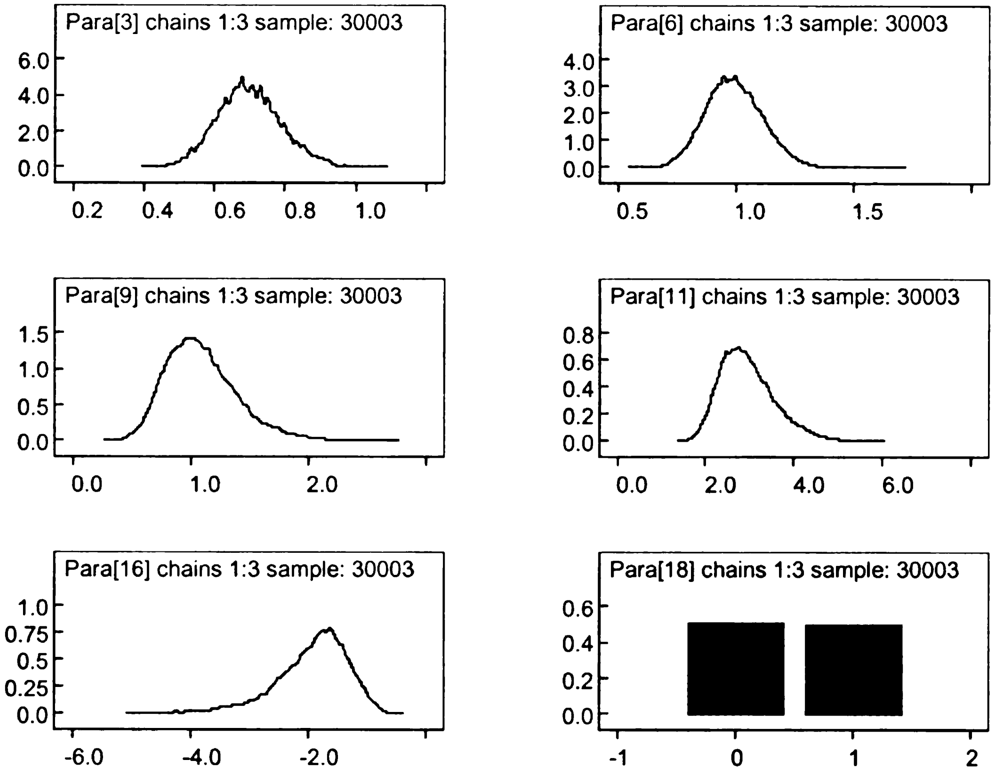

Figure 4.2.3

Kernel density for the restricted data: Three independent chains (excerpted).

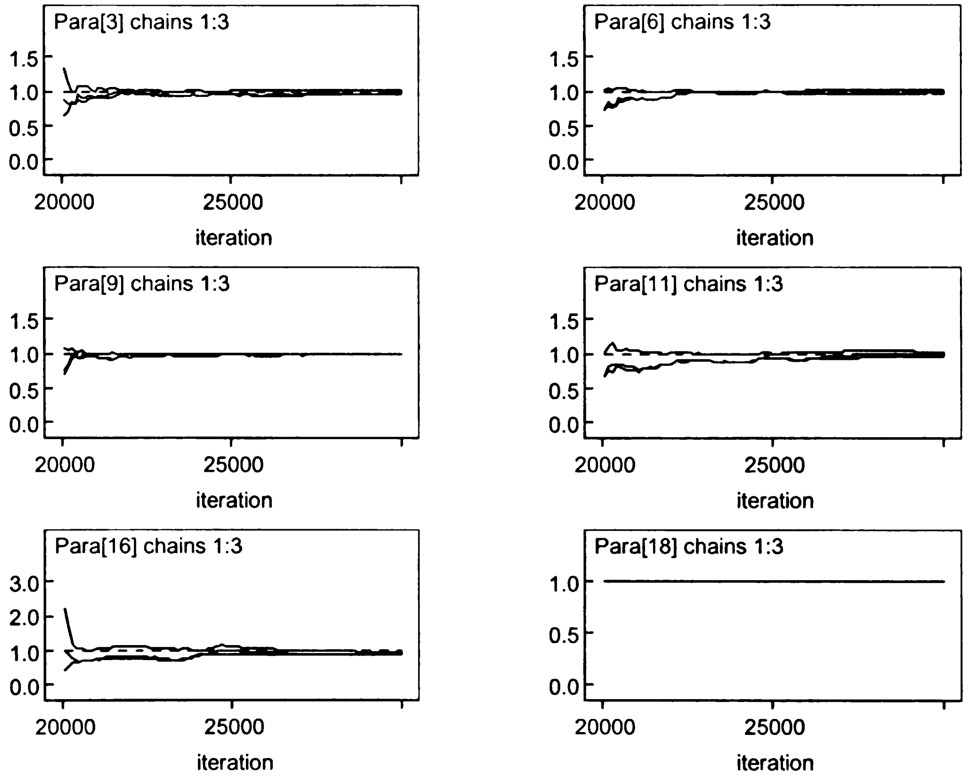

Figure 4.2.4

Gelman-Rubin statistic for the restricted dataset: Three independent chains (excerpted).

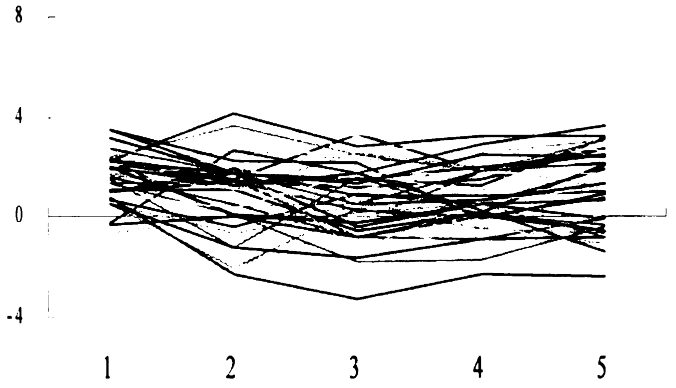

Figure 4.3.1a

Perceived social isolation across five occasions (*n*=44).

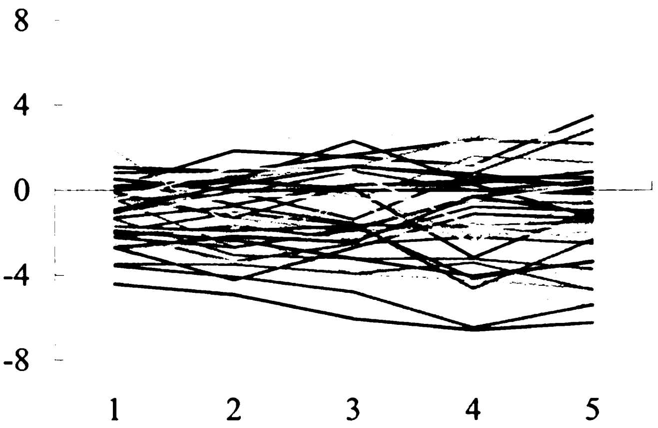

Figure 4.3.1b

Perceived extent of exposure to delinquent peers across five occasions (*n*=44).

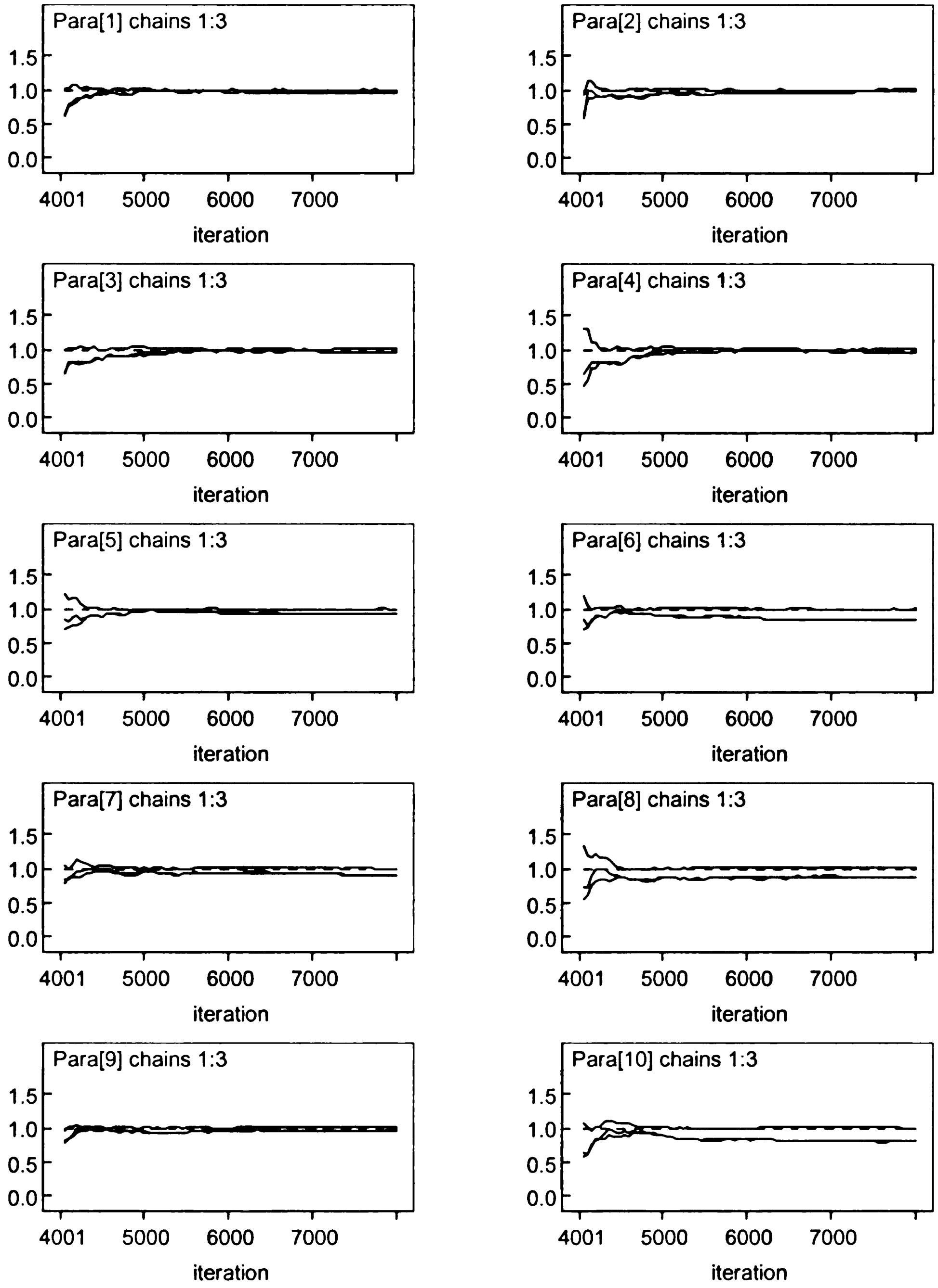

Figure 4.3.2

MCMC convergence diagnostics: Gelman and Rubin statistics.

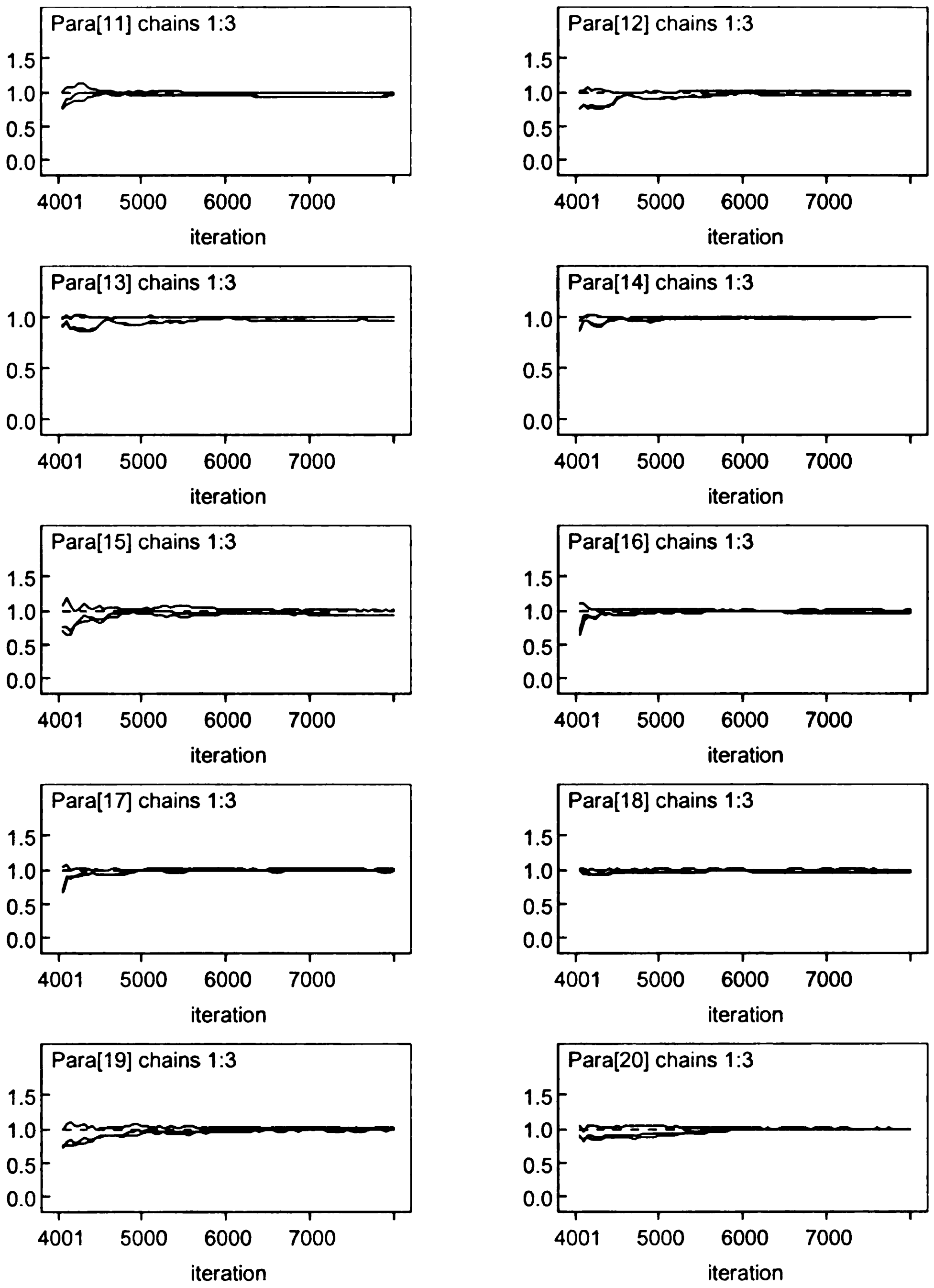

Figure 4.3.2 (cont'd)

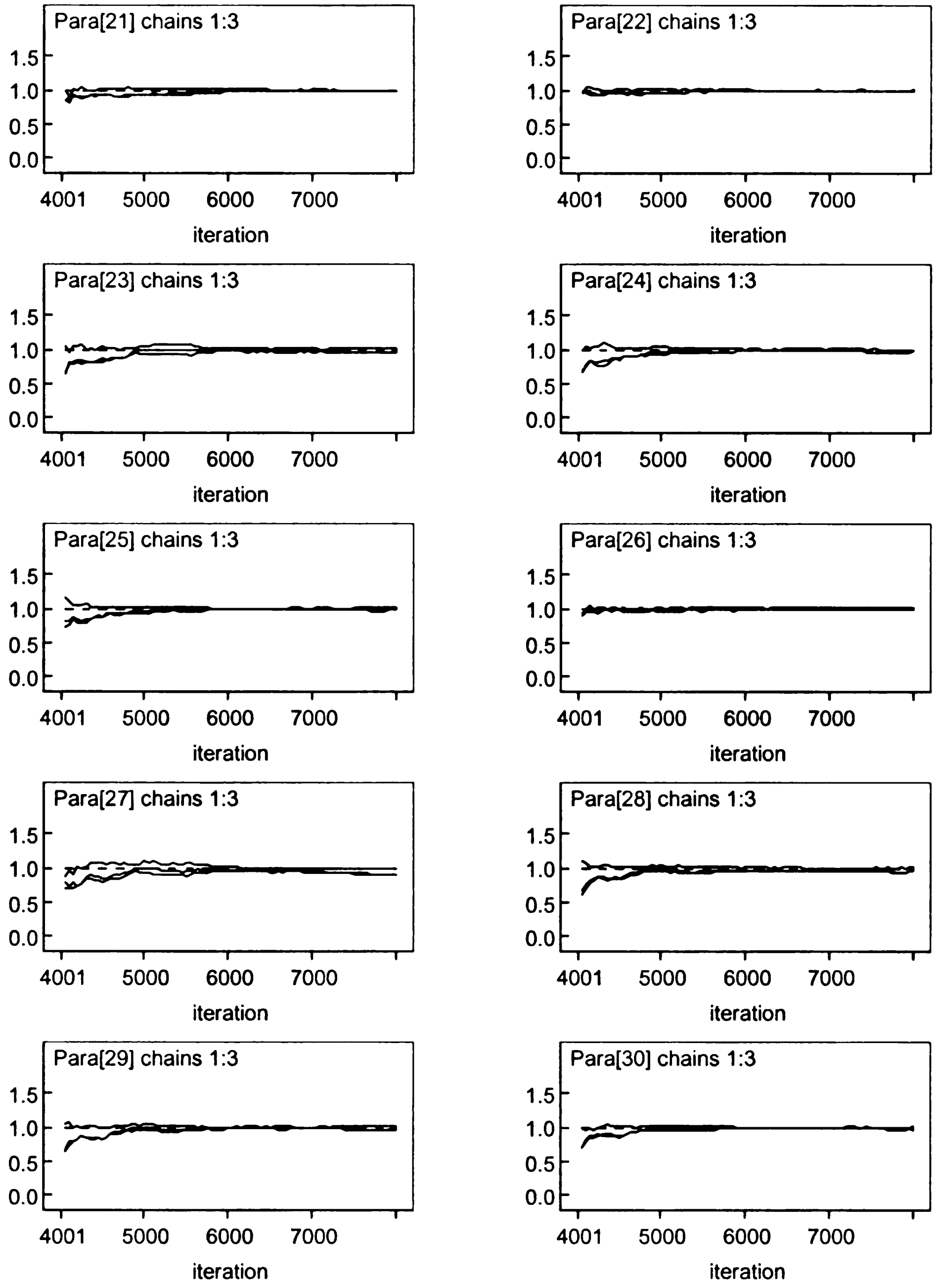

Figure 4.3.2 (cont'd)

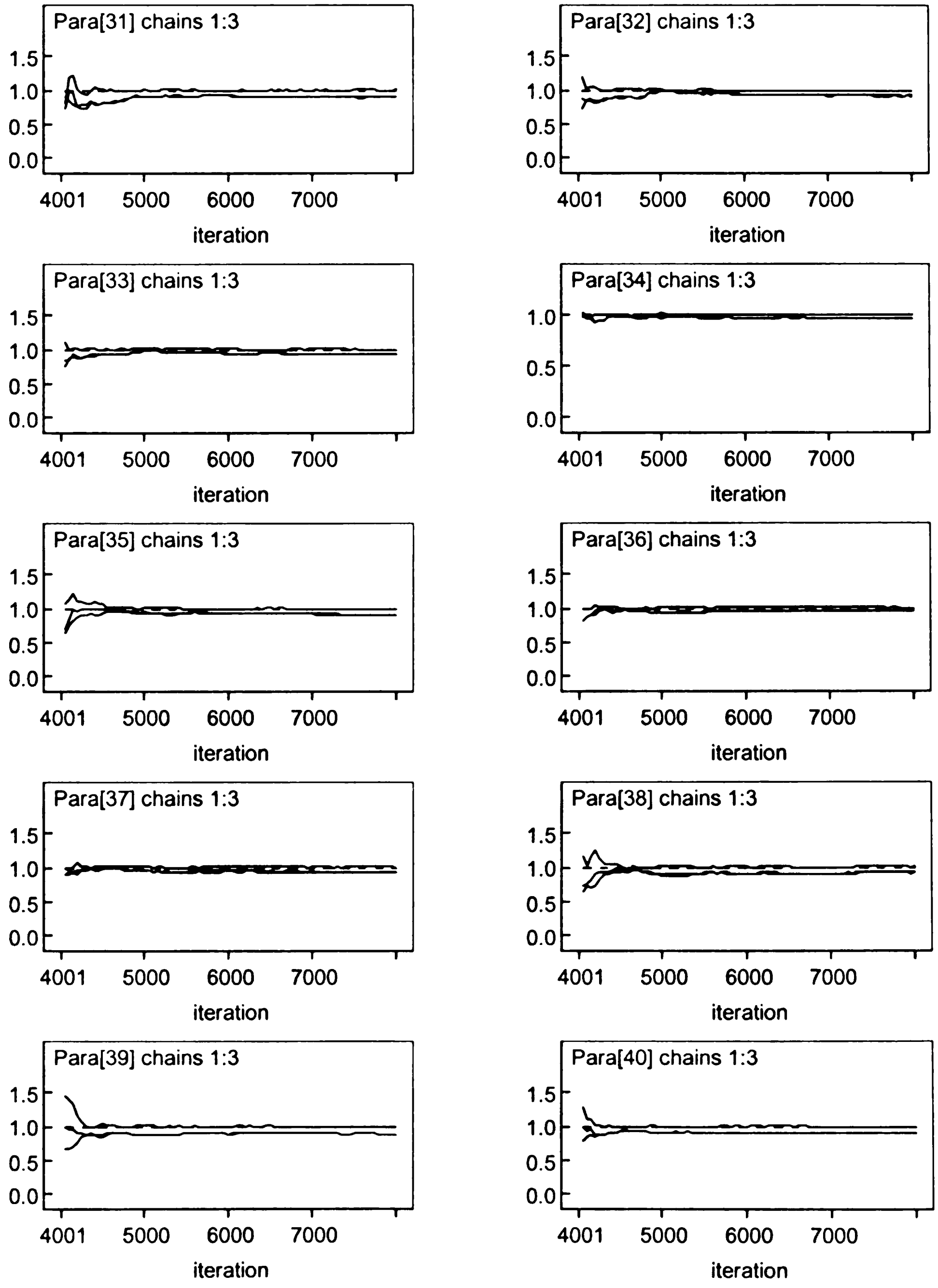

Figure 4.3.2 (cont'd)

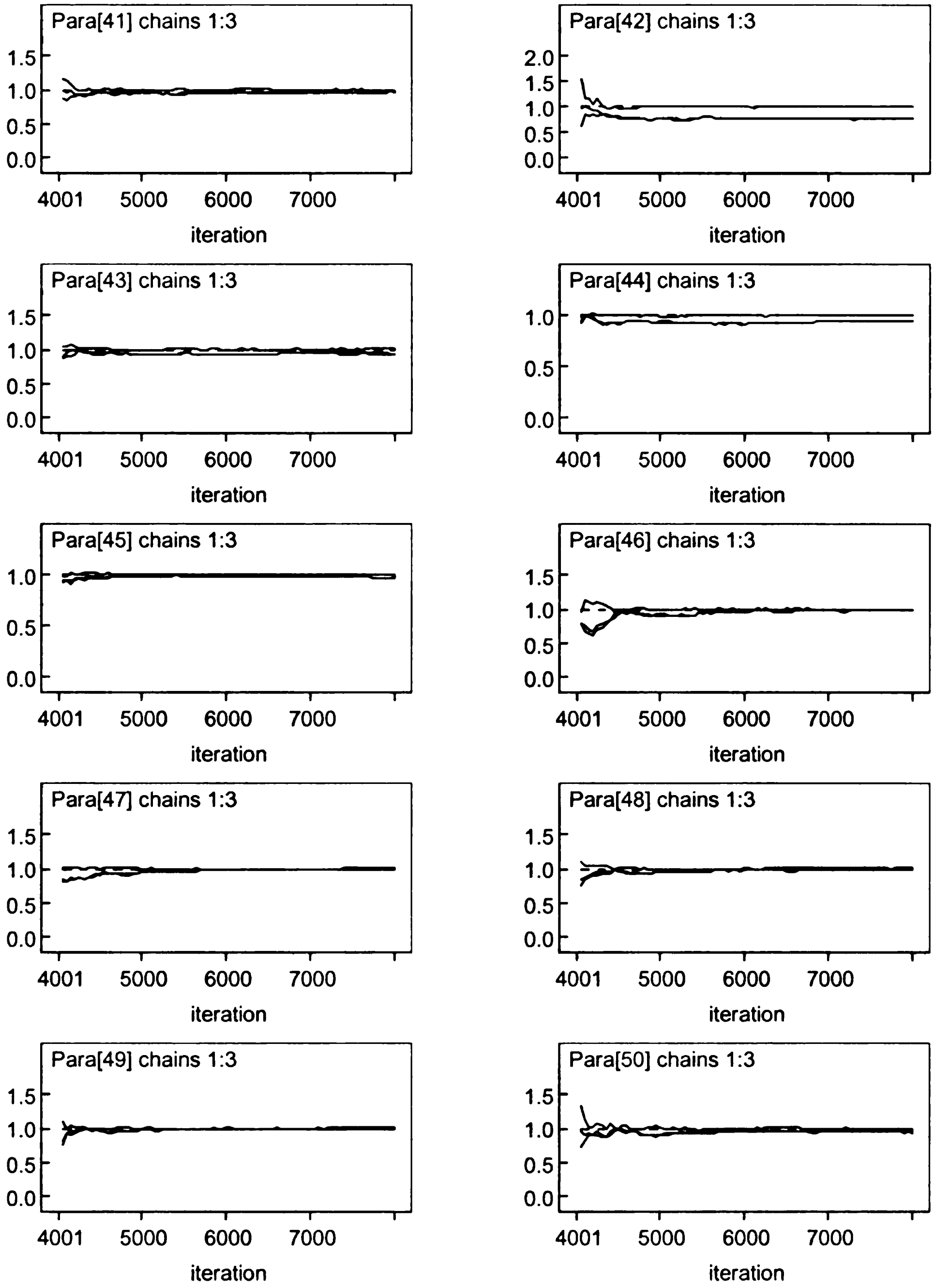

Figure 4.3.2 (cont'd)

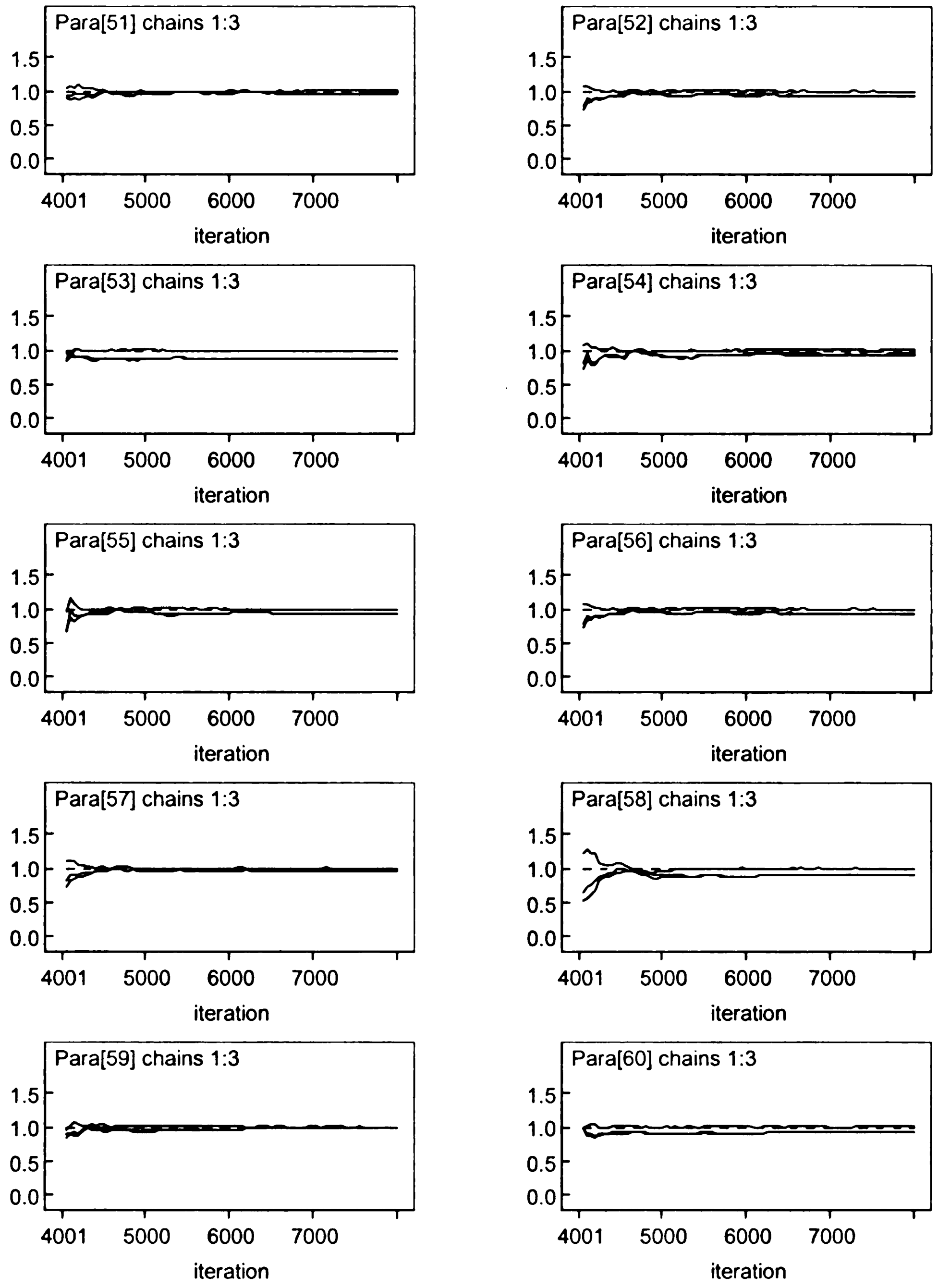

Figure 4.3.2 (cont'd)

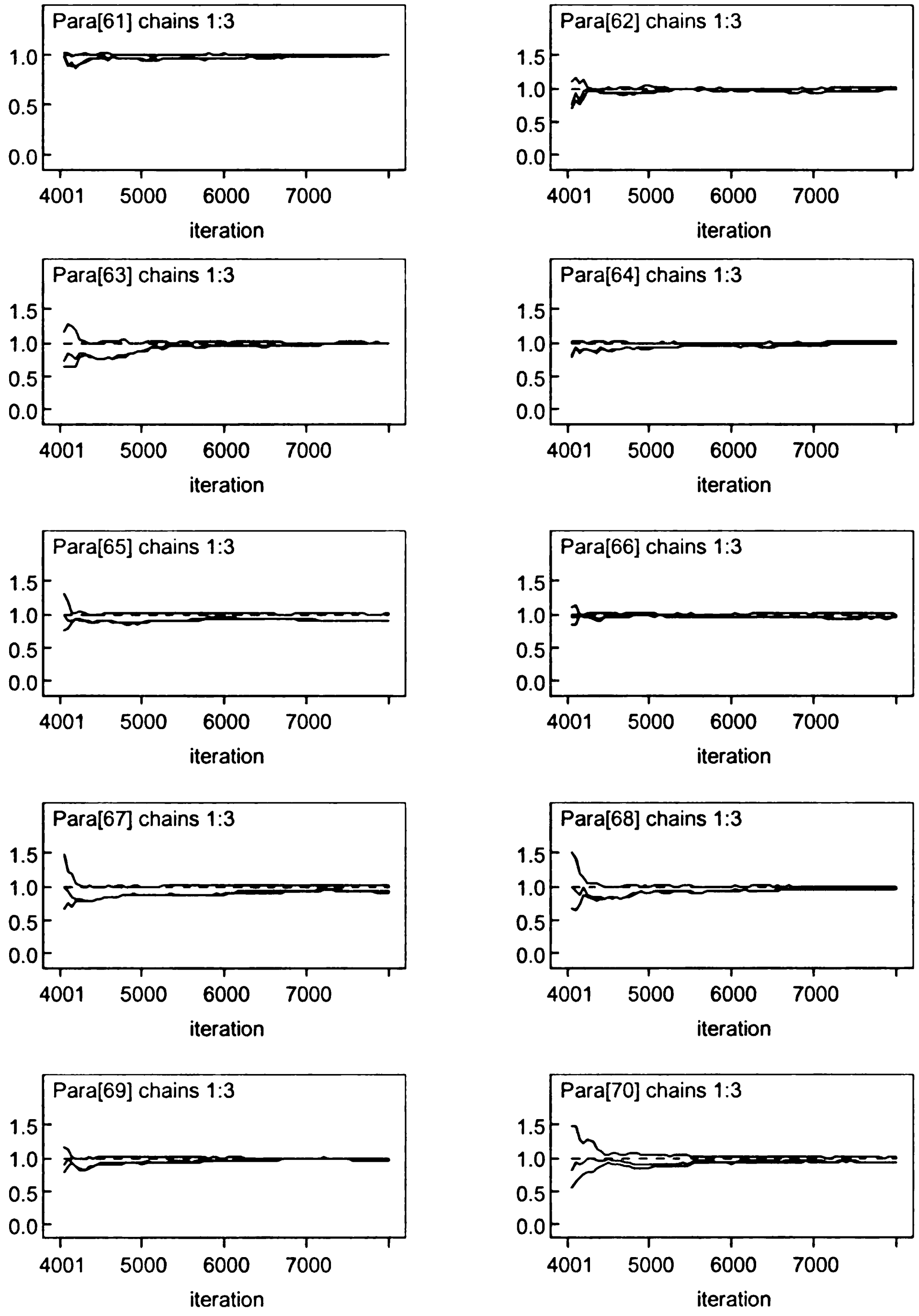

Figure 4.3.2 (cont'd)

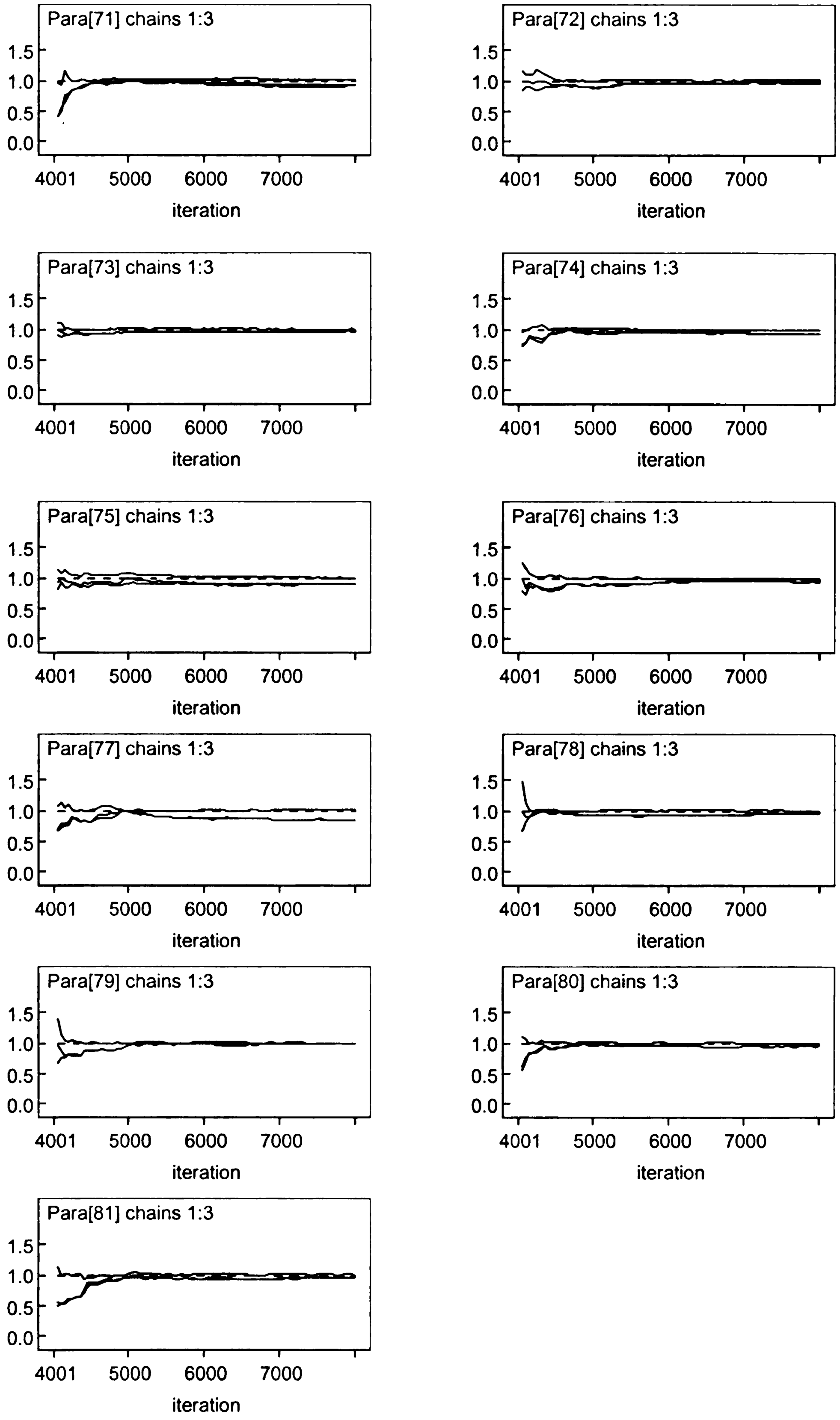

Figure 4.3.2 (cont'd)

REFERENCES

REFERENCES

Adams, J. A., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21 (1), 1-23.

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.

Anderson, E. B. (1985). Estimating latent correlations between repeated testing. *Psychometrika*, 50 (1), 3-16.

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

Bauer, D. J. (2009). A note on comparing the estimates of models for cluster-correlated or longitudinal data with binary or ordinal outcomes. *Psychometrika*, 74 (1), 97-105.

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 (Eds.), *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). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Bolt, D. M., & Kim, J.-S. (2005). Hierarchical IRT models. In B. S. Everitt & D. C.

Howell (Eds.), *Encyclopedia of statistics in behavioral science* (vol. 2, pp. 805-810). London: John Wiley & Sons.

Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement*, 27 (6), 395-414.

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.

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 (2), 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 (3), 247-266.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, E., Reinherz, H. Z., & Frost, A. K. (1994). Self-perceptions of unpopularity in adolescence: Links to past and current adjustment. *Child and Adolescent Social Work Journal*, 11 (1), 37-52.

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* (vol. 3, pp. 1134-43). London: John Wiley & Sons.

Congdon, P. (2006). *Bayesian statistical modeling*. NJ: 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). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Curran, P. J., Obeidat, K., & Losardo, D. (in press). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development.*

De Ayala, R.J. (1994). The influence of multidimensionality on the graded response model. *Applied Psychological Measurement*, 18 (2), 155-170.

De Boeck, P. (2008). Random item IRT models. *Psychometrika*, 73 (4), 533-559.

de la Torre, J. & Patz, R. J. (2005). Making the most of what we have: A practical application of multidimensional item response theory in test scoring. *Journal of Educational and Behavioral Statistics*, 30 (3), 295-311.

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

Duncan, S. C., Duncan, T. E., & Strycker, L. A. (2000). Risk and protective factors influencing adolescent problem behavior: A multivariate latent growth curve analysis. *Annals of Behavioral Medicine*, 22 (2), 103-109.

Duncan, S. C., Duncan, T. E., & Strycker, L. A. (2001). Qualitative and quantitative shifts in adolescent problem behavior development: A cohort-sequential multivariate latent growth modeling approach. *Journal of Psychopathology and Behavioral Assessment*, 23 (1), 43-50.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Duncan, T. E., Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy*, 35 (2), 333-363.

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.

Elliott, D. *National Youth Survey (NYS) Series, 1976-1987* [computer file]. ICPSR08375-06542. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2008-08-01. Retrieved 21 March, 2009, from http://www.icpsr.umich.edu/cocoon/ICPSR/SERIES/00088.xml.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change, *Psychometrika*, 56 (3), 495-515.

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.

Embretson, S. E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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). Mahwah, 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* (vol. 2, pp. 1098-1101). London: John Wiley & Sons.

Ferrer, E., & McArdle, J. J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 10 (4), 493-524.

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* (2nd Ed.). London: Chapman & Hall.

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

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science,* 7 (4), 457-472.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

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*, 53 (4), 1527-1537.

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.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice,* 12 (2), 38-47.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97-109.

Hedeker, D. (2005). Generalized linear mixed models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (vol. 2, pp. 729-738). London: John Wiley & Sons.

Hertzog, C., von Oertzen, T., Ghisletta, P., and Lindenberger (2008). Evaluating the power of latent growth curve model to detect individual differences in change. *Structural Equation Modeling: A Multidisciplinary Journal*, 15 (4), 541-563.

Holland, P. W., & Wainer, H. (1993). Differential item functioning (1st Ed.). Hillsdale, 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* (vol. 3, 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.

Hsieh, C., & von Eye, A. A. (in press). The best of both worlds: A joint modeling approach for the assessment of change across repeated measurements. *International Journal of Psychological Research.*

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.

Jamshidian, M., & Jennrich, R. I. (2000). Standard errors for EM estimation. *Journal of the Royal Statistical Society*, Series B (Statistical Methodology), 62 (2), 257-270.

Johnson, C., Raudenbush, S. W. (2006). A repeated measures, multilevel Rasch model with application to self-reported criminal behavior. In C. S. Bergeman and S. M. Boker (Eds.) *Methodological Issues in Aging Research* (pp. 131-164). Notre Dame Series on Quantitative Methods. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Johnson, M. S., Sinharay, S., & Bradlow, E. T. (2007). Hierarchical item response theory models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Psychometrics* (vol. 26, 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.

Keller, L. A. (2005). Markov chain Monte Carlo item response theory estimation. In B. S. Everitt and D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (vol. 3, pp. 1143-1148). London: John Wiley & Sons.

Kim, J-S., & Bolt, D.M. (2007). Markov chain Monte Carlo estimation of item response models. *Educational Measurement: Issues and Practice*, 26 (4), 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 (3), 662-668.

Lee, S-K. (2007). *Structural equation modeling: A Bayesian approach*. NJ: Wiley.

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 University Paper series on Quantitative Applications in the Social Sciences, 07-101. Thousand Oaks, CA: Sage.

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.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Luke, D. A. (2004). *Multilevel modeling.* Sage University Paper series on Quantitative Applications in the Social Sciences, 143. Thousand Oaks, CA: Sage.

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

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, 32 (3), 15-53.

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

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

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.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561-614). NY: Plenum Press.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

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

Meredith, W. & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In Sayer, A.G. & Collins, L.M. (Eds.), *New Methods for the Analysis of Change* (pp. 201-240). Washington, D.C.: American Psychological Association.

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

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087-1092.

Mislevy, R. J. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters. *Applied Psychological Measurement*, 11 (1), 81-91.

Moustaki, I., & Knott, M. (2000). Generalized latent trait models. *Psychometrika*, 65 (3), 391-411.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16 (2), 159-176.

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, B. O., & Curran, P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, 2, 371-402.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9 (4), 599-620.

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide (5th* Ed.). Los Angeles, CA: Muthén & Muthén.

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: Psychometrics* (vol. 26, 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.

Phillips, L. D. (2005). Bayesian statistics. In B. S. Everitt and D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (vol. 1, pp. 146-150). London: John Wiley & Sons.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling.* Sage University Paper series on Quantitative Applications in the Social Sciences, 157. Thousand Oaks, CA: Sage.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

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

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

Raudenbush, S. W. & Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6 (4), 387-401.

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 (1), 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.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9 (4), 401-412.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21 (1), 25-36.

Reckase, M. D. (2009). *Multidimensional item response theory.* NY: Springer.

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

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, 8 (2), 185-205.

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.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, no. 17, Richmond, VA: Psychometric Society.

Samejima, F. (1997). Graded response model. In W. J. van der Linden and R. K. Hambleton. (Eds.), *Handbook of modern item response theory* (pp. 85-100). NY: Springer.

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, D.C.: 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.

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64 (1), 37-52.

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* (vol. 2, 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). NY: Springer.

Snijders, T. A. B. (1996). Analysis of longitudinal data using the hierarchical linear model. *Quality and Quantity*, 30, 405-426.

Snijders, T. A. B., & Bosker, R.J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, 18, 237-259.

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, Psychometrics* (vol. 26, 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* (vol. 3, pp. 1608-1610). London: John Wiley & Sons.

Swahn, M. & Donovan, J. (2003). Correlates and predictors of violent behavior among adolescent drinkers. *Journal of Adolescent Health*, 34 (6), 480-492.

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.

Tucker, L. R. (1966). Learning theory and multivariate experiment: Illustration of determination of generalized learning curves. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 476-501). NY: Rand McNally.

Tuerlinckx, F., & Wang, W. C. (2004). Models for polytomous data. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response model: A generalized linear and nonlinear approach* (pp. 75-110). NY: Springer.

van den Oord, E. J. C. G. (2005). Estimating Johnson curve population distribution in MULTILOG. *Applied Psychological Measurement*, 29 (1), 45-64.

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). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wesserman, L. (2003). *All of statistics: A concise course in statistical inference*, NY: Springer.

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

Western, B., & Jackman, S. (1994). Bayesian inference for comparative research. *American Political Science Review*, 88 (2), 412-423.

Wiggins, R. D., Ashworth, K., & O'Muircheartaigh, C. A. (1990). Multilevel analysis of attitudes to abortion. *The Statistician*, 40 (2), 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.