SAMONTE, KELLI MARIE, Ph.D. Examining the Impact of Differential Item Functioning on Growth Models. (2017)
Directed by Dr. John T. Willse. 307 pp.

Longitudinal data analysis assumes that scales meet the assumption of longitudinal measurement invariance (i.e., that scales function equivalently across measurement occasions). This simulation study examines the impact of violations to the assumption of longitudinal measurement invariance on growth models and whether modeling the invariance violations improves the outcomes of interest. The four conditions were varied in the study: percent of non-invariant items, magnitude of invariance violation, type of invariance violation, and test length. Six latent growth models (first- and second-order) were estimated to examine the impact of invariance violations under varying degrees of model misspecification. The results suggest that the proportion of non-invariant items and the size of intercept invariance violations have the most significant impact on results. In addition, modeling the partial measurement invariance did not improve growth model parameter recovery. Ultimately, researchers should use extreme caution when estimating growth models when measurement invariance violations are present as it may lead to spurious conclusions about change over time.

EXAMINING THE IMPACT OF DIFFERENTIAL ITEM FUNCTIONING ON

GROWTH MODELS

by

Kelli Marie Samonte

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2017

Approved by

Dr. John Willse
Committee Chair

APPROVAL PAGE

This dissertation written by KELLI MARIE SAMONTE has been approved by the following committee of the Faculty of the Graduate School at The University of the North Carolina at Greensboro.

Committee Chair ______________________________
Dr. John Willse

Committee Members ______________________________
Dr. Randall Penfield

____
Dr. Ric Luecht

____
Dr. Robert Henson

____
Date of Acceptance by Committee

____
Date of Final Oral Examination

ACKNOWLEDGEMENTS

Throughout my time as a graduate student at UNCG I have had immeasurable support from a group of phenomenal individuals that deserve much more than a shout out in the acknowledgement section, but here we are. I'm thankful for my advisor, Dr. John Willse, for always making time to help me work through seemingly impossible problems and providing the perfect amount of encouragement and nudging to keep me on track. I'm grateful to my dissertation committee, Dr. Randall Penfield, Dr. Ric Luecht, and Dr. Bob Henson, for their thoughtful comments, insight, and support over the past several years.

The friends I have made during graduate school are remarkable and inspiring. I am eternally appreciative for their role in my academic development and for keeping me (mostly) sane. Jessica Flake and Allison Ames are essentially unicorns. These ladies are the most incredible friends, colleagues, and role-models I've ever met and I am lucky to have them in my life.

Finally, I am thankful for my family for providing me endless support and love. Pat Treadwell has been my biggest cheerleader and provided unwavering love and support every step of the way. His patience, love, and ability to make me laugh are unparalleled and there are no words to describe how grateful I am to have him in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Educational research is often focused on comparing student abilities between groups of individuals. Researchers may be interested in examining whether one group of students outperforms another group of students on a knowledge assessment. For example, a researcher may ask whether a class with an intensive math intervention has higher math scores than a similar class that did not receive the intervention. Longitudinal analyses are a special case of group comparison in which an individual's performance is compared to their own performance at other measurement occasions. Instead of comparing across independent groups, longitudinal analyses compare individuals' scores to their previous scores. In this sense, individuals act as their own control and each measurement occasion can be considered a different "group." Often researchers are interested in how ability (or some other construct) changes over time or when a particular event occurs. Related to the math intervention example above, a researcher may want to examine the pattern of change in math scores throughout the course of a semester and whether the change is larger when the students receive an intensive math intervention. Students' math abilities would be tested at multiple occasions throughout the semester and changes in their performances would be examined. The current study focuses on these types of longitudinal analyses.

The current chapter will begin with a brief sample of growth modeling examples in the educational research literature. This sample of studies includes a variety of models for longitudinal data and is intended to demonstrate the importance of growth models in educational research. Following the overview of longitudinal analyses in the literature, brief overviews of three common approaches to longitudinal analysis are presented in a general linear model (GLM) framework: repeated measures analysis of variance (ANOVA), hierarchical linear modeling (HLM), and structural equation modeling (SEM). This overview serves as a general introduction to the multitude of ways longitudinal data can be analyzed. Ultimately, SEM will be used for the study and a more thorough discussion of growth models in SEM is presented in chapter two. Measurement invariance is introduced as one of the necessary assumptions in growth modeling and an example of a construct that may violate the assumption of longitudinal measurement invariance is provided. Finally, the purpose and research questions for the current study are outlined.

Growth Models in Educational Research

Studies involving longitudinal data are common in educational research. A few examples of longitudinal studies in educational research are provided in the following sections. This section serves to orient the reader to the types of questions answered by longitudinal data in practice and solidify the importance of longitudinal analyses in educational research contexts. The studies are organized based on typical research questions addressed by educational researchers.

Educational researchers often examine how beliefs and attitudes change over time. Bible and Tadros (2014) examined how values change over time for business majors. The researchers were interested in determining whether ethics and values changed as students gained new experiences and more education. Results of the study suggested that values change for business majors over time. Specifically, business majors place more importance on values over time. The authors also reported differences in individual and higher order values between males and females. Jaakkola, Wang, Yli-Piipari, and Liukkonen (2015) tested changes in motivational regulations in physical education during students' transition from elementary through middle school. The authors were interested in individual and classroom level differences in motivational regulations change over time. The results suggested that some types of students' motivational regulations in physical education developed at different rates over time, whereas other motivational regulations remained stable over time. The authors note that identified regulation increased across grades 6 through 9 and is influenced by environmental factors. Amotivation increased from grade 6 to grade 7 and change was due to individual factors rather than environmental factors. King and McInerney (2014) examined how students' English and math self-concepts changed over time. The authors also investigated whether initial self-concept and changes in self-concept over time differed by gender. Results suggested that students' English self-concept increased over time, whereas math self-concept declined over time. The authors note that initial English and math self-concept differed between males and females. Males and females also differed in the extent to which their English and math self-efficacy changed over time.

Results of studies focused on student attitudes over time may help teachers plan interventions or simply feel better prepared for shifts in attitudes over time.

In addition to changes in beliefs and attitudes over time, educational researchers are frequently interested in how student ability or performance changes over time. Ryoo et al. (2014) investigated early math ability growth for high-, average-, and low-performing children in two U.S. cities and China. The results of the study suggested that the students in China had higher initial math ability scores than the students in the two U.S. cities. Results also suggest differences between the student abilities in the two U.S. cities. The authors found that change in math ability over time is non-linear and depended on the location of the student. The authors discussed potential implications for math research and education. Ouweneel, Schaufeli, and Le Blanc (2013) examined whether changes in students' self-efficacy were related to changes in student engagement and student performance. The results of the studies suggested that students' self-efficacy was related to engagement and performance. Specifically, increases in self-efficacy were related to increases in study engagement and task performance. The authors note that examining changes in self-efficacy may be an important component for investigations of student performance over time. Lamote, Pincten, Van Den Noortgate, and Van Damme (2014) conducted a study in which they explored differences in growth between students who had and had not been retained. The authors explored whether there were differences in language achievement and academic self-concept growth between the two groups. The results suggested that in the short-term (i.e., the year of retention) there was no effect on language acquisition and a positive effect on academic self-concept for individuals in the retention group. In the long-term, however, the authors observed declines in achievement for the individuals in the retention group. The authors note that academic self-concept did not change in the long-term. The results of the study may help to inform decisions about whether or not retaining students would be beneficial or harmful. Generally, results from analyses predicting performance change over time may help develop more targeted interventions and make more informed decisions in practice.

The studies presented above are a small sample of the literature examining change over time in educational contexts. Given the sizeable presence of these analyses in the literature, the types of models used to analyze the data and the assumptions made by the models are discussed in the following sections.

Types of Longitudinal Analyses

There are a multitude of approaches for analyzing longitudinal data. Methods range from observed variable approaches to latent variable approaches. In the current section, three analytic approaches are briefly described: repeated measures (ANOVA), HLM, and SEM. While item response theory (IRT) approaches do exist for growth models, they are not common in longitudinal educational research. Given the sparse use of IRT growth models in applied research, the discussion of latent variable growth models is limited to SEM approaches. The connections between IRT and SEM approaches will be briefly discussed in subsequent chapters. All models in the current section use continuous items and can be introduced under a general linear model (GLM) framework. This framework will allow for an easier transition to models that have dichotomous items. The transition to dichotomous items will be useful when IRT and

SEM connections are discussed in Chapter 3. Ultimately, all of these longitudinal models aim to examine changes in a construct across measurement occasions. The models also may explore whether change over time depends on covariates (e.g., gender).

Linear Model. One of the first models students learn in introductory statistics courses is the linear model, in which an outcome variable (the dependent variable) is modeled by some set of predictor variables (the independent variables) and an error term. The error term represents the variability in the dependent variable unaccounted for by the independent variables. The linear model is also referred to as the general linear model. One of the most common examples of the linear model is multiple regression shown in Equation 1 below.

$$y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i \tag{1}$$

Thus, the predicted value of the dependent variable for person i, y_i, is a function of an intercept, β_0, several predictor variables, X_{i1} through X_{ik} (weighted by the β_1 through β_k regression coefficients), and an error term for individual i, ε_i. The β coefficients are the relationship between each predictor variable, X, and the dependent variable, y, controlling for other variables in the model. These β coefficients are considered “fixed effects” and are assumed to be constant, or fixed, across all individuals. Thus, the relationships between the predictors and dependent variable are the same for all individuals. Fixed effects will be discussed in more detail in the section that introduces general linear mixed models. Notably, the predictors in the linear model can be either categorical or continuous, but the dependent variable must be continuous. The general linear model can also be expressed in matrix form, as shown in Equation 2 below.

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{2}$$

In Equation 2 above, **Y** represents the vector of length *i* containing the scores on the dependent variable, where *i* represents the number of individuals. The **X** parameter represents an *i* by *k* matrix of observed values for the independent variables, where *i* is the number of individuals and *k* is the number of independent variables plus one (to include the intercept). Note that the first column in the **X** matrix is equal to one for every individual to represent the intercept. The **B** parameter is a vector of length *k* including the regression coefficients. The **E** parameter represents a vector of length *i* containing the error for each individual. The GLM assumes that errors are uncorrelated with one another (i.e., observations are independent) and follow a multivariate normal distribution.

Repeated Measures Analysis of Variance. The GLM can be extended to a general linear mixed model (GLMM) which includes both fixed and random effects. Fixed factors are variables (predictors) that are assumed to include all levels of interest. Fixed factors are often the main interest in the model and are chosen to make contrasts or represent conditions within a study (West, Welch, & Galecki, 2006). Fixed effects are the relationships between fixed factors and the dependent variables. Fixed effects are considered constant quantities, suggesting that the relationships between fixed factors and dependent variables are the same across all individuals within the sample. Random factors are variables that only include a sample from a larger population of levels of interest, and are used to generalize to the entire population. Random effects are included to examine variation in the dependent variable across levels of the random factor and can be included to account for nesting within the data (West et al., 2006). Random effects are the relationships between random factors and the dependent variable. Longitudinal analyses contain random effects to account for the fact that the assumption of independent observations is violated. Longitudinal data cannot meet the assumption of independent observations because it contains multiple records from the same individual. Consequently, some records (i.e., two from the same person) are more related than others (i.e., two records from two different individuals).

Often, repeated measures ANOVA is discussed in terms of partitioned variance, sums of squares, and F-statistics. The model can also be discussed as a general linear mixed model with categorical predictors (in this case, measurement occasions are the predictors), as will be the case in the current study. Repeated measures ANOVA is a type of GLMM in which the initial measurement (intercept) is random and the change over time is fixed.

To demonstrate repeated measures ANOVA as a general linear mixed model, consider a scenario in which we were interested in student motivation throughout a semester. In this scenario, we measure student motivation at three measurement occasions. Equation 3 below can be used to model the repeated measures data.

$$y_{ti} = \beta_0 + \beta_1 t1_{ti} + \beta_2 t2_{ti} + u_{0i} + e_{ti} \tag{3}$$

In the equation above, y_{ti} represents individual *i*'s predicted motivation score at time *t*. The $t1_{ti}$ and $t2_{ti}$ parameters are dummy codes for the three measurement occasions. The β_0 parameter represents the intercept. The intercept is the typical score across all individuals for the measurement occasion that is coded as the reference time point (i.e., the dummy codes are both zero). In longitudinal contexts, it may be useful to code the reference time point as the initial measurement occasion, thus making the intercept the average score at the initial time point. The β_1 and β_2 parameters are the typical score across all students at the two dummy coded measurement occasions. If the initial time point is coded as the reference time point, these values represent the second and third time points, respectively. The u_{0i} parameter is the random effect, or the "person effect," and indicates the extent to which individual *i*'s average score deviates from the overall average score. Including this parameter is what makes the equation a "repeated measures" analysis. Without the person effect, researchers are assuming that each observation is independent and an independent samples approach (i.e., traditional ANOVA) would be appropriate. The e_{ti} parameter is the difference between individual *i*'s observed and predicted scores (including the person effect) at time *t*.

In sum, an individual's score is best predicted by the average score at each measurement occasion (β_0, β_1, and β_2), the individual's disposition (u_{0i}), and other random sources of error (e_{ti}). A more general parameterization of the GLMM is presented in Equation 4 below.

$$\mathbf{Y} = \mathbf{XB} + \mathbf{ZU} + \mathbf{E} \tag{4}$$

In the equation above, **Y** is a vector of length *i* containing the scores on the dependent variable. The **X** matrix is an *i* by *p* design matrix for the fixed effects, where *p* represents the number of fixed effects. If an intercept is modeled, the number of fixed effects, *p*, would be the number of independent variables plus one. The **B** parameter is a vector of length *p* containing the estimated parameters for the fixed effects. The **Z** matrix is an *i* by *r* design matrix for the random effects in the model, where *r* represents the number of random effects. If the intercept is modeled and all variables are considered random, the number of random effects, *r*, is the number of independent variables plus one. The **U** parameter is a vector of length *r* that includes the estimated effects of the random effects. The **E** parameter is a vector of length *i* containing the residuals (or error) for each individual. Thus, the β parameters from Equation 3 are contained in the **B** matrix, the *t* design codes are contained in the **X** matrix, the *u* parameter is contained in the **U** matrix, and the *e* parameter is contained in the **E** matrix. Repeated measures ANOVA is considered a "random-intercepts" model, therefore the **Z** matrix is simply a vector of ones (i.e., only the intercept randomly varies across individuals). Notably, the mixed model approach to repeated measures ANOVA allows for more relaxed assumptions (i.e., does not require sphericity), but the model can be constrained to equal the results of a traditional repeated measures ANOVA analysis.

Hierarchical Linear Modeling. HLM is typically used with data that violates the assumption of independent observations (i.e., nested data). Often nested data is discussed in terms of "levels" of nesting. With longitudinal data, measurement occasions (level one) are considered nested within individuals (level two) (Singer & Willett, 2003). The first level contains information about individuals' predicted scores at each measurement occasion. The parameters at level one include an individual intercept and an individual slope as well as within-person variation (i.e., error). The intercept represents the individual's score when time equals zero (typically coded as the first measurement occasion). The slope represents the change in an individual's score for each unit increase in time. Within-person variation refers to the variability of individuals' scores around their own predicted trajectory. The second level contains information about the predicted scores across the entire sample (intercept and slope) as well as information regarding between-person variability. The intercept at the second level describes the typical score, across all individuals, when time is equal to zero (typically coded as the first measurement occasion). The slope at level two describes the typical change, across all individuals, for each unit increase in time. The between-person variability parameters describe the variability in intercepts and slopes around the overall intercept and slope, respectively.

HLM is another form of the general linear mixed model. HLM is often expressed as a hierarchical general linear model (HGLM), and although the literature tends to discuss the HGLM and the GLMM separately, they are incredibly similar. The difference between the HGLM and GLMM is that the HGLM models nested data with several simpler equations using different "levels" corresponding to each level of nesting (Setzer, 2008). Ultimately, the HGLM and the GLMM have the same capabilities, but the HGLM may make it more straightforward to include predictors into the different levels. One possible two-level model for growth (specified as an HGLM) is provided in Equations 5 and 6 below.

$$y_{ti} = \pi_{0i} + \pi_{1i}t_{ti} + e_{ti} \tag{5}$$

$$\begin{aligned} \pi_{0i} &= \beta_{00} + u_{0i} \\ \pi_{1i} &= \beta_{10} + u_{1i} \end{aligned} \tag{6}$$

Equation 5 is the level one model. The value of y_{ti} is the predicted score for individual i at time t, π_{0i} is the individual's intercept. The intercept represents an individual's score when time is equal to zero. Notably, the coding of time can change the interpretation of the intercept value. For example, if the initial time point is set to 0, then the intercept would equal an individual's score at the initial time point. The π_{1i} term is the slope for each individual. The slope indicates the amount of change in in dependent variable for each unit change in time. Again, the coding of time (e.g., days, months, years) can alter the interpretation of the slope parameter. The value of e_{ti} is the residual, or the deviation of each individual's observed score from their predicted score. The residual indicates the amount of the within-person variation left unexplained by time. If there is a sizeable amount of unexplained variance, other models that include additional time-varying predictors may help to explain the remaining variation.

The second level of the model is presented in Equation 6. Note that the dependent variables at the second level are the intercept and slope parameters from level one. The β_{00} parameter is the overall intercept and represents the average initial score (assuming the initial score is coded as zero) across all individuals. The residual for this equation, u_{0j}, indicates how much an individual's intercept deviates from the overall intercept. The between-person variation in intercepts can also be estimated and is often represented by τ_{00}. The variation in intercepts describes how similarly (or differently) individuals score at the initial time point. The second equation estimates an overall slope, which indicates, on average, the extent to which participants' scores change per unit change in time. The residual for this equation, u_{1j}, indicates how much an individual's slope differs from the average slope. The between-person variation in slopes can also be estimated and is often represented by τ_{11}. The variation in slopes describes how similarly (or differently) individuals change over time. Notably, this parameter is not modeled in repeated measures ANOVA. HLM allows for more interpretive power compared to repeated measured ANOVA with regard to how individual's vary in their change over time. In addition, the covariance between the intercept and slope parameters can be estimated and is often represented by τ_{10}. This covariance describes the extent to which an individual's score at the initial time point is related to how the individual changes over time. This parameter is also not estimated in repeated measures ANOVA.

Several variations of the model presented in Equations 5 and 6 can be made depending on the scenario and underlying theory. Predictors can be added to either level to explain variability in scores above and beyond what can be explained by time. At the first level, predictors would be those that vary over the measurement occasions. For example, researchers may want to model measurement occasions in which students were and were not receiving an intervention. Including the intervention as a time-varying predictor allows researchers to examine whether the predicted scores differed between intervention measurement occasions and control measurement occasions. Predictors at the second level help to explain between-person variability in intercepts and slopes. For example, researchers may be interested in whether intercepts and/or slopes differ between males and females, treatment conditions, etc. Another variation on the model presented in Equations 5 and 6 would be to constrain the random effects. For example, the residual terms for the intercept and/or slope equations at the second level could be constrained to be zero. Constraining the intercept residual to be zero postulates that every individual has the exact same score at the initial time point. A constrained slope parameter would suggest that all individuals change the same way over time. Notably, constraining the random effect for the slope parameter to be zero would result in a model resembling repeated measures ANOVA. The next section delineates connections between HLM and repeated measures ANOVA.

Connections Between HLM and Repeated Measures ANOVA. As previously noted, the slope random effect in Equation 6 can be constrained to zero, as shown in Equation 7 below.

$$\begin{aligned} \pi_{0i} &= \beta_{00} + u_{0i} \\ \pi_{1i} &= \beta_{10} \end{aligned} \tag{7}$$

Note that Equation 5 and Equation 7 can be written as one equation by substituting the level 2 (Equation 7) equations into the level 1 (Equation 5) equation. Notably, this allows the HLM model previously expressed as an HGLM to be expressed as a GLMM. The resulting equation (with some slight reorganization) is shown below.

$$y_{ti} = \beta_{00} + \beta_{10}t_{ti} + u_{0i} + e_{ti} \tag{8}$$

Note that Equation 8 looks very similar to the one for repeated measures ANOVA shown in Equation 3. The distinction between the model presented in Equation 8 and the model in Equation 3 is that the time variable (t) is treated as continuous in Equation 8 and as categorical in Equation 3. The β_{00} parameter represents the overall score on the dependent variable when time is equal to zero. This HLM parameter is analogous to the β_0 parameter in repeated measures ANOVA when the first measurement occasion is coded as the reference measurement occasion. The u_{0i} parameter represents the random effect for intercepts in both models. The variability of the u_{0i} parameter provides an estimate of between person variability. Omitting the u_{0i} parameter suggests that there is no "person effect" that needs to be included in the model and that each observation is independent (i.e., one-way ANOVA). The e_{ti} parameter in both equations represents the individual residual. The variance of this term represents the within person variability. The general parameterization of the HLM for modeling change over time looks identical to Equation 4.

Structural Equation Modeling. The specifications for latent growth models (LGM) in the SEM approach mirror the specifications in the HLM approach described above. As with HLM, the LGM can be expressed as an HGLM or a GLMM. In this context, I will present it as an HGLM to help make clear connections to HLM and to allow for clearer transition to second-order LGMs in Chapter 3. In SEM, change over time is often characterized by two factors: initial status and slope. Thus, an individual's observed score can be represented based on the two latent variables and error, as shown in Equation 9 below.

$$y_{ti} = \beta_0 \eta_{0i} + \beta_1 \eta_{1i} + e_{ti} \tag{9}$$

In the equation above, y_{ti} represents the observed score for individual i at time t. The individual estimates for the latent variables, initial status and slope, are represented by η_{0i} and η_{1i}, respectively. The β values are fixed to constants. For $\beta_{0,}$ all values are fixed to one to represent the intercept. For $\beta_{1,}$ the values are fixed to constants that represent changes in time. For example, with 4 measurement occasions the values might be specified to 0, 1, 2, and 3 to represent the initial status (time 0), time 1, time 2, and time 3, respectively. This approach to coding time assumes that all individuals have the same data collection schedule and that measurement occasions are equally spaced. The estimated values on the latent variable for each individual can be represented by an overall mean and each individual's deviation from the overall mean, as shown in Equations 10 below.

$$\begin{aligned} \eta_{0i} &= \alpha_0 + \zeta_{0i} \\ \eta_{1i} &= \alpha_1 + \zeta_{1i} \end{aligned} \tag{10}$$

In the equations above, the α_0 and α_1 values represent the overall means for the initial status and slope latent variables, respectively. Thus, α_0 describes the typical score for individuals when time is equal to zero (typically the initial measurement occasion). The α_1 parameter represents the typical change in score per unit increase in time. The ζ_{0i}

parameter represents the difference between an individual's score at the initial status and the typical initial status across all individuals. The ζ_{1i} parameter represents the difference between an individual's slope and the overall slope.

Connections Between SEM and HLM. The parameters in the SEM and HLM approaches are parallel. Note that Equation 5 and Equation 9 look very similar. Both equations predict the observed scores for an individual with an intercept, or initial status, and slope. The intercept and slope (π_{0i} and π_{1i}) parameters in Equation 5 are analogous to the intercept and slope (η_{0i} and η_{1i}) parameters in Equation 9. The t_{ti} parameter in Equation 5 corresponds to the β_1 parameter in Equation 9. Both parameters describe time in the models. Note that the β_0 parameter in Equation 9 is a vector of ones. While the parameter is not explicitly stated in Equation 5, it could easily be included (it would simply multiply the π_{0i} parameter by one). Equation 6 and Equation 10 are also analogous. In both approaches the individual intercept and slope parameters at the first level are modeled at the second level by an overall estimate of the intercept and slope across all individuals and an individual residual. The β_{00} parameter from Equation 6 and the α_0 parameter from Equation 10 both represent the overall intercept across all individuals. The u_{0i} and ζ_{0i} parameters from Equation 6 and Equation 10, respectively, represent the difference between an individual's intercept and the overall intercept. The β_{10} parameter from Equation 6 and the α_1 parameter from Equation 10 both represent the typical slope across all individuals. The u_{1i} and ζ_{1i} parameters from Equation 6 and Equation 10, respectively, represent the difference between an individual's slope and the overall slope. Again, the general parameterization of the SEM growth model is equivalent to Equation 4.

Raudenbush and Bryk (2002) note that the HLM shown in Equations 5 and 6 can be considered a specific “covariance structure” model and is able to be estimated by standard SEM software. Willett and Sayer (1994) demonstrate how longitudinal HLM can be rewritten as a structural equation model (referred to as a covariance structure model in the paper) and estimated in LISREL. Chou, Bentler, and Pentz (1998) provide a summary of the similarities and differences between growth models in SEM and HLM contexts. They note that when the β_0 and β_1 parameters are fixed to the aforementioned constants in SEM, the results between the two models are analogous. Specifically, the regression coefficients (β in HLM and α in SEM) are equivalent. The estimates’ standard errors, the variances of the factors (initial status/intercept and slope), and covariances between initial status and slope are slightly different between the two models. The authors note that this difference is likely due to different estimation methods used in the programs they chose rather than true differences between the models. Ultimately, the constrained LGM and the longitudinal HLM presented in the above sections are equivalent models.

Summary. Each of the aforementioned approaches has their own set of strengths and weaknesses. One advantage of repeated measures ANOVA is the simplicity of the analysis. Results of repeated measures ANOVA may be easier to convey to individuals without a strong statistical background. One of the disadvantages of repeated measures ANOVA is the inability to disattenuate the parameters in the model for measurement error. In addition, repeated measures ANOVA only models overall initial status and slope and does not allow for individual variation in the slope parameter. In other words, repeated measures ANOVA makes the assumption that all individuals change the same way over time. Finally, repeated measures ANOVA makes more strict assumptions about the data (i.e., sphericity) than HLM and SEM. One disadvantage of HLM is that, like repeated measures ANOVA, the model does not easily allow for the parameters to be disattenuated for measurement error. One advantage of an HLM approach is the ability to handle data in which the data collection schedule differs across individuals. HLM allows for individuals to have varying time lengths between measurement occasions, making longitudinal data collection more feasible. HLM is advantageous over repeated measures ANOVA such that it relaxes the assumption of sphericity and allows for individuals to vary in their intercept *and* slope parameters. One potential drawback of SEM is that varying measurement schedules (i.e., varying time between measurement occasions across individuals) may be difficult to include in the model. One advantage of SEM is the ability to easily include a measurement model for the scores at each measurement occasion. In traditional LGMs, the observed total score at each time point is included in the model as the observed score for each measurement occasion. Instead, observed item responses can be modeled with a measurement model and the factor scores can be used in a second-order model to examine change over time in scores that have been disattenuated for measurement error (Ferrer, Balluerka, & Widaman, 2008). For the current study, an SEM approach will be used due to the ability to easily incorporate a second-order factor model. Notably, measurement models can be indirectly included in repeated measures

ANOVA and HLM. To include these models, researchers would first estimate factor scores using an item response theory (IRT) or confirmatory factor analysis (CFA) model and then use those scores as input for the repeated measures ANOVA or HLM. This two-step process, however, does not allow for the entire model (including the measurement model) to be estimated at the same time.

Given the wide-spread use of models for longitudinal data, the assumptions associated with the model(s) being used must be considered. The assumption of measurement invariance is essential to all longitudinal analyses. This assumption is further discussed in the subsequent section.

Measurement Invariance

Comparing scores across measurement occasions relies on the assumption that the scores can be interpreted consistently at each measurement occasion. The scales must function equivalently across all measurement occasions. This assumption is often called the assumption of measurement invariance, though it may be called factorial invariance or measurement equivalence in the literature. For simplicity, I will use the term "measurement invariance" to denote examination of scale comparability across measurement occasions. Deviations from measurement invariance suggest that the scores from an assessment may have different interpretations depending on the measurement occasion from which the score originated. In other words, a violation of measurement invariance suggests that scores from one measurement occasion represent something different than those from another measurement occasion. Notably, the IRT literature often terms invariance violations as differential item functioning (DIF). The terms “measurement invariance violation” and “DIF” will be used interchangeably throughout the study.

SEM literature distinguishes between “measurement invariance” and “structural invariance.” Measurement invariance examines equality of item intercepts, factor loadings, and error/uniqueness variances across measurement occasions (or groups). There are several levels of measurement invariance discussed in the literature. Configural invariance is the most lenient level of invariance and only assumes that the pattern of items loading on factors is the same across measurement occasions. Weak invariance adds the additional assumption that the factor loadings for each item are the same across measurement occasions. Strong invariance assumes that, in addition to equivalent factor loadings, the item intercepts are equivalent across measurement occasions. Strict invariance adds the assumption that the error variances across measurement occasions are equivalent. A more thorough discussion of each level of invariance is provided in a later section. Structural invariance explores the equality of factor means, factor variances, and factor covariances across measurement occasions (Byrne, Shavelson, & Muthén, 1989). Generally, examinations of measurement invariance determine equality of constructs across measurement occasions, whereas examinations of structural invariance tend to provide answers to more substantive questions. For the current study, I will focus exclusively on issues related to violations of measurement invariance.

Longitudinal Measurement Invariance. Measurement invariance is most often tested in independent group comparison contexts using manifest grouping variables such as gender or ethnicity. The current body of research on invariance tends to focus on cross-group comparisons with independent groups and continuous observed variables. Generally, examinations of measurement invariance are fairly uncommon. Borsboom (2006) notes that, “often, measurement invariance is tacitly assumed rather than investigated” (p. S178).

Situations in which longitudinal data is used to examine change over time have been largely ignored in measurement invariance literature. Pentz and Chou (1994) note that, “Virtually no studies have applied systematic tests of measurement invariance to longitudinal data in the context of evaluating intervention effects on multiple groups.” More recently, Schmitt and Kuljanin (2008) conducted a review of invariance studies between 2000 and 2008 and only 18% of the studies made any mention of examining invariance across time (i.e., age, cohort, retest). In longitudinal situations measurement invariance is often overlooked completely and researchers assume scales function equivalently across measurement occasions. The lack of investigations into longitudinal measurement invariance is concerning given that there are several instances in which a construct’s operational definition is predicted to change over time (Wirth, 2008). Wirth (2008) notes that anxiety, temperament, reading ability, aspects of personality, and antisocial behaviors are theorized to change in their manifestation throughout development.

Second language literature contains examples of scales that change across measurement occasions. Hulstijn (2001) provides a general overview of second language vocabulary learning. The author notes that coding new vocabulary words may be more or less difficult depending on a learner’s prior phonetic knowledge. Consider a situation where the researcher is interested in measuring a second language learner's ability to acquire new vocabulary in an introductory language course. In the beginning of the course, item responses on a vocabulary exam may depend on two constructs: actual vocabulary knowledge *and* familiarity with the phonetics. This secondary construct (phonetics) can be considered unintended multidimensionality and may hinder accurate measurement of vocabulary knowledge (Ackerman, 1992). As the course progresses, phonetic language becomes more automatic and have less of an impact on vocabulary acquisition. Thus, scores on a vocabulary test later in the year may only depend on one construct: actual vocabulary knowledge. In this scenario, differences across the measurement occasions represent differences in vocabulary knowledge *and* phonetic knowledge. If the researcher is only interested in vocabulary growth, inferences about score changes without acknowledging the role of phonetic knowledge in early measurement occasions would be inappropriate.

Another approach to conceptualizing changes in scales across time is that the item parameters are "drifting" over time. Item parameter drift is defined as changes in item parameters over measurement occasions due to factors other than sampling error (Goldstein, 1983). Comparing scores that are not invariant across measurement occasions is akin to comparing apples and oranges. If the assumption of measurement invariance is not met, researchers should be extremely cautious when interpreting comparisons across measurement occasions.

Purpose

Growth models are often used in educational contexts to examine changes over time in students' knowledge or attitudes. Repeated measurements assume that the scale is invariant across measurement occasions. Given that measurement invariance is often assumed rather than explicitly tested, researchers need to consider the implications violations of measurement invariance may have on growth model parameters. In situations where measurement invariance is explicitly tested, researchers often allow for small violations of measurement invariance (i.e., partial measurement invariance), but little research has been conducted to determine what constitutes a "small" violation of invariance. Widaman, Ferrer, and Conger (2010) note that, "Few guidelines have been developed for comparing and interpreting models that have partial measurement invariance, even though partial invariance is not unexpected." The purpose of the current study is to examine how varying degrees of longitudinal measurement invariance violations impact growth model parameters. The results of this study will help to determine the extent to which growth models are robust to violations of measurement invariance. The study will examine the impact varying levels of invariance violations have on growth model parameters (i.e., the estimated intercept, slope, and variance components), and the estimated shape of change over time (i.e., linear vs. quadratic). In addition, the study will examine whether explicitly modeling the invariance violations (i.e., modeling partial measurement invariance) allows researchers to accurately model growth. The overarching goal of the study is to determine whether or not violations of measurement invariance compromise the validity of results from growth models. Thus, this study is meant to establish whether or not measurement invariance violations in growth models are an issue, not how to solve the issue.

Organization of Study

The remaining chapters provide a review of the literature, study methodology, results, and a summative discussion. Chapter two reviews the relevant literature. The chapter begins with an overview of potential growth models in SEM contexts. Next, measurement invariance concepts are more thoroughly discussed. A review of the literature examining the impact of invariance violations and the research questions for the current study are also provided. Chapter three focuses on the methods used to answer the research questions. The connections between SEM and IRT models are briefly discussed as they relate to data generation and the interpretation of results. The third chapter also describes the data simulation design, modeling approach, and criteria by which the results are evaluated. Chapter four displays the results of the study. The final chapter provides a general discussion of the results, implications for researchers, study limitations, and potential future directions.

CHAPTER II

LITERATURE REVIEW

Growth Modeling in SEM

Growth modeling in an SEM context was briefly described in chapter one. Latent growth models (LGM) are often similar to the one presented in Equations 9 and 10 such that they use total scores at each measurement occasion. Typically, models in SEM are presented in a path diagram. An example path diagram for a traditional growth model such as the one presented in Equations 9 and 10 is presented in Figure 1. To maintain a clear figure, item intercepts and factor variances are omitted from the path diagram and all other path diagrams in the current chapter.

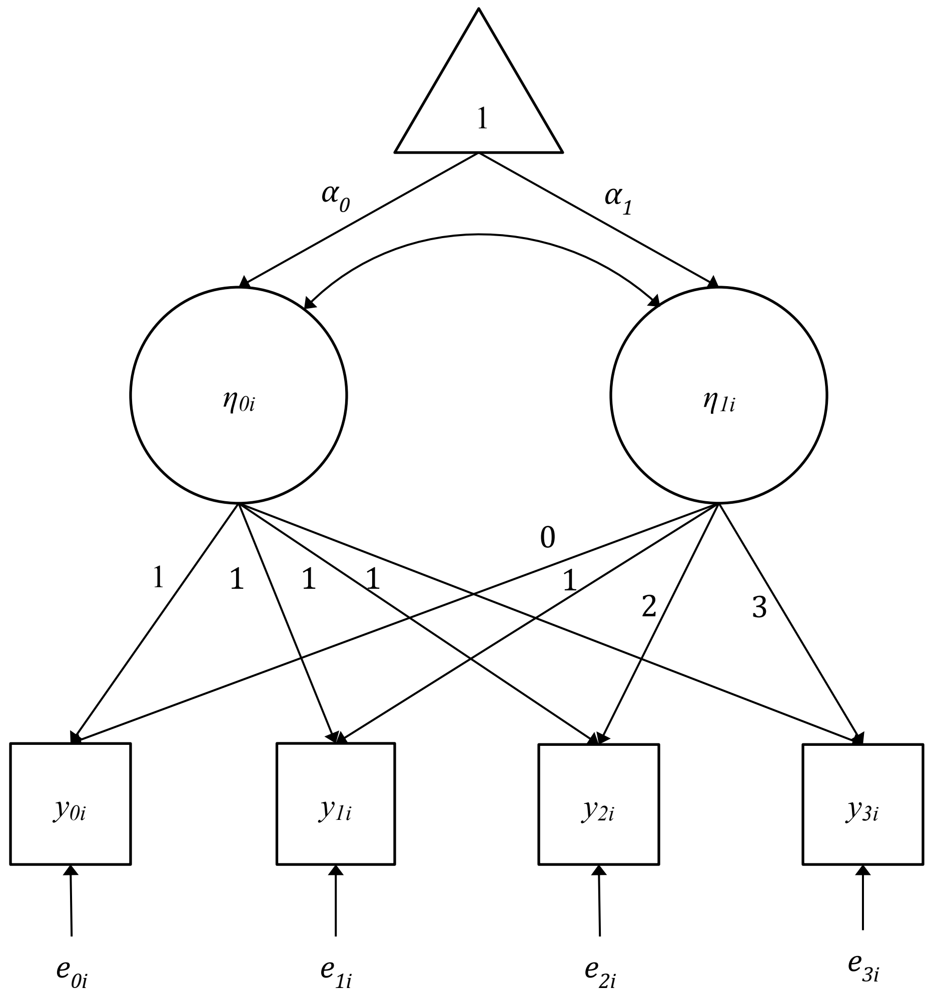

Figure 1. Path Diagram for a First-Order LGM

As previously noted, another approach to growth modeling in SEM would be to include a measurement model for each measurement occasion. This approach is often called a second-order LGM and allows researchers to use factor scores as input for the growth component of the model. Including a measurement model and using factor scores for the growth model allows researchers to examine both change over time and the measurement invariance across measurement occasions. To illustrate the second-order

LGM consider a scenario with a three item scale measured at several time points. The second-order LGM can be estimated using Equations 11, 12, and 13 below.

$$\begin{aligned} X_{ti} &= \tau_{Xt} + \lambda_{Xt} y_{ti} + e_{Xti} \\ W_{ti} &= \tau_{Yt} + \lambda_{Wt} y_{ti} + e_{Yti} \\ Z_{ti} &= \tau_{Zt} + \lambda_{Zt} y_{ti} + e_{Zti} \end{aligned} \tag{11}$$

$$y_{ti} = \beta_0 \eta_{0i} + \beta_1 \eta_{1i} + \varepsilon_{ti} \tag{12}$$

$$\begin{aligned} \eta_{0i} &= \alpha_0 + \zeta_{0i} \\ \eta_{1i} &= \alpha_1 + \zeta_{1i} \end{aligned} \tag{13}$$

Note that Equations 12 and 13 are identical to those described in chapter one. Note that in this instance, however, y_{ti} is a factor score (i.e., latent variable) rather than the observed total score at each measurement occasion. The distinction is the inclusion of Equation 11. In Equation 11, X_{ti}, W_{ti}, and Z_{ti} represent individual *i*'s responses to item X, Y, and Z, respectively, at measurement occasion *t*. The τ parameters represent the intercepts for each item and the λ parameters represent the loadings for each item. The subscript *t* on the τ and λ parameters suggests that item intercepts and slopes can differ across measurement occasions. Often intercepts and slopes are constrained to be equivalent across measurement occasions but, as will be further discussed later, these parameters can be freely estimated for each measurement occasion if measurement invariance does not hold. As previously noted, y_{ti} is the factor score for person *i* at measurement occasion *t*. The general equation for Equations 12 and 13 were presented in

Equation 4. The general equation for the measurement model (Equation 11) is presented in Equation 14 below.

$$\mathbf{W} = \mathbf{T} + \mathbf{\Lambda Y} + \mathbf{E} \tag{14}$$

In Equation 14 above, the **T** matrix contains the item intercepts for each item. The **Λ** matrix contains the item loadings for each item. The **Y** matrix contains the factor scores and the **E** matrix contains the residual item variance for each item. Note that this measurement model is another parameterization of a general linear mixed model in which fixed effects (item characteristics) and random effects (individual factor scores) predict responses to items. The path diagram for the second-order growth model presented in Equations 11, 12, and 13 is presented in Figure 2 below.

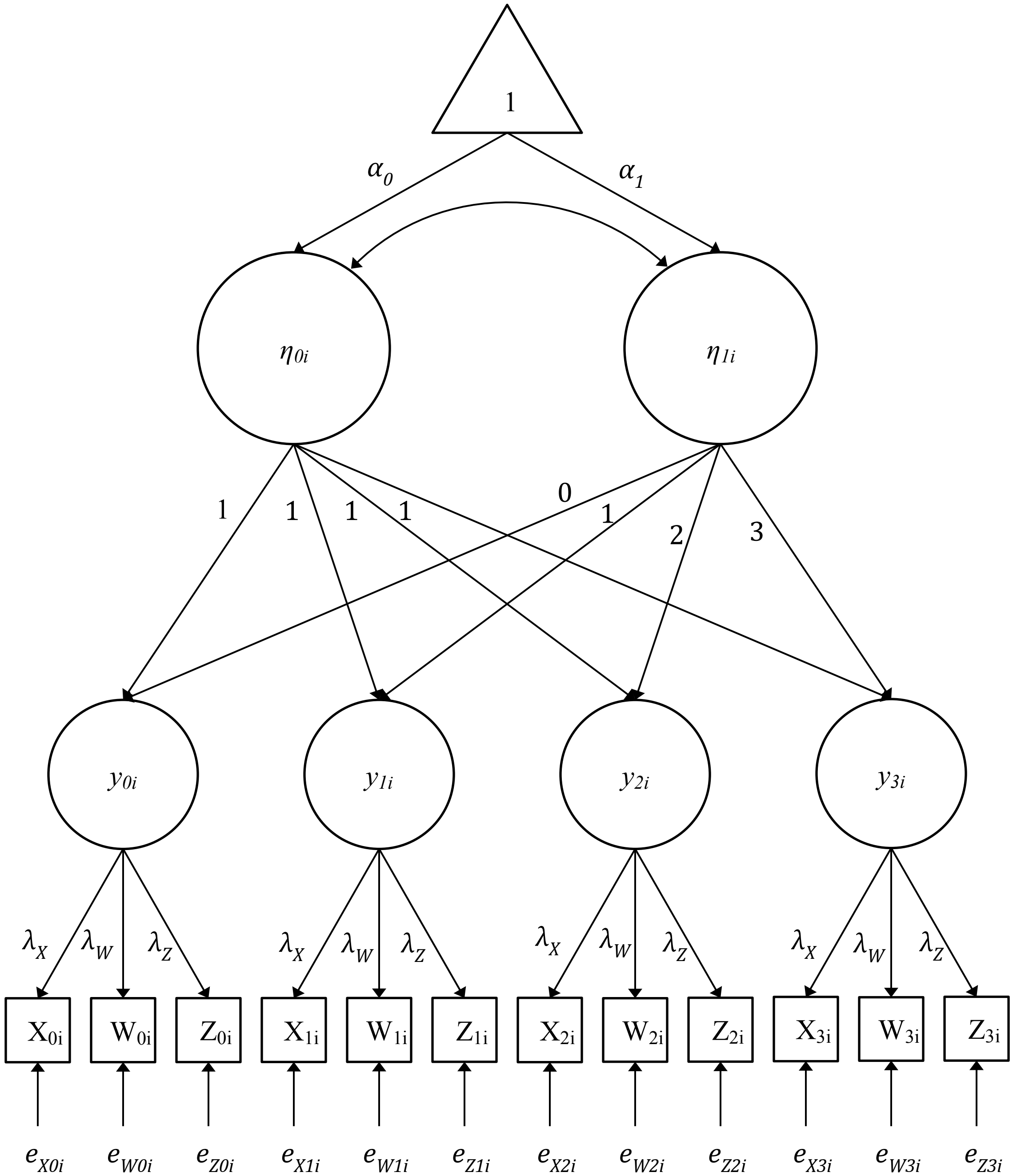

Figure 2. Path Diagram for Second-Order LGM.

In addition to including the measurement model for each measurement occasion, non-linear change in scores over time can be incorporated to latent growth models in an

SEM context. In a non-linear LGM, additional slope parameters are included to capture non-linear changes in scores over time. If researchers wanted to model quadratic change, a parameter for the linear slope and a parameter for the quadratic slope are included. Equations 15 and 16 display a first-order non-linear LGM with a quadratic slope.

$$y_{ti} = \beta_0 \eta_{0i} + \beta_1 \eta_{1i} + \beta_2 \eta_{2i} + e_{ti} \tag{15}$$

$$\begin{aligned} \eta_{0i} &= \alpha_0 + \zeta_{0i} \\ \eta_{1i} &= \alpha_1 + \zeta_{1i} \\ \eta_{2i} &= \alpha_2 + \zeta_{2i} \end{aligned} \tag{16}$$

In Equation 15, the y_{ti} parameter represents individual *i*'s observed score at time *t*. The β parameters are fixed and represent time. The β_0 parameter is the intercept and is fixed to a vector of one's. The β_1 parameter is the linear slope and is fixed to the units of time (e.g., 0, 1, 2, and 3). The linear slope in a non-linear model represents the slope of the line at the intercept. The β_2 parameter is fixed to the squared values of the time scores in the β_1 vector (e.g., 0, 1, 4, and 9). The η_{0i}, η_{1i}, and η_{2i} parameters represent the intercept, linear slope, and quadratic slope for individual *i*. In Equation 16, the first equation models the individual intercept, η_{0i}, as a function of the overall intercept, α_0, and the difference between the individual's intercept and the overall intercept, ζ_{0i}. The second equation models the individual linear slope, η_{1i}, as a function of the overall linear slope, α_1, and the difference between the individual's linear slope and the overall linear slope, ζ_{1i}. The final equation models the quadratic slope for each individual, η_{2i}, as a function of the overall quadratic slope, α_2, and the deviation of the individual's quadratic slope from the overall quadratic slope, ζ_{2i}. The model for the first-order non-linear LGM with a quadratic slope is presented in the path diagram in Figure 3.

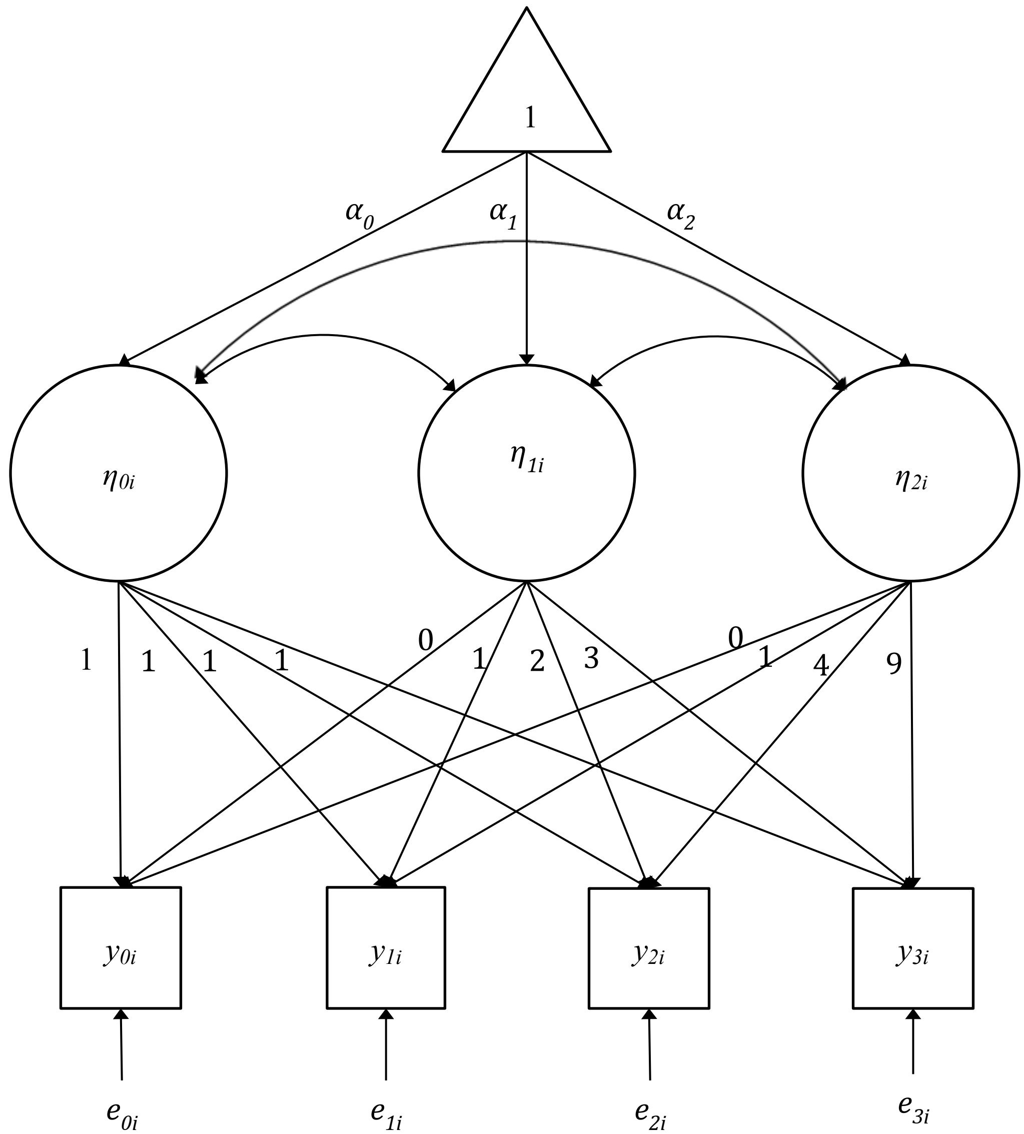

Figure 3. Path Diagram for First-Order Non-Linear LGM with Quadratic Slope

As with the first-order model, non-linear slopes can be included in the second-order LGM. An example of the second-order non-linear LGM with a three item scale at each measurement occasion and a quadratic slope is presented in Equations 17, 18, and 19 below.

$$\begin{aligned} X_{ti} &= \tau_{Xt} + \lambda_{Xt} y_{ti} + e_{Xti} \\ W_{ti} &= \tau_{Wt} + \lambda_{Wt} y_{ti} + e_{Yti} \\ Z_{ti} &= \tau_{Zt} + \lambda_{Zt} y_{ti} + e_{Zti} \end{aligned} \tag{17}$$

$$y_{ti} = \beta_0 \eta_{0i} + \beta_1 \eta_{1i} + \beta_2 \eta_{2i} + e_{ti} \tag{18}$$

$$\begin{aligned} \eta_{0i} &= \alpha_0 + \zeta_{0i} \\ \eta_{1i} &= \alpha_1 + \zeta_{1i} \\ \eta_{2i} &= \alpha_2 + \zeta_{2i} \end{aligned} \tag{19}$$

Note that Equations 18 and 19 look identical to Equations 15 and 16. Including the measurement models in Equation 17 distinguishes the first-order model from the second-order model. Including the measurement models in Equation 17 means that the y_{ti} parameter in Equation 18 is interpreted as a factor score for individual *i* at time *t*. The path diagram for the second-order non-linear LGM with a quadratic slope is presented in Figure 4.

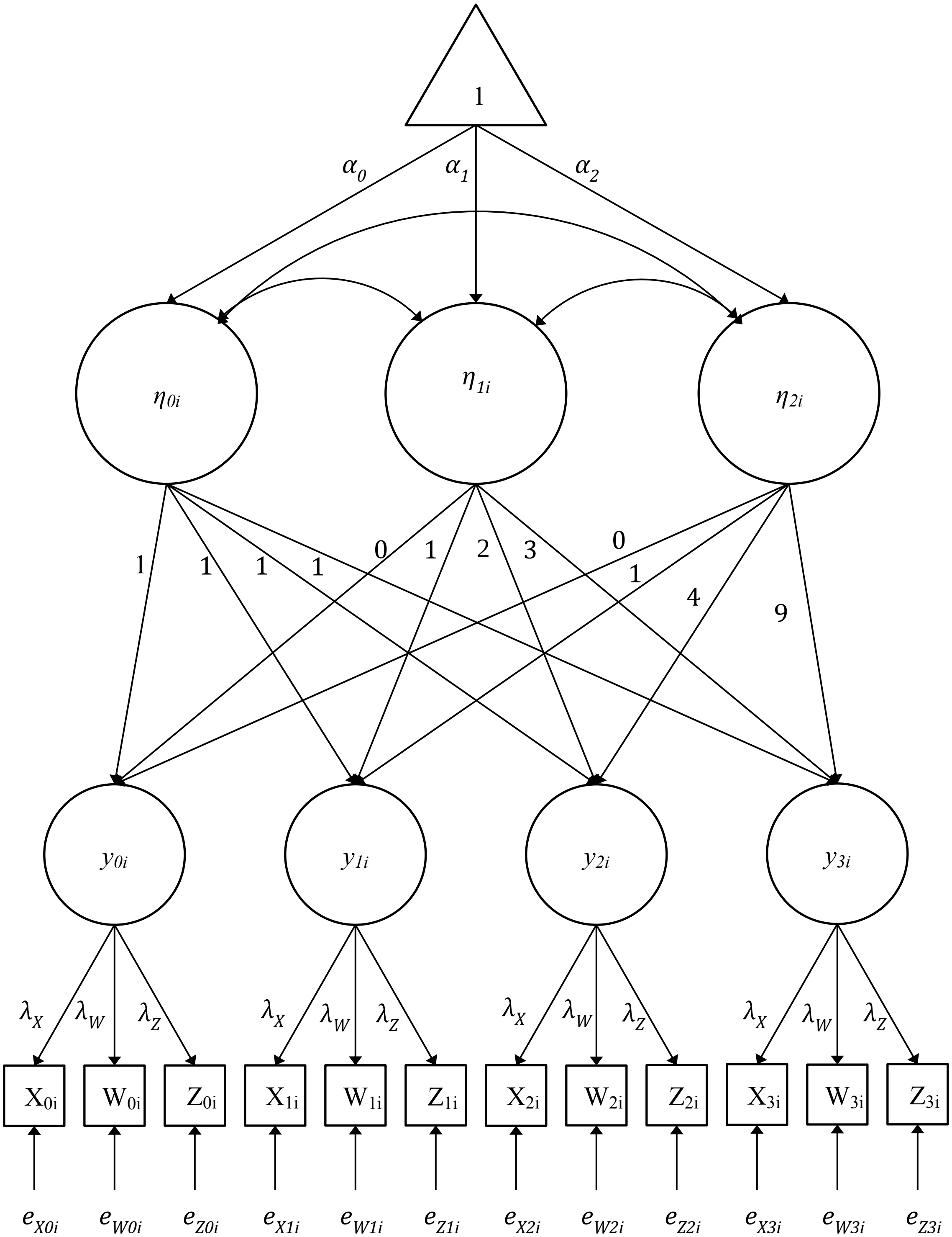

Figure 4. Path Diagram for Second-Order Non-Linear LGM with a Quadratic Slope.

As with the addition of a second slope for quadratic change over time, a third slope parameter can be added to model cubic change over time. The non-linear LGM with a cubic slope will not be presented as it is not the focus of the current study. In addition, predictors can be added to all of the models described above to help explain individual variability in intercepts and slopes. Adding predictors helps researchers to better understand individual variations in how people change over time. Despite the utility, the current study will not focus on models that have predictors of intercepts and slopes and thus, further description of adding predictors to the model is not provided. As will be shown in the next section, a dichotomous confirmatory factor analysis (CFA), like the one at the first level of the 2LGM, is equivalent to a 2 parameter logistic (2PL) IRT model. The choice of framework (SEM/CFA vs. IRT) can be used at the discretion of the researcher. In this case, the SEM framework is more common for longitudinal analyses and thus, allows us to model growth in the same way most researchers would approach growth modeling

Connections between SEM and IRT Models

SEM approaches are commonly used to examine measurement invariance in longitudinal models and to model longitudinal data. SEM is typically used with continuous, normally distributed observed variables, whereas IRT is typically used with binary or categorical observed variables. Notably, modern approaches to SEM allow for the analysis of discrete data, but IRT is still the preferred method with unidimensional constructs. There are several IRT models commonly used in practice, but for the current study the focus will be on the two-parameter logistic (2PL) model. The 2PL predicts item responses (i.e., the probability of a correct response) as a function of item difficulty and item discrimination. Equation 20 shows a traditional 2PL model.

$$P(Y_i = 1 | \theta) = \varphi = \frac{e^{1.7a_i(\theta_j - b_i)}}{1 + e^{1.7a_i(\theta_j - b_i)}} \tag{20}$$

In Equation 20 above, $P(Y_i = 1|\theta)$ represents the probability of a correct response for item i, after controlling for ability, θ. In the equation, 1.7 is used as a scaling factor to approximate the normal ogive. The a_i parameter is the discrimination for item i. The θ_j parameter is the ability estimate for individual j. The difficulty estimate for item i is represented by the b_i parameter.

Different vocabulary is used in the SEM and IRT literatures to discuss item characteristics. In SEM, the term "factor," represented as η in the current manuscript, is often used to reference the construct of interest. In IRT, the construct is often referred to as "ability" and is represented by θ in the current manuscript. The parameter that represents the relationship between each item and the construct of interest is called a "factor loading" (often represented by λ) in SEM and "item discrimination" (often represented by a) in IRT. The expected item response when the construct is equal to zero is called the "item intercept" (often represented by τ) in SEM and the "item difficulty" (often represented by b) in IRT. In IRT, the b parameter represents the point on the ability scale at which a respondent had a 50% chance of a correct response. The current study will examine longitudinal measurement invariance violations in dichotomous observed variables, making IRT the first choice for analysis. Because growth models are most often estimated in practice within a SEM framework, however, the current study will use a SEM framework in order to most closely approximate real world scenarios. Ultimately, some SEM and IRT models are equivalent and thus, the framework in which the models are estimated should have minimal impact on the results. The subsequent paragraphs demonstrate the equivalencies between confirmatory factor analysis (CFA) in SEM and the 2PL in IRT.

Recall that the measurement component of the LGM was presented in terms of the general linear mixed model. The general linear mixed model is appropriate when the dependent variable is continuous and normally distributed. When the dependent variable is categorical, in this case dichotomous, the *generalized* linear mixed model most appropriately models the data. The generalized linear mixed model is simply an extension of the general linear mixed model. Specifically, it allows for continuous *and categorical* dependent variables. Thus, the general linear mixed model can be considered a special case of the generalized linear mixed model in which a continuous dependent variable is used. Notably, when categorical (nominal or ordinal) data are modeled using the linear model, it is possible to obtain predicted values outside of the plausible range. For example, if a dichotomous outcome variable (e.g., right=1/wrong=0) is modeled with a linear model, it is possible to obtain predicted values less than 0 and/or greater than 1.

In order to appropriately model categorical dependent variables, the generalized linear mixed model uses a link function. A link function transforms the expected value of the dependent variable so that a linear relationship can be modeled between the independent and dependent variables. There are several different link function options depending on the type of dependent variable. For dichotomous data, the logit (a.k.a., log-odds) link function is often used (Raudenbush & Bryk, 2002). The HGLM described in chapter one can be used to model dichotomous item response data. Kamata (2001) highlights that item response data is considered nested such that items are repeated measurements within an individual. Thus, items, modeled at level 1, are nested within people, modeled at level 2. In order to model a linear relationship between the dichotomous responses and predictors (person ability and item parameters), a logit link function can be used. To obtain the log-odds of a correct response, we can first calculate the odds of a correct response. The odds of a correct response is defined as the probability of getting and item correct divided by the probability of getting an item incorrect. The calculation of the odds of a correct response for the 2PL model is shown in Equation 21.

$$\begin{aligned}\left(\frac{\varphi}{1-\varphi}\right) &= \frac{\dfrac{e^{1.7a_i(\theta_j-b_i)}}{1+e^{1.7a_i(\theta_j-b_i)}}}{\dfrac{1}{1+e^{1.7a_i(\theta_j-b_i)}}} \\ &= \frac{e^{1.7a_i(\theta_j-b_i)}}{1} \\ &= e^{1.7a_i(\theta_j-b_i)}\end{aligned} \tag{21}$$

In order to model the log-odds (or logit) of a correct response, the log can be taken for both sides of Equation 21 as shown in Equation 22.

$$\log\left(\frac{\varphi}{1-\varphi}\right) = \log\left[e^{1.7a_i(\theta_j - b_i)}\right]$$
$$= 1.7a_i(\theta_j - b_i) \qquad (22)$$
$$= -1.7a_i b_i + 1.7a_i\theta_j$$

The equation above may not initially look like the equation for the measurement model in SEM, but the parameters are comparable. The first component of the equation, -1.7$a_i b_i$, represents the item intercept and is analogous to τ in the SEM measurement models. The second part of the equation, 1.7$a_i\theta_j$, multiplies the item discrimination by the individual's score on the factor. This component is analogous to the factor loading (λ) multiplied by the factor score (η) in the SEM measurement models. Takane and de Leeuw (1987) demonstrate the equivalence between CFA and the normal ogive model in IRT. The addition of the 1.7 scaling factor to the 2PL makes the logistic model practically equivalent to the normal ogive (Hambleton & Swaminathan, 1985). Brown (2006) notes that to calculate SEM factor loadings from IRT discrimination parameters, Equation 23 below can be used.

$$\lambda_i = \frac{a_i}{\sqrt{\left(1 + a_i^2\right)}} \qquad (23)$$

To obtain SEM item intercepts, Equation 24 below can be used.

$$\tau_i = -a_i b_i \qquad (24)$$

Muthén (2012) notes that often the differences between IRT and categorical CFA are due to estimator differences rather than true model differences. Specifying a unidimensional, dichotomous CFA in SEM is equivalent to specifying a 2PL model in IRT. One advantage of specifying the model in the SEM framework is the ability to seamlessly incorporate the measurement model into a full structural model. This advantage allows for a categorical CFA (i.e., 2PL) model to be incorporated into a second-order LGM. Because of this possibility, the current study will be conducted in an SEM framework.

As noted in chapter one, all longitudinal models assume that the scales function equivalently across measurement occasions. The following section provides a more thorough overview of measurement invariance concepts and how to test for different types of measurement invariance.

Measurement Invariance

As noted in chapter one, examinations of measurement invariance in SEM literature are often categorized into structural invariance and measurement invariance. Measurement invariance can be further categorized into configural, weak, strong, and strict measurement invariance (Millsap & Meredith, 2004). Configural invariance refers to the equivalence of the factor pattern matrices across measurement occasions. If configural invariance holds, the pattern of zero and non-zero loadings across measurement occasions is equivalent (i.e., the same items measure the same factor(s) across all measurement occasions). Weak measurement invariance implies that the values within the factor pattern matrix are equivalent across measurement occasions. In other words, weak measurement invariance examines whether the relationship between each item and the factor is the same across measurement occasions. If weak measurement invariance is tenable, the loadings for each item on a given factor are the same for all measurement occasions. Because factors are often defined or named by the items that have the highest loadings, weak measurement invariance is particularly important from a substantive perspective. Strong measurement invariance posits that, in addition to configural and weak measurement invariance, the item intercepts for all measurement occasions are equal. Strong measurement invariance relies on the assumption that item difficulty is equal for all measurement occasions. If strong measurement invariance holds, the pattern, loadings, and intercepts are equivalent across all measurement occasions. Strict measurement invariance builds upon strong measurement invariance to include the equivalence of uniquenesses (i.e., error variances) across all measurement occasions (Millsap & Meredith, 2004). Note that for all four conditions of measurement invariance, the previous condition must be met before continuing with more stringent invariance tests (i.e., configural invariance must hold before testing for weak measurement invariance). Notably, some researchers argue that small departures from invariance, often referred to as partial measurement invariance, may permit researchers to assess more constrained invariance conditions (Byrne et al., 1989; Millsap & Meredith, 2004). There is little research that provides guidelines to define "small" departures of invariance. Partial measurement invariance will be further discussed later in this chapter.

The goals of a given study define the level of measurement invariance necessary for researchers to be confident in the interpretations of the results. For situations in which researchers are only interested in the measurement of the factor variance-covariance structures, only the factor loadings need to be invariant across the groups. That is, weak measurement invariance must hold if researchers are interested in the relationships between factors or the variability of factor scores. For situations in which factor means are of interest, the factor loadings and intercepts must be equivalent across groups (Byrne et al., 1989). Thus, to compare means across measurement occasions (i.e., examining whether scores change over time) and have confidence in the interpretations, strong measurement invariance must hold.

Tests of measurement invariance are conducted with latent variable (confirmatory factor analysis; CFA, and item response theory; IRT), or observed variable (i.e., Mantel-Haenszel) approaches. For the current study, I will focus on describing test of invariance for latent variable approaches. For clarity, CFA literature often refers to the examination of measurement invariance as measurement or factorial invariance testing. The IRT literature tends to refer to these analyses as examinations of DIF or bias (i.e., lack of invariance). CFA and IRT literature bases also use different terms for the different levels of measurement invariance. IRT typically focuses on unidimensional scales and thus, configural invariance is rarely discussed in the IRT literature. Table 1 provides common vocabulary to describe different levels of measurement invariance in the CFA and IRT literature bases.

Table 1

Vocabulary Used to Describe Levels of Measurement Invariance

Millsap & Meredith (2007)	CFA Literature	IRT Literature	Description (Equivalence Constraint Across Time)	Notation
Configural	Configural	--	Pattern matrix	
Weak	Metric	Non-uniform DIF	Loadings (a-parameters)	$(\mathbf{\Lambda}_{\text{Time1}} = \mathbf{\Lambda}_{\text{Time2}})$
Strong	Scalar	Uniform DIF	Intercepts (b-parameters)	$(\mathbf{\Lambda}_{\mathbf{Time1}} = \mathbf{\Lambda}_{\text{Time2}}$ and $\mathbf{T}_{\text{Time1}} = \mathbf{T}_{\text{Time2}})$
Strict	Strict	N/A	Error variance	$(\mathbf{\Lambda}_{\mathbf{Time1}} = \mathbf{\Lambda}_{\text{Time2}}$ and $\mathbf{T}_{\text{Time1}} = \mathbf{T}_{\text{Time2}}$ and $\mathbf{E}_{\text{Time1}} = \mathbf{E}_{\text{Time2}})$

The current study will use the terminology outlined in Millsap and Meredith (2007) (i.e., configural, weak, strong, strict) to describe levels of measurement invariance. While the general definitions of invariance are the same for all modeling approaches, the methods for assessing invariance differ. One of the largest differences between the methods is the assumptions the models make on the data. For example, in traditional CFA methods, the models assume that the observed variables are continuous and multivariate normally distributed. Traditional IRT models, on the other hand, assume that the observed variables are categorical and do not need to be multivariate normal. Because there are many different approaches to assessing invariance with many different assumptions, assessing how different analytic approaches impact detection of measurement invariance is difficult (Borsboom, 2006). For a more thorough description of the similarities and differences in assessing measurement invariance between CFA and IRT see Raju, Laffitte, and Byrne (2002). Ultimately, both IRT and CFA approaches to identifying DIF examine the magnitude of the differences in item parameters between measurement occasions to determine whether they are significant.

Given that an SEM framework will be used for this study and that the 2PL is equivalent to a categorical CFA model, SEM approaches to identifying measurement invariance violations are briefly described here. Nested model comparisons are the most common approach to examining weak, strong, and strict invariance violations in SEM. To test for weak measurement invariance, two models are estimated. The first model allows all item loadings to be freely estimated across measurement occasions. The second model constrains corresponding item loadings to be equivalent across measurement occasions and represents a model in which the assumption of weak measurement invariance holds. The second model is considered nested within the first model such that constraining the item loadings in the first model leads to the second model. The two models can be statistically compared through a chi-squared difference test and fit indices to determine whether constraining the item loadings (i.e., assuming weak invariance) results in significantly worse fit. If the constrained model has significantly worse fit, the assumption of weak measurement invariance is violated. If weak measurement invariance holds, strong measurement invariance can be tested. Similar approaches are used to test strong and strict measurement invariance. For strong measurement invariance, a model with equal item loadings, but freely estimated item intercepts is compared with a model that constrains item loadings and item intercepts to be equivalent across measurement occasions. If strong measurement invariance holds, strict measurement invariance can be tested. For strict measurement invariance, a model with equal item loadings, equal item intercepts, and freely estimated item residuals is compared with a model that constrains item loadings, intercepts, and residuals across measurement occasions. If the constrained model fits significantly worse than the less constrained model with which it is being compared, that level of measurement invariance cannot be assumed. In these situations, researchers may further explore partial measurement invariance, as is described in the next section.

Partial Measurement Invariance. Partial measurement invariance was introduced by Byrne, Shavelson, and Muthén (1989) and allows non-invariant item parameters to be freely estimated between groups (or measurement occasions).

Steenkamp and Baumgartner (1998) note that full measurement invariance may be unlikely in tests of weak, strong, and, especially, strict measurement invariance. Further, Horn (1991, as cited by Steenkamp & Baumgartner, 1998) suggests that metric invariance is “a reasonable ideal...a condition to be striven for, not one expected to be fully realized.” Overall, partial measurement invariance at the weak, strong, and strict levels seems to be a common occurrence and, generally, well accepted in the literature. Some researchers caution against the use of empirically derived partial measurement invariance with little or no theoretical basis for freeing parameters across groups (Byrne, Shavelson, & Muthén, 1989). Given the prevalence of partial measurement invariance, researchers may want to consider how partial measurement invariance is handled in CFA and IRT approaches to measurement.

In the CFA approach, partial measurement invariance is often included in the model and thus, is accounted for when interpreting values on the latent variable. In the IRT literature, there is little discussion of partial measurement invariance. The lack of discussion in IRT contexts may be because DIF is often examined in educational testing scenarios where exam scores are often used to make high stakes decisions at the individual level. DIF may be problematic in many situations, but in scenarios involving individual-level decision making, DIF is particularly problematic. In an Educational Testing Service (ETS) research report on DIF procedures, Zwick (2012) describes the current system used by ETS to classify DIF. Items are classified into three groups: A (negligible or non-significant DIF), B (slight to moderate DIF), or C (moderate to large DIF). The article notes that the reasoning behind the cutoffs used to classify items into the A, B, or C categories is that small values are undesirable, but tolerable, whereas larger values should be avoided. Notably, this suggests that some small violations of invariance are tolerable (i.e., partial measurement invariance), but other, more severe cases of invariance violations should be excluded from scoring and other analyses. The article goes on to examine the performance of the current classification system to accurately classify items with DIF. The author suggests devising more effective rules for identifying DIF and notes that, "the first step should be to reconsider the issue of minimal DIF magnitude that is of concern (and is therefore important to detect) as well as the level of false positives that can be tolerated." (p. 10). The discussion of minimal DIF magnitude of concern suggests that some "small" instances of partial measurement invariance is tolerated in IRT contexts. Notably, if an item displays small amounts of DIF (i.e. a "tolerable" amount), it is typically included in the assessment, but is treated as if it were invariant. Thus, small amounts of DIF are typically ignored in IRT contexts. As small amounts of DIF (i.e., partial measurement invariance) may be common in IRT contexts, researchers should ensure that the models we use in practice are robust to these "small" violations of measurement invariance.

The examination of model performance under partial measurement invariance is the focus of the current study. For the current study, dichotomous items will be modeled within a SEM framework to examine IRT invariance violations on growth models. Grounding the current study in IRT, but modeling the data in a SEM framework provides several benefits. The SEM framework allows for comparison between models with partial measurement invariance included and models in which invariance violations are ignored.

Invariance Testing in the Literature

Borsboom (2006) describes situations in which measurement invariance is necessary for valid interpretations across groups and situations in which measurement invariance is less of a necessity. Ultimately, a violation of measurement invariance is always an issue; however, there may be some scenarios in which measurement invariance violations are less of a concern. Borsboom describes a situation in which an invariance violation is an order of magnitude smaller than the targeted effect size in between-group comparisons. When the invariance violation is small in relation to the anticipated effect, there may be less serious concern. This scenario assumes that researchers can accurately predict the effect size, which is often unrealistic in applied research. Several researchers have examined the impact of invariance violations on results.

Vandenberg and Lance (2000) provide a representative summary of applied studies in which measurement invariance was tested using CFA approaches in conjunction with a substantive hypothesis. The authors identified 67 studies that report some form of invariance testing before investigating substantive hypotheses. The authors note a variety of motivations driving examinations of measurement invariance (e.g., to supplement and extend traditional examinations of validity, to examine cross cultural generalizability of a scale). Of the 67 studies identified, 88% of the studies included a test of configural invariance. Tests for weak (i.e., metric) invariance were reported in 99% of the studies. Fewer studies reported tests for strict invariance (i.e., uniquenesses) (49%), factor variances (33%), factor covariances (58%), latent mean differences (21%), and strong (i.e., scalar) invariance (12%). Notably, strong invariance is necessary for analyses of latent mean differences, however, strong invariance was tested less frequently than latent mean differences.

Schmitt and Kuljanin (2008) provide an update to the review of measurement invariance applications using CFA completed by Vandenberg and Lance (2000). The authors identified 75 articles published after 2000 that conducted empirical analyses of measurement invariance with CFA models. The authors note that most of the papers investigate measurement invariance to support the use of the instrument across groups (e.g., gender, ethnicity). All of the articles reported testing configural and weak measurement invariance. This is consistent with the high occurrence of configural and weak measurement invariance found in Vandenberg and Lance (2000). Unlike the Vandenberg and Lance article that reported a 12% occurrence of strong measurement invariance testing, Schmitt and Kuljanin found that 54% of the articles reported tests of strong invariance. This suggests an increase in strong measurement invariance testing between the times of the articles included in the Vandenberg and Lance summary and the articles included in the Schmitt and Kuljanin summary. The authors note that this increase may be due to researchers becoming more aware of the latent mean testing available or the Vandenberg and Lance review that outlined best practices for group comparison.

The Vandenberg and Lance review did not mention the prevalence of partial measurement invariance testing. Schmitt and Kuljanin found that 50% of the studies included tests of partial measurement invariance. The Schmitt and Kuljanin article suggests that allowing for partial measurement invariance is fairly common in practice.

The measurement invariance reviews conducted by Vandenberg and Lance (2000) and Schmitt and Kuljanin (2008) focus on examinations of measurement invariance from a CFA approach and most studies focused on independent group comparisons rather than longitudinal comparisons.

Impact of Invariance Violations

A large body of research exists surrounding the detection of non-invariant item parameters, whereas fewer studies have examined the consequences of specifying a model that ignores non-invariant items. Researchers may be interested in examining how failing to model partial invariance affects longitudinal research results. Currently, there is no consensus on how to proceed under conditions of partial measurement invariance in longitudinal (or multi-group) analyses. This lack of consensus is likely due to the generally sparse body of research investigating the impact of invariance violations under controlled circumstances (i.e., simulation-based research). The research on invariance violations can be categorized into studies using real data (i.e., applied approaches) and studies using simulated data.

Applied Studies. Several researchers have examined the impact of invariance violations in varying situations. Schmitt and Kuljanin (2008) summarize measurement invariance literature and provide a small study with real data examining the impact violations of strong invariance have on factor mean estimates. The authors found that five item intercepts on a scale were not invariant and fit three multiple group models (unconstrained, constraining all intercepts to be equal, and allowing the five intercepts to be estimated freely) to examine the impact of non-invariance on factor mean estimates.

The results suggest that there is little impact on the factor means regardless of the model used. The authors note that had the differences in intercepts been in the same direction or larger, there may have been a more notable bias in factor means. This study suggests that in the presence of some invariance violations there is minimal impact on group comparisons.

Schmitt, Golubovich, and Leong (2011) examined the relationship college and high school GPAs have with career interest and personality constructs in models that do and do not account for lack of measurement invariance. The authors specified a partially invariant model that modeled differences in item parameters across groups and a fully invariant model that constrained item parameters to be equivalent across groups. The authors found minimal differences in factor correlations between the partially invariant and fully invariant models. The results also suggested minimal differences in factor means for both modeling approaches. Only one of the latent mean estimates was notably different across the two modeling approaches. The authors note that, while this difference is not statistically significant, it is practically significant and should be avoided, if possible. The differences for the regression slopes between models were also non-significant. Schmitt et al. conclude that, because the differences between subgroups were practically significant, modeling partial measurement invariance is important. Notably, the authors used item parcels in their models which may have masked larger violations of invariance in the measures.

Fleishman, Spector, and Altman (2002) examined the differences in group comparison results between models that did and did not account for DIF. The authors focused on differences in functional disability between gender and age groups. The authors first examined differences between groups using a multiple-indicator/multiple-cause (MIMIC) model without taking into account intercept DIF across groups. Results suggested that, without accounting for intercept DIF, young women, middle-aged women, and middle-aged men were significantly less functionally disabled than elderly men. Notably, young men were not significantly less disabled than elderly men. In addition, young women were significantly less disabled than young men, and middle-aged women were significantly less disabled than middle-aged men. When the authors accounted for DIF, results suggested that there was no significant difference in functional disability when middle-aged men and women were compared to elderly men. In addition, the results with adjustments for DIF suggested that the young women *and* young men were significantly less disabled than elderly men. Finally, the results for the model accounting for DIF suggest that the differences between genders (young men compared to young women and middle-aged men compared to middle-aged women) were not significant. The authors also examined the results when the two items with the largest DIF were removed from the analysis. Results suggested that an adjustment for DIF was still necessary for valid interpretations of group differences. Ultimately, the results of this study suggest that failing to account for DIF can drastically impact the results of group comparisons.

Jones and Gallo (2002) studied the effects of DIF on education and gender differences in the Mini-Mental State Examination (MMSE). The authors used a MIMIC model to compare differences in MMSE scores between levels of education and between males and females when DIF was and was not included in the model. The results suggest that observed differences between educational levels is minimally impacted by DIF. Conversely, gender differences were heavily impacted by the presence of DIF. When DIF is ignored, results suggest that there are significant differences in cognitive dysfunction between males and females. When DIF was included in the model, the differences between males and females were not significant. This study suggests that the impact of DIF on group comparisons may or may not be problematic. This conclusion makes it difficult for researchers to know whether their group comparisons can be trusted when there is DIF present and unmodeled.

Ferrer, Balluerka, and Widaman (2008) examined measurement invariance using second-order growth models. As previously discussed, second-order growth models include a measurement model for each time point. The authors used real data to examine the impact of using different indicators to identify the latent variable. The authors used two sets of data for the study. For one dataset the scale was invariant across measurement occasions. For the second dataset, neither weak nor strong factorial invariance held. The authors first examined the fit of second-order growth models when different reference indicators were used. The results of the study suggest that if measurement invariance does not hold (i.e., in the second dataset), the interpretation of the results may drastically differ depending on which item is used to set the scale for the factor. Specifically, the predicted trajectories for individuals when measurement invariance was violated were notably different depending on which item was chosen as the reference indicator. When measurement invariance held (i.e., in the first dataset), the trajectories were essentially the same regardless of the reference indicator chosen. The authors also compared the trajectories between a first-order LGC (in which measurement invariance is ignored) and a second-order LGC (in which item intercepts and loadings are freely estimated). For the dataset in which measurement invariance held (the first dataset), the trajectories were essentially the same for the two models. For the dataset in which measurement invariance was violated (the second dataset), the trajectories (both intercept and slope) differed between the two models. These results suggest that ignoring the measurement invariance violations may alter the estimated trajectory of a traditional (first-order) growth model.

The lack of consensus among applied research studies suggests the need for a more systematic approach to examining the impact of invariance violations. The next section describes the simulation studies that have been conducted to systematically examine the impact of violating the assumption of measurement invariance

Simulation Studies. Several researchers have examined the impact of invariance violations on substantive comparisons under controlled situations, such as a simulation study. Few of these researchers conducted these controlled studies in relation to longitudinal analyses (e.g., using a LGM). Millsap and Meredith (2004) indicate that researchers may ignore violations of measurement invariance (i.e., allow for partial measurement invariance) if the size of the violations and the number of non-invariant items are small. The authors also note that researchers have very little information for defining a "small" violation or a "small" number of invariant items. Several researchers have called for a more thorough examination of the impact of partial measurement invariance on substantive tests (Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). Simulation studies can help provide such information.

Chen (2008) conducted a series of studies to examine the impact of invariance violations across groups on regression slopes and factor means. The authors' first study examined the impact of weak factorial invariance violations (for predictors) on regression slopes. The author varied the proportion of non-invariant items (87.5%, 75%, 50%, and 25%), the pattern of invariance (same direction vs. mixed), and the ratio of the sample sizes across groups (equal vs. 4 to 1). The authors also examined scenarios in which the dependent variable lacked invariance. When the criterion lacked weak factorial invariance and the direction was the same for all non-invariant items, the regression slopes were overestimated for the reference group and underestimated for the focal group. The authors note that this creates a pseudo-interaction between the predictor and the grouping variable. For conditions in which the direction of bias was mixed, the bias was reduced.

Chen (2008) included a second and third study examining the impact of weak and strong factorial invariance violations on factor means. The model for this study was a one factor measurement model with no predictors. The conditions for the second study were the same as those in the first study. Artificial group differences in factor means were created when the factor loadings differed across groups. In addition, as the proportion of non-invariant items increased, bias in mean estimates increased. The conditions changed slightly for study three such that the proportion of non-invariant items was set to 100%, 75%, 50% and 25%. Note that in this simulation the factor loadings were invariant, but the intercepts were not. The results suggest that factor mean bias depends on the proportion of non-invariant items and the ratio of the sample sizes between groups. The bias in means was notably lower for the conditions in which the direction of bias in the intercepts differed across items, suggesting that some of the bias cancelled out.

Steinmetz (2013) conducted a simulation study to examine whether violations of invariance could lead to erroneous differences between groups when compared on composite means (i.e., when invariance violations were ignored). The author varied the number of non-invariant item intercepts, number of non-invariant item factor loadings, sample size per group, the total number of items in the composite, and whether or not there were true group differences on the latent mean. The results suggested that invariance violations in factor loadings had minimal impact on composite mean differences. Non-invariant intercepts, however, substantially impact the differences in composite means and, in turn, the probability of significant composite mean differences.

Olivera-Aguilar (2013) conducted a simulation study to examine the impact of longitudinal measurement invariance violations on latent growth models and autoregressive quasi-simplex models. The author varied the sample size per group, total number of items, proportion of non-invariant items, magnitude of non-invariance in item factor loadings, and magnitude of non-invariance in item intercepts. The patterns of non-invariance included conditions in which both intercepts and factor loadings were invariant, conditions with invariant intercepts and non-invariant factor loadings, and conditions with non-invariant intercepts and invariant factor loadings. Conditions with non-invariant intercepts *and* non-invariant factor loadings were not examined. Item response data for continuous items was simulated for five measurement occasions and composite values at each measurement occasion were created. The composite values were used in the LGM and autoregressive quasi-simplex model. The results of the study suggested that in conditions with non-invariant factor loadings, absolute relative bias for the slope factor mean, the slope factor variance, and the intercept-slope covariance were larger than the suggested cutoff value (0.05). In the conditions with non-invariant intercepts, only the slope factor mean had an absolute relative bias value larger than the suggested cutoff value. Relative bias increased with increases in the proportion of non-invariant items and the magnitude of non-invariance.

Leite (2007) examined the ability of the LGM to recover parameter estimates, standard errors, chi-square statistics, and adequate fit indices when composites of the observed variables (i.e., mean scores) are used as input variables. The author compared the performance of traditional LGM, LGM with fixed error variances, and a second-order LGM (which he refers to as the "curve-of-factors model"). The author varied the number of measurement occasions, the number of items, sample size, item types (i.e., essentially congeneric or essentially tau-equivalent), reliability, and level of non-invariance. For the levels of non-invariance configural, weak, or strict invariance was simulated. The univariate LGMs generally produced biased parameter estimates, but unbiased standard errors. The curve-of-factors models produced unbiased estimates in all conditions, but required larger sample sizes for accurate chi-square and fit indices.

Wirth (2008) conducted a study examining the roles of measurement invariance in studying stability and growth. The author varied sample size, time-adjacent unique factor correlations, the patterns of non-invariance (e.g., invariant loadings and intercepts, invariant loadings and non-invariant intercepts, etc.). Mean score and factor score models were used to examine the impact of invariance violations on LGMs. For the mean score models, the observed mean for the variables at each measurement occasion was used as input for the LGM. For the factor score models, factor scores were calculated (regression-based or constrained-covariance) and used as input for the LGM. In addition, the LGMs were either specified to have linear growth (i.e., the loadings for the slope factor were constrained) or to have freely estimated growth (i.e., loadings for the slope factor were freely estimated). The results suggest that non-invariant measurement structures consistently led to biased estimates of almost all parameters for both mean score and factor score approaches. In addition, the author noted that the presence of non-invariance led to estimates of non-linear growth trajectories. Finally, the author found that the use of factor scores in LGMs led to biased fit statistics.

In sum, the research suggests that ignoring measurement invariance tends to lead to biased estimates and fit statistics. While some of the applied studies suggest minimal impact on substantive results when invariance is violated, all of the simulation studies found that invariance violations may have an impact on substantive results. Results suggest that intercept invariance violations had more of an impact on composite mean differences than loading invariance violations. Ultimately, the research on the impact of invariance violations is relatively scant. Within the realm of invariance violations research there are even fewer studies examining invariance violations for dichotomously scored data in longitudinal contexts. The current study aims to add to this body of literature and help provide guidelines for handling longitudinal measurement invariance violations.

Current Study

The current study aims to examine the impact varying levels of measurement invariance has on LGMs. This simulation study adds a systematic investigation of the impact invariance violations have on LGMs to a relatively small body of research. A simulation study allows researchers to know and vary truth to clearly observe the situations in which invariance violations may and may not be problematic. In addition, unlike the study by Olivera-Aguilar (2013), the current study examines the impact of invariance violations with dichotomous items, reflecting common measurement practice in educational settings. Chen (2008) noted that the impact of measurement invariance violations on substantive tests has not been thoroughly examined in categorical or dichotomous variables. Similar to Wirth (2008), the current study examines the potential for invariance violations to impact the estimated shape of change over time (i.e., linear vs. quadratic).

This study investigates the potential issues that may arise if partial measurement invariance is ignored and observed total scores (without modeling the invariant parameters) at each measurement occasion are used in a latent growth model. Several models are specified to examine the impact of varying degrees of longitudinal measurement invariance violations. Two first-order LGMs, two second-order LGMs, and two second-order non-linear latent growth models (NLGM) were specified. The first order models use observed scores to model change. One second-order LGM and one second-order NLGM constrain item parameters to be equivalent across measurement occasions (i.e., assuming strong measurement invariance). The other second-order LGM and NLGM allow non-invariant item parameters to be freely estimated across measurement occasions (i.e., explicitly modeling partial measurement invariance). As many researchers feel that small violations of measurement invariance (i.e. partial measurement invariance) are acceptable, the current study aims to examine how much measurement non-invariance is too much and whether modeling it provides more accurate estimates of growth. The type of invariance violation (weak or strong), the proportion of non-variant items, the size of the invariance violation, and the total length of the assessment will be varied to examine if and under which conditions growth models are robust to invariance violations. The study aims to answer five general research questions outlined below.

Research Questions

1. To what extent do varying degrees of strong measurement invariance violations impact the estimated parameters of a latent growth model?
2. To what extent do varying degrees of weak measurement invariance violations impact the estimated parameters of a latent growth model?
3. To what extent do varying degrees of strong measurement invariance violations impact the shape of the growth curve (i.e., linear, quadratic) in a latent growth model?

4. To what extent do varying degrees of weak measurement invariance violations impact the shape of the growth curve (i.e., linear, quadratic) in a latent growth model?
5. Does modeling invariance violations (i.e., modeling partial measurement invariance) result in more accurate results than those observed in Research Question #1 – Research Question #4?

CHAPTER III

METHODS

The current study aims to examine the extent to which varying degrees of longitudinal measurement invariance violations impact growth model parameters. The current section outlines the data simulation design, including each of the conditions examined. The models estimated are presented. Finally, the criteria by which the results were evaluated are described.

Simulation Design

A first-order LGM was used to simulate individual theta values. The theta values for the first measurement occasion (the intercept of the growth model) follow a normal distribution with a mean of -1.0 and a variance of 1, $N(-1.0,1.0)$. The growth parameters were simulated such that the linear growth was normally distributed with a mean of 0.5 and a variance of 0.2, $N(0.5, 0.2)$. Muthén and Muthén (2002) note that the ratio between the intercept and slope variance is commonly 5 to 1. Given that the intercept variance was set to 1.0, the slope variance was set to 0.2. To simulate variance around individual trajectories, a normally distributed error term with a mean of 0.0 and a variance of 1.5, $N(0.0,1.5)$, was added to each individual's simulated theta value at each time point. The intercept, magnitude of growth, and variability in growth and intercept parameters did not vary across conditions. Item level responses for the thetas were then generated from a 2PL IRT model. Data were simulated for 3000 simulees on a dichotomously scored assessment at four measurement occasions. Simulating data for 3000 individuals allowed for accurate and precise estimation of parameters, thus any errors should be attributable to model misfit, rather than sampling error. Four measurement occasions allowed for linear and non-linear (i.e., quadratic) slopes to be estimated for the LGM. Invariant item intercept parameters were drawn from a uniform distribution ranging from -2.0 to 2.0. Item loading parameters were drawn from a uniform distribution ranging from 0.6 to 1.5. These values represent typical IRT item parameter values and ensure that the addition of DIF will not result in extreme item parameters. Twenty-five replications within each condition were simulated. The conditions varied within the study are outlined in Table 2 below.

Table 2

Conditions Varied in the Study

Condition	**Levels**
Percent of non-invariant items	0%, 15%, 30%, 45%
Magnitude of invariance violation	Loading: 0.2, 0.3, 0.4
	Intercept: 0.4, 0.7, 1.0
Type of invariance violation	Loading, Intercept, Both
Test length	Short (20 items), Long (40 items)

The conditions above are similar to those from several other studies examining the impact of invariance violations. The literature has explored several conditions that may influence the impact of invariance violations in group comparisons or latent growth models. The conditions previously explored include sample size (Chen, 2008; Leite, 2007; Olivera-Aguilar, 2013; Steinmetz, 2013; Wirth, 2008), number of items on the assessment (Leite, 2007; Olivera-Aguilar, 2013; Steinmetz, 2013), proportion of non-invariant items (Chen, 2008; Olivera-Aguilar, 2013; Steinmetz, 2013), magnitude of invariance violations in intercepts (Olivera-Aguilar, 2013; Steinmetz, 2013), magnitude of invariance violations in loadings (Olivera-Aguilar, 2013), and patterns of invariance violations (Chen, 2008; Wirth, 2008). Many of these studies, however, examined the impact of invariance violations using continuous variables. Little work has considered the impact of invariance violations in dichotomous indicators on growth models. In addition, only two studies examined how violations of invariance in intercepts *and* loadings simultaneously influence the estimation of growth models(Leite, 2007; Wirth, 2008). Several studies have tangentially examined test information changes on group comparisons and growth models. While no study explicitly examined changes in test information, several of the studies examined systematically decreasing loadings and intercepts, which may also be interpreted as systematic decreases in information. Given the limitations in the literature related to invariance, other factors were given less emphasis in the current research. Instead, the focus was placed on the impact of various invariance violation scenarios.

Proportion of Non-Invariant Items. Several studies have included the proportion of non-invariant items in their investigation of invariance violations (Chen, 2008; Olivera-Aguilar, 2013; Steinmetz, 2013). The results of the studies suggest that the proportion of non-invariant items impacts comparisons between groups and growth models. Notably, the studies generally used a small number of continuous items, ranging between 4 and 15, more commonly seen in psychology literature. In the current study, the levels for the proportion of non-invariant items condition are 0%, 15%, 30%, or 45%.

While one study did note that 26 of the 97 comparisons in the author's literature review had 90% or more of the loadings higher for one group than another, this did not seem to be a realistic condition for an educational assessment context.

Magnitude of Invariance Violations. Item loadings and intercepts were simulated to either be invariant across measurement occasions, or to have some proportion of items with small, medium, or large violations of invariance across measurement occasions. Olivera-Aguilar (2013) described a process to calculate systematic decreases in item intercepts over time. In this process, the total amount of change in a parameter (i.e., the degree of the invariance violation) is divided by the number of measurement occasions. This approach represents a gradual change in item parameters across time. In this study, the difference between the first and fourth measurement occasion will equal the small, medium, or large invariance violation specified. Nye (2011) outlined small, medium, and large values of DIF for item intercepts and slopes based on previous simulation research. Based on his results, the small, medium, and large DIF magnitudes for item intercepts were simulated to be 0.4, 0.7, and 1.0 logits, respectively. Small, medium, and large DIF magnitudes for item slopes were 0.2, 0.3, and 0.4, respectively.

Type of Invariance Violation. Researchers have examined the impact of weak (i.e., item loading) and strong (i.e., item intercept) invariance violations, separately (Olivera-Aguilar, 2013; Steinmetz, 2013). These studies tend to suggest that violations of strong (intercept differences only) measurement invariance are more detrimental than violations of weak (loading differences only) measurement invariance violations. Only two studies included a condition in which intercepts *and* loadings were non-invariant across measurement occasions (Leite, 2007; Wirth, 2008). Wirth (2008) notes that when all items are constrained to be equal, despite partially invariance loadings and intercepts, the observed trajectories were non-linear. The type of invariance violation condition included weak, strong, and a combination of weak and strong invariance violations to further examine the relative importance of the type of violation in growth modeling contexts.

Test Length. Several researchers have examined invariance violations with varying test lengths (Leite, 2007; Olivera-Aguilar, 2013; Steinmetz, 2013). Leite (2007) found that with more items some fit indices suggested a relatively well-fitting model, regardless of invariance violations. Steinmetz (2013) found that having fewer items may increase the chance of finding spurious differences. Notably, in the studies discussed, the total number of items is related to the proportion of non-invariant items. In these studies, the number of non-invariant items was fixed. Fixing the number of non-invariant items and varying the test length effectively changes the proportion of non-invariant items. The current study includes a preliminary investigation of whether having more items in total, but an equivalent proportion of non-invariant items, leads to differences in growth model results. Test lengths of 20 items and 40 items were chosen such that they represent typical test lengths for academic assessments and allow for integer values when combined with the proportion of non-invariant items condition. The 20 item exam was crossed with all conditions. Time constraints on estimation limited the number of conditions for which the

40 item exam could be examined. The conditions in which we examined the 40 item exam were chosen based on the results of the 20 item exam.

Modeling Approaches

All analyses were estimated in Mplus and results were summarized in R. Six modeling approaches were used in the proposed study. For four approaches a second-order LGM was estimated. The first model constrained the intercepts and loadings to be equal across all measurement occasions (2LGMC). This model suggests a situation in which the latent variables are used in the LGM without testing for invariance across measurement occasions. Second, a second-order LGM was specified such that non-invariant item parameters are freely estimated and invariant item parameters were constrained to be equivalent across measurement occasions (2LGMF). This model simulates a situation in which measurement invariance has been examined and partial measurement invariance is incorporated into the model. This model is the "true" model and would, ideally, perform best. Third, a non-linear (quadratic) second-order LGM was estimated. In this model, like the first model, the item parameters were constrained to be equivalent across measurement occasions (2NLGMC). Wirth (2008) found that the presence of invariance violations may lead to spurious non-linear trajectories. This model allows for further examination of this phenomenon. Finally, the fourth model estimated was a non-linear (quadratic) second-order LGM in which non-invariant item parameters were freely estimated and invariant item parameters were constrained to be equivalent across measurement occasions (2NLGMF). To identify the second-order models, the mean of the growth intercept factor and the variance of the time one measurement model factor were set to their true values (-1.0 and 1.5, respectively). Other forms of model identification were possible and perhaps more traditional. This form of identification was chosen because it allowed for interpretation of results on the metric of the simulation and did not change overall model fit from what would have been found with alternative forms of identification. A summary of the four second-order modeling approaches is provided in Table 3 below.

Table 3

Summary of Second-Order Latent Growth Model Approaches

		Item Constraints	
		All Items Equal	*Invariant Equal, Non-Invariant Free*
Trajectory	*Linear*	Model 1	Model 2
	Quadratic	Model 3	Model 4

In addition to the four second-order LGMs specified above, two first-order LGMs were estimated. The first model estimated linear growth (LGM), whereas the second estimated non-linear (quadratic) growth (NLGM). Identifying the second-order models with the growth intercept mean and time one factor variance made it difficult to replicate in the first-order models. In order to ensure the first-order models were identified and scaled commensurate with the second-order models, the sum score at each time point was treated as the single indicator for a time factor, as shown in Figure 5. The loadings and intercepts for the sum score on each time factor were constrained to be equal. The residual variance of each single indicator was constrained to be zero (i.e., all the variability in the sum score is explained by the time factor). As in the second order models, the intercept factor mean and time one factor variance were fixed to their true values (-1.0 and 1.5, respectively). Specifying the first order models this way allows the first order models to be directly compared to the second-order models. In addition, the fit of the single-indicator factor model fits identically to a first-order growth model with the observed sum score at each time point (Figure 1). Table 4 displays the fit statistics for one replication of each first-order model specification approach and demonstrates the equivalency across the two approaches. All second-order models were estimated using weighted least squares means and variance adjusted (WLSMV) and first-order models were estimated with maximum likelihood (ML).

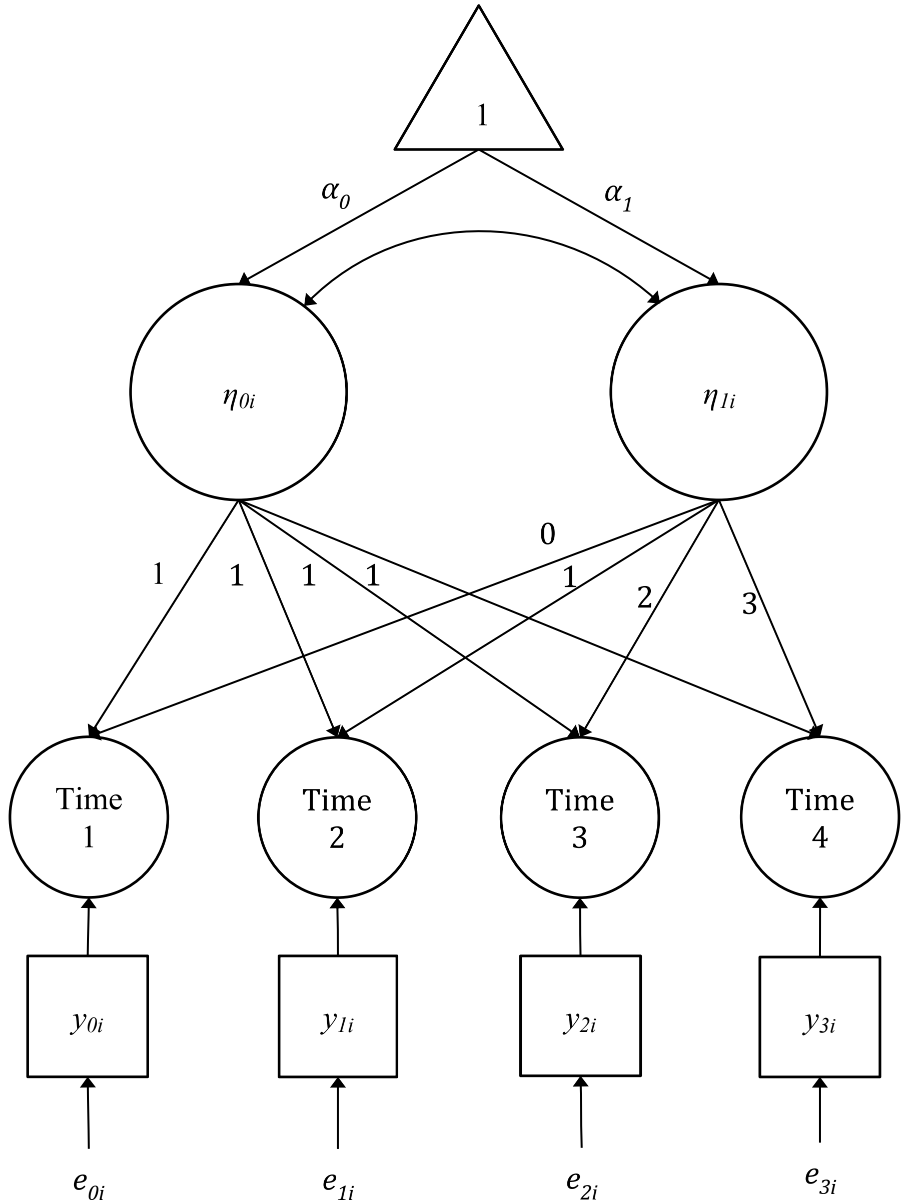

Figure 5. Single Indicator Latent Growth Model.

Table 4

Fit for Observed Sum Score and Single-Indicator Factor Specifications of First-Order Model

	X^2	Degrees of Freedom	*p*-value	RMSEA	CFI
Observed Sum Score	8.652	5	0.1238	0.016	0.998
Single-Indicator Factor	8.652	5	0.1238	0.016	0.998

Evaluation Criteria

Several measures were used to judge the acceptability of model results from the six modeling approaches within each condition. Model fit was assessed using a combination of model fit criteria. Several model fit indices were used to evaluate model fit. Each index was chosen because it adds a unique perspective as to how well the model fits the data. Taken together, the indices provide a relatively holistic view of model-data fit. The first index is the χ^2 statistic. The χ^2 statistic examines the exact differences between the observed and model-implied covariance matrices. A non-significant χ^2 would indicate that the model fits the data well and that the model-implied covariance matrix is not significantly different than the observed covariance matrix. Notably, the χ^2 statistic is sensitive to sample size and thus may reject models with small differences between the observed and reproduced covariance matrices due to the large sample.

The root mean square error of approximation (RMSEA) is an absolute fit index sensitive to misspecification of factor loadings and detects misfit solely due to model misspecification, not due to random sampling error. It provides an estimate of model misspecification per degree of freedom, and values of 0.06 or less are encouraging (Hu &

Bentler, 1999). The comparative fit index (CFI) compares the fit of the proposed model to a baseline, or null, model in which all paths are set to be zero. In other words, the CFI indicates how much model fit improves when our model is compared to no model. This index ranges from 0 to 1 but is different than the other indices as larger values, 0 .95 or greater, indicate adequate model-data fit (Hu & Bentler, 1999).

Fit was compared between nested models. Nested models are constrained versions of another model. For example, Model 1 is nested within Model 2 such that if the non-invariant parameters in Model 2 were constrained to be equal within item across measurement occasions, Model 2 would equal Model 1. For nested models, the chi-square test can be used to examine whether the more complex model provides significant improvement in model fit.

To examine the extent to which model parameters are accurately and precisely recovered, relative bias and root mean square error (RMSE) were evaluated for the growth model factor means (i.e., intercept and slope) and variances. The variance for the intercept factor represents whether all individuals score similarly at the initial time point or whether they vary in their scores. The slope variance describes whether individuals change similarly (i.e., little to no variability) or differently (i.e., a fair amount of variability). Relative bias is calculated by subtracting the true value from the estimated value and dividing the resulting difference by the true value, as shown in Equation 25 below (Hoogland & Boomsma, 1998). Relative bias puts the bias on a percentage metric and makes comparing across conditions straightforward. Hoogland and Boomsma (1998) suggest that, for parameter estimates, relative bias values of 0.05 or less are considered acceptable.

$$Relative\ Bias = \frac{\hat{\eta} - \eta}{\eta} \tag{25}$$

In Equation 25, η represents the true value and $\hat{\eta}$represents the estimated value. The RMSE takes into account the bias and variability of estimates and is calculated using the formula in Equation 26 below.

$$RMSE = \sqrt{(\hat{\eta} - \eta)^2} \tag{26}$$

Finally, to examine the parameters estimated in the over-parameterized models (i.e., the non-linear models), the significance tests for the incorrectly modeled parameters were examined. A significant estimate for a parameter that was not included in the generating model (e.g., a quadratic slope) would suggest that the model provided results that could lead to incorrect conclusions about growth (i.e., Type I error).

The type, size, and proportion of non-invariant items impact the information function for the assessment across measurement occasions. These changes in information should be considered when evaluating the results. While the issue of test information cannot be untangled from the conditions varied in the study, it can help add context to the results.

CHAPTER IV

RESULTS

Convergence

Each replication was checked for convergence before analyzing results. A summary of convergence issues is provided in Table 5. None of the linear growth models (LGM, 2LGMC, and 2LGMF) had issues converging on a solution. All of the non-linear growth models (NLGM, 2NLGMC, and 2NLGMF), however, frequently had convergence issues. More specifically, most of the convergence issues were due to a non-positive definite PSI matrix. Convergence concerns were prevalent in all conditions for the non-linear models and were seemingly unrelated to any particular condition varied in the study.

Table 5

Summary of Non-Convergence Issues Across All Six Growth Modeling Approaches

Loading DIF	Intercept DIF	Percent of DIF	LGM	NLGM	2LGMC	2NLGMC	2LGMF	2NLGMF
None	None	N/A	0	11	0	14	0	14
None	Small	15%	0	11	0	13	0	13
None	Small	30%	0	10	0	13	0	13
None	Small	45%	0	10	0	12	0	11
None	Medium	15%	0	11	0	14	0	14
None	Medium	30%	0	9	0	11	0	11
None	Medium	45%	0	6	0	8	0	9
None	Large	15%	0	9	0	13	0	13
None	Large	30%	0	8	0	13	0	14
None	Large	45%	0	7	0	12	0	12
Small	None	15%	0	11	0	12	0	12
Small	None	30%	0	12	0	13	0	13
Small	None	45%	0	16	0	18	0	17
Small	Small	15%	0	9	0	13	0	12
Small	Small	30%	0	10	0	16	0	16
Small	Small	45%	0	12	0	16	0	16
Small	Medium	15%	0	11	0	13	0	13
Small	Medium	30%	0	11	0	13	0	13
Small	Medium	45%	0	10	0	16	0	16
Small	Large	15%	0	10	0	12	0	12
Small	Large	30%	0	8	0	11	0	10
Small	Large	45%	0	13	0	18	0	18
Medium	None	15%	0	10	0	13	0	13
Medium	None	30%	0	10	0	13	0	13
Medium	None	45%	0	14	0	17	0	17
Medium	Small	15%	0	9	0	10	0	10
Medium	Small	30%	0	7	0	8	0	7
Medium	Small	45%	0	10	0	12	0	12
Medium	Medium	15%	0	14	0	18	0	18
Medium	Medium	30%	0	13	0	18	0	18
Medium	Medium	45%	0	9	0	12	0	12
Medium	Large	15%	0	14	0	16	0	16
Medium	Large	30%	0	8	0	10	0	11
Medium	Large	45%	0	9	0	13	0	13
Large	None	15%	0	14	0	16	0	16
Large	None	30%	0	9	0	15	0	15
Large	None	45%	0	12	0	14	0	15
Large	Small	15%	0	6	0	11	0	11
Large	Small	30%	0	12	0	13	0	13
Large	Small	45%	0	13	0	17	0	18
Large	Medium	15%	0	14	0	16	0	16

Large	Medium	30%	0	10	0	14	0	15
Large	Medium	45%	0	11	0	13	0	13
Large	Large	15%	0	12	0	13	0	13
Large	Large	30%	0	13	0	17	0	17
Large	Large	45%	0	12	0	15	0	15

Model Fit

Model fit statistics were used to examine how well each model fit the data. The chi-squared statistic, RMSEA, and CFI estimates are presented for each modeling approach in the graphs below. In addition, nested model chi-squared statistics are presented.

Chi-Squared. A significant chi-squared *p*-value suggests that the model-implied covariance matrix is significantly different than the observed covariance matrix. A significant difference between the two matrices suggests that the model does not adequately fit the data. The *p*-values for the chi-squared statistics are presented in Figures 6-11. A red line at 0.05 represents the typical critical value to which chi-squared *p*-values are compared. The mean and standard deviation values for the *p*-values are included in Appendix A. For the first order linear model, results exhibited a wide range of *p*-values within all conditions. The chi-square *p*-values for the first-order linear model were more variable in conditions where there is a small percent of DIF and/or small intercept DIF. Conditions with a small percent of DIF and/or small intercept DIF had several *p*-values that suggested that the chi-square was non-significant (i.e., the model fit the data). Aggregating across the loading DIF size levels, the chi-squared *p*-values for the condition with 15% of items with small intercept DIF were significant 65% of the time. Conversely, conditions with substantial (in proportion and size) intercept DIF consistently had chi-squared values that suggested the model did not fit the data well. Aggregating across the loading DIF size levels, the chi-squared *p*-values for the condition with 45% of items with large intercept DIF were significant 95% of the time. For the second-order linear modeling approaches, the median chi-squared *p*-values all suggested that the models did not fit the data (i.e., $p < 0.05$). The chi-squared *p*-values for the first-order non-linear model were unrelated to the conditions varied in the study and were often not significant, suggesting the model fit the data well. On average, across all levels of all conditions, the chi-squared *p*-value for the first-order non-linear model was only significant 11% of the time. For the second-order non-linear modeling approaches, the median chi-squared *p*-values all suggested that the models did not fit the data (i.e., $p < 0.05$). Notably, the chi-squared statistic is known to be sensitive to sample size and should be interpreted with caution.

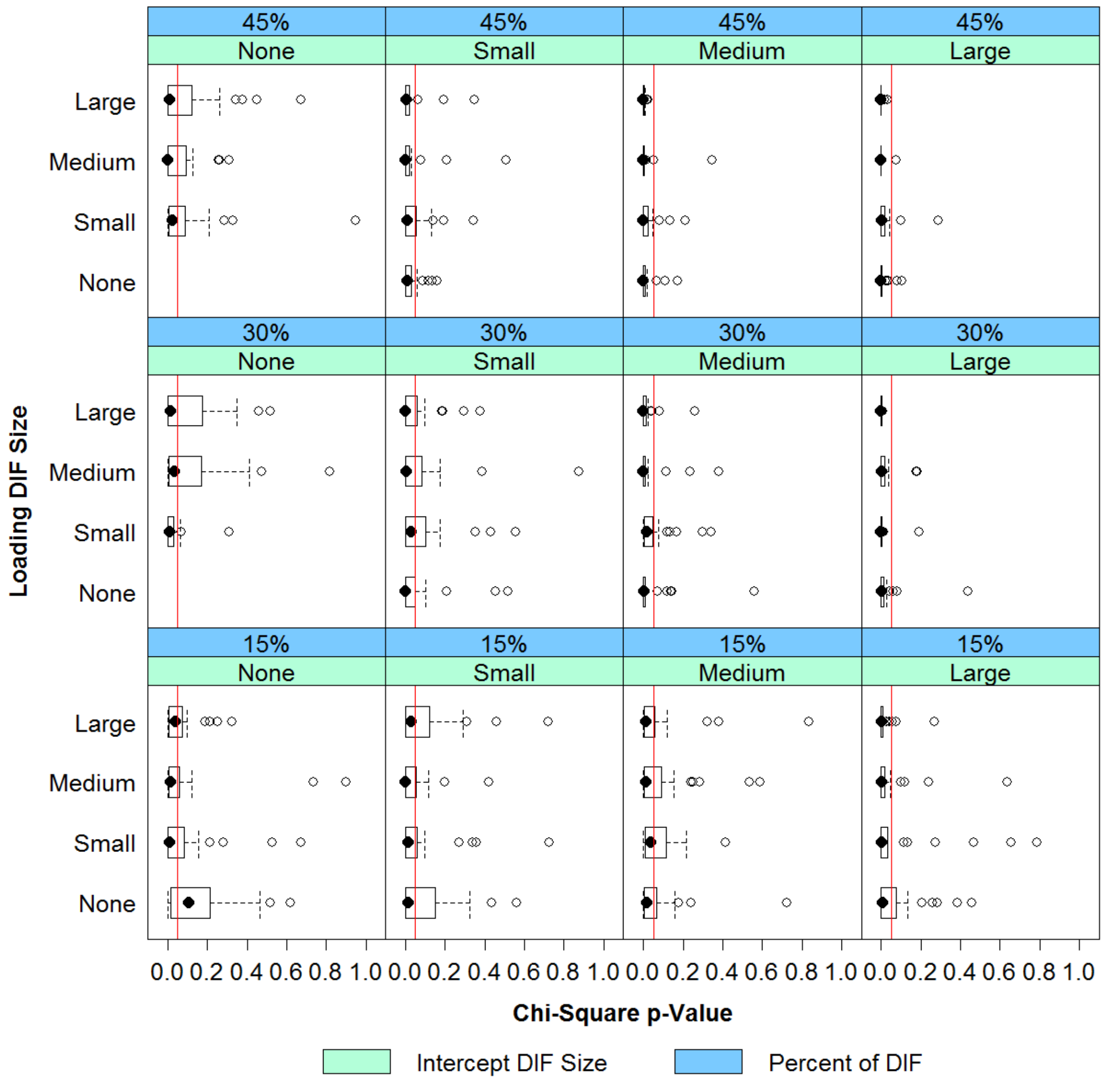

Figure 6. Chi-Squared *p*-Values for the First-Order Latent Growth Model.

Figure 7. Chi-Squared *p*-Values for the Constrained Second-Order Latent Growth Model.

Figure 8. Chi-Squared *p*-Values for the Free Second-Order Latent Growth Model.

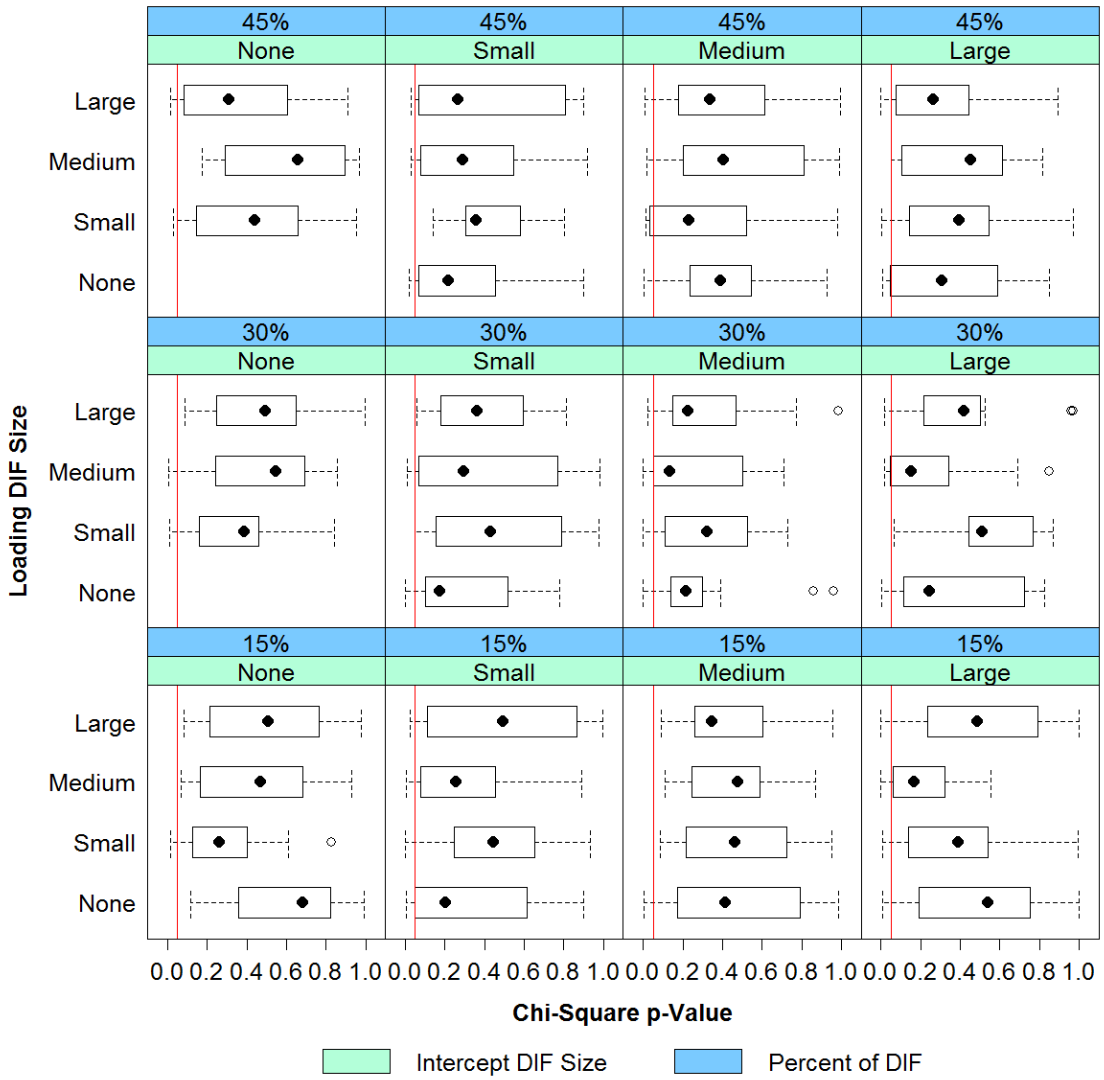

Figure 9. Chi-Squared *p*-Values for the First-Order Non-Linear Latent Growth Model.

Figure 10. Chi-Squared *p*-Values for the Constrained Second-Order Non-Linear Latent Growth Model.

Figure 11. Chi-Squared *p*-Values for the Free Second-Order Non-Linear Latent Growth Model.

Root Mean Square Error of Approximation. The RMSEA estimates for all modeling approaches are presented in Figures 12-17. A red line at 0.06 represents the typical value to which RMSEA values are compared. The mean and standard deviation values for the RMSEA are included in Appendix B. As a reminder, lower RMSEA values represent better model-data fit. For all modeling approaches and all conditions, the RMSEA values were below the suggested cutoff for RMSEA suggesting that all models fit the data well. The RMSEA values for all modeling approaches and all conditions had very little variability. The first-order models (linear and non-linear) were more variable than the second-order models.

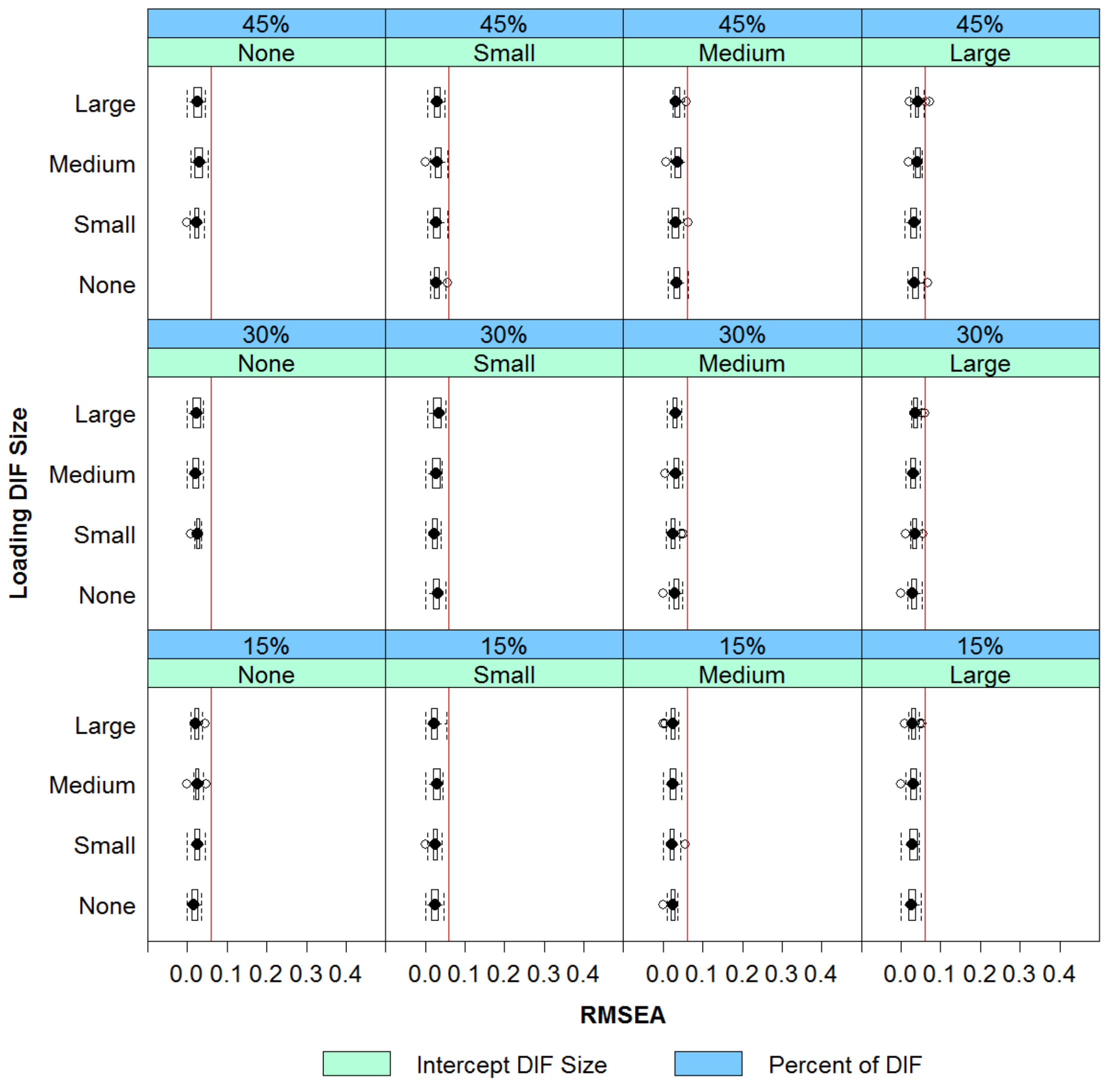

Figure 12. RMSEA Values for the First-Order Latent Growth Model.

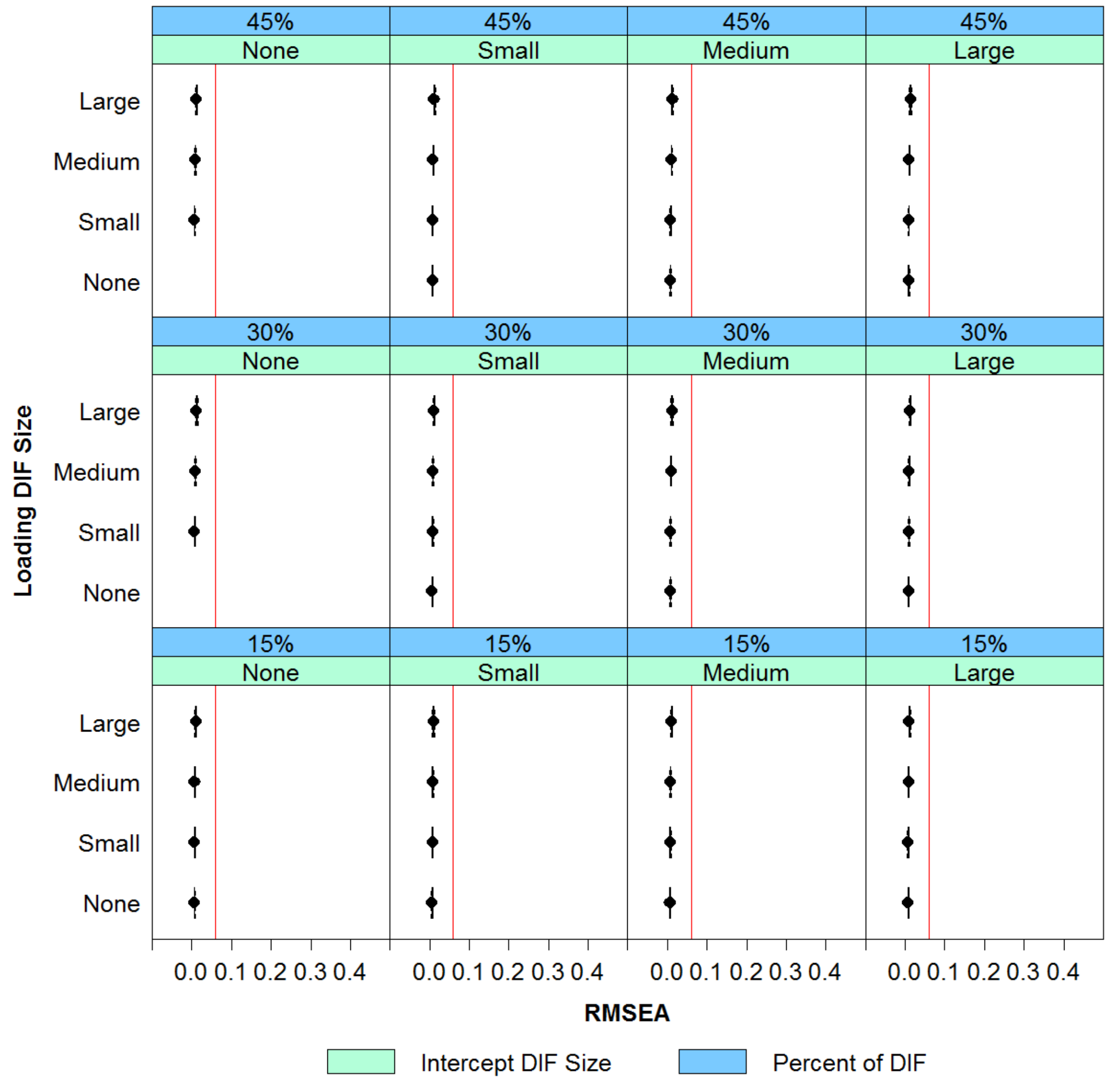

Figure 13. RMSEA Values for the Constrained Second-Order Latent Growth Model.

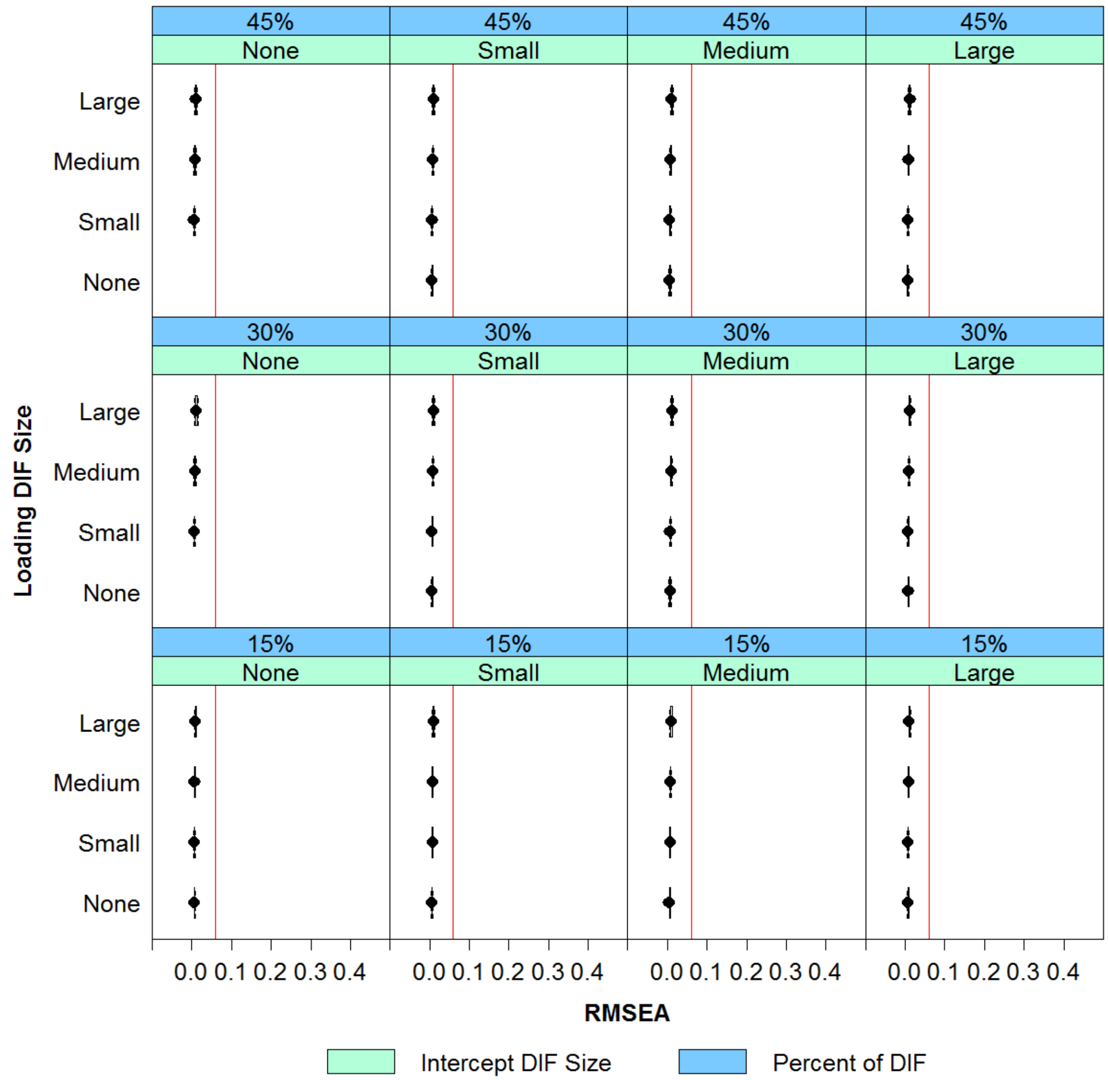

Figure 14. RMSEA Values for the Free Second-Order Latent Growth Model.

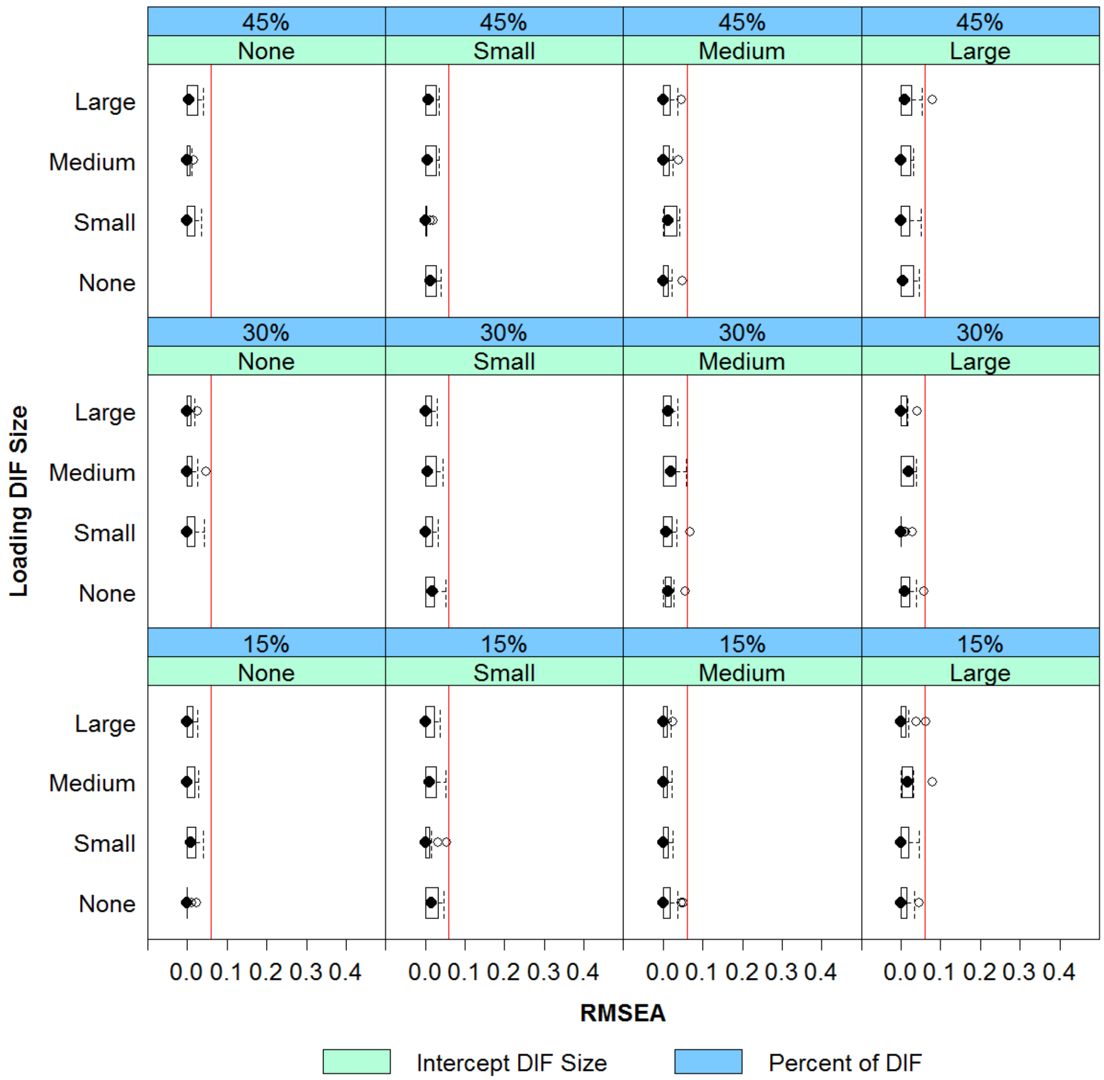

Figure 15. RMSEA Values for the First-Order Non-Linear Latent Growth Model.

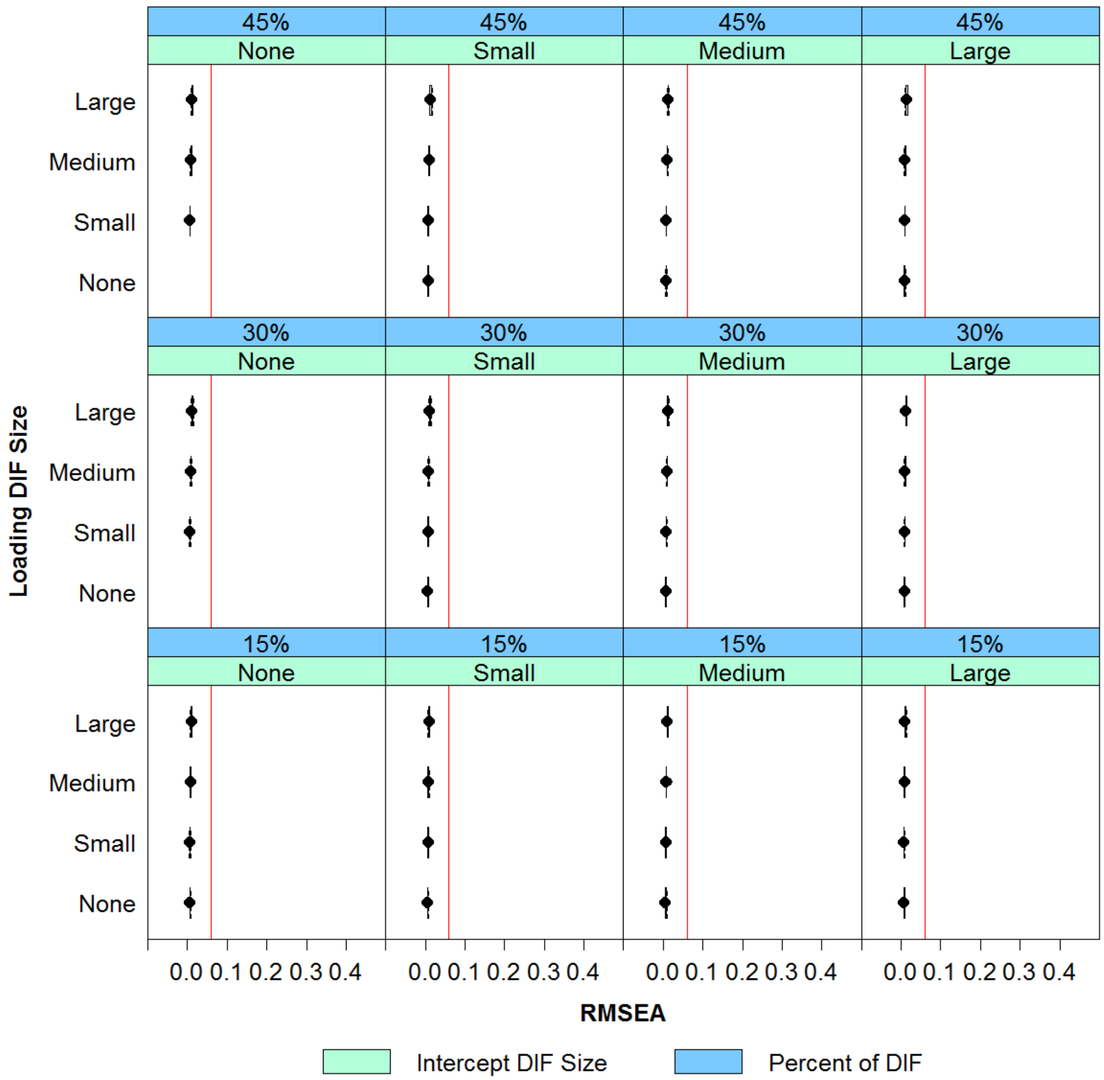

Figure 16. RMSEA Values for the Constrained Second-Order Non-Linear Latent Growth Model.

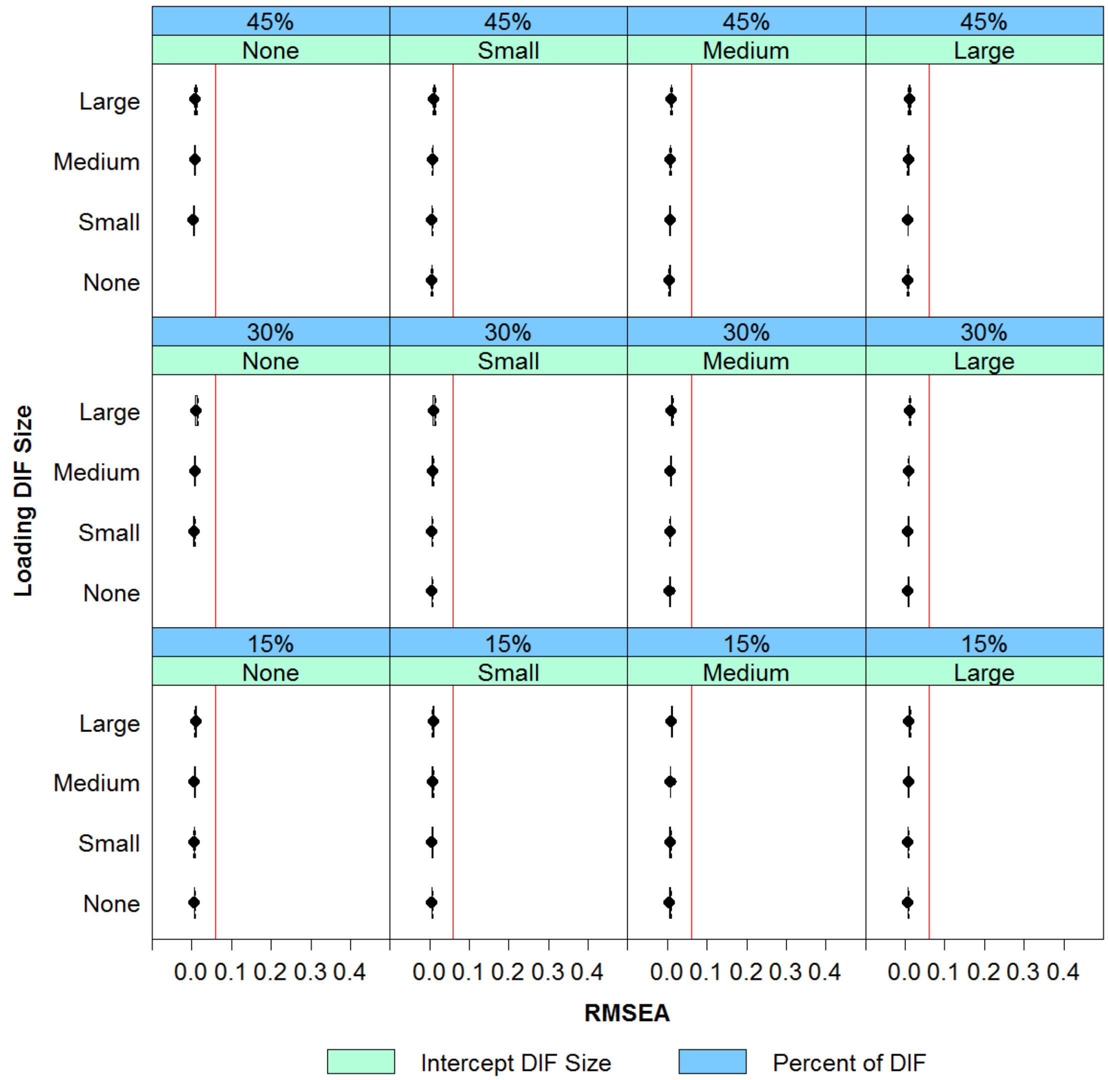

Figure 17. RMSEA Values for the Free Second-Order Non-Linear Latent Growth Model.

Comparative Fit Index. The CFI estimates for all modeling approaches are presented in Figures 18-23. A red line at 0.95 represents the typical value to which CFI values are compared. The mean and standard deviation values for the CFI are included in Appendix C. As a reminder, higher CFI values represent better model-data fit. The CFI values for all modeling approaches and all conditions were well above the cutoff value of

0.95 suggesting that all models fit the data well. In addition, the CFI values had very little variability across replications.

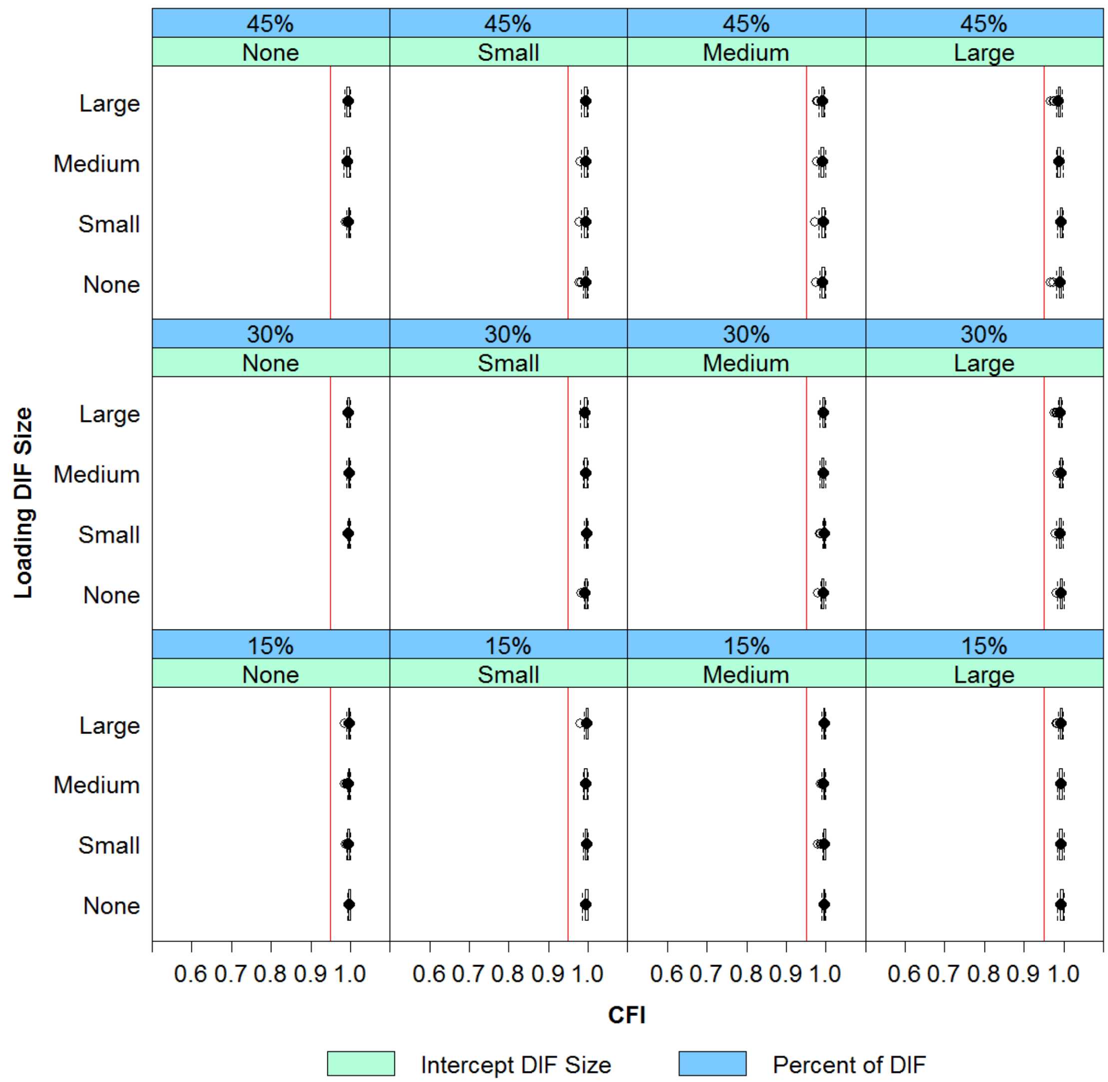

Figure 18. CFI Values for the First-Order Latent Growth Model.

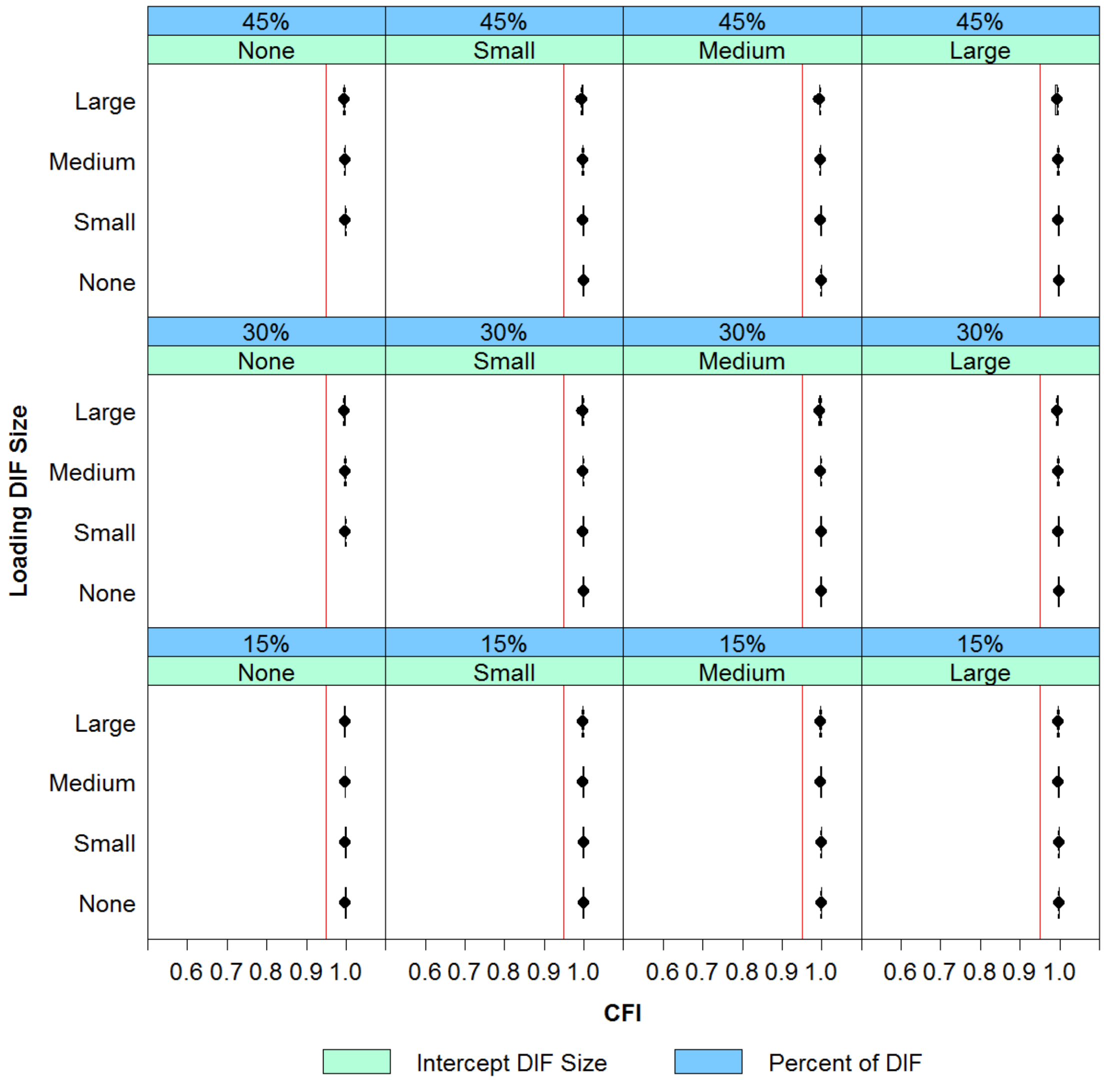

Figure 19. CFI Values for the Constrained Second-Order Latent Growth Model.

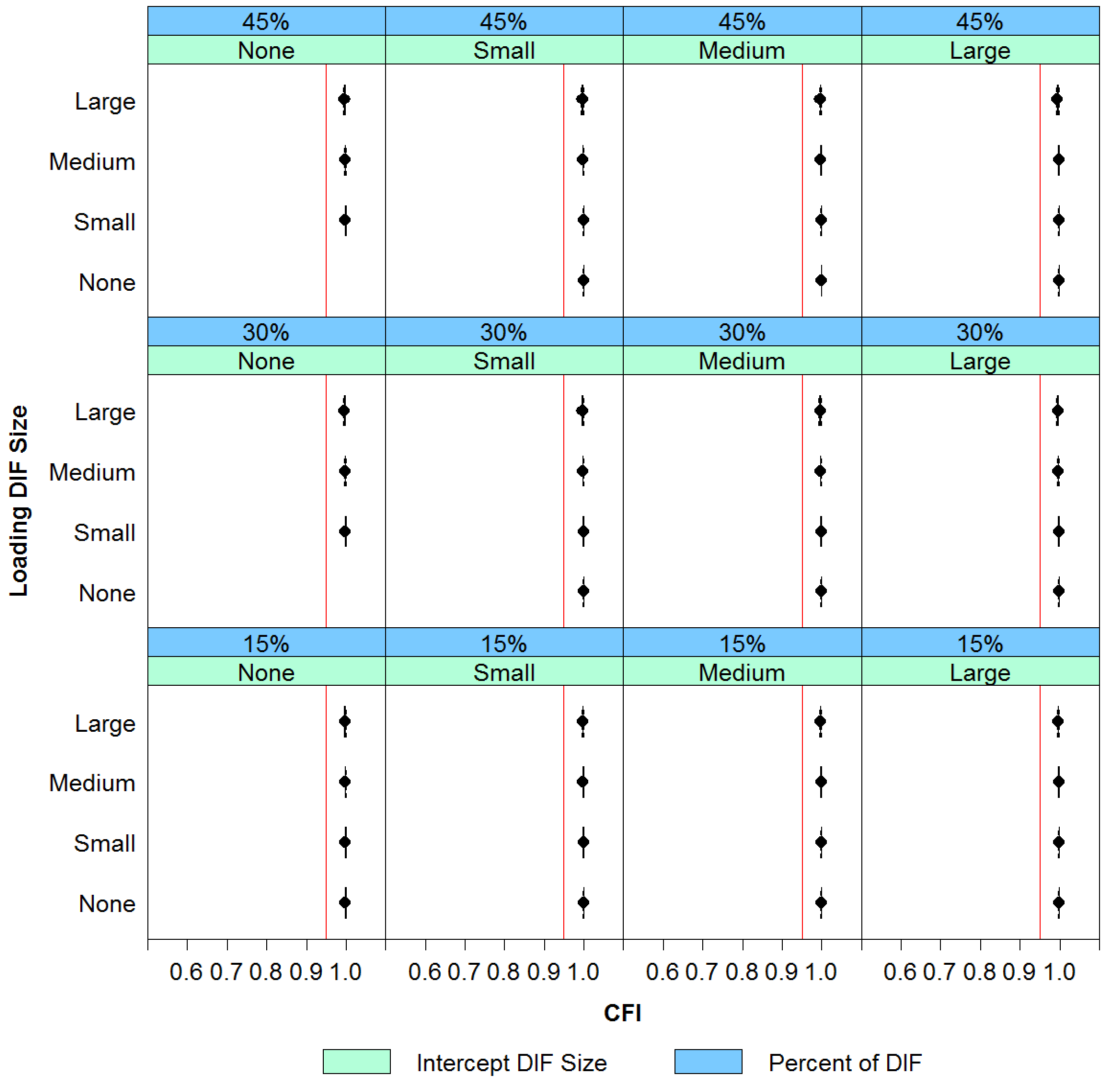

Figure 20. CFI Values for the Free Second-Order Latent Growth Model.

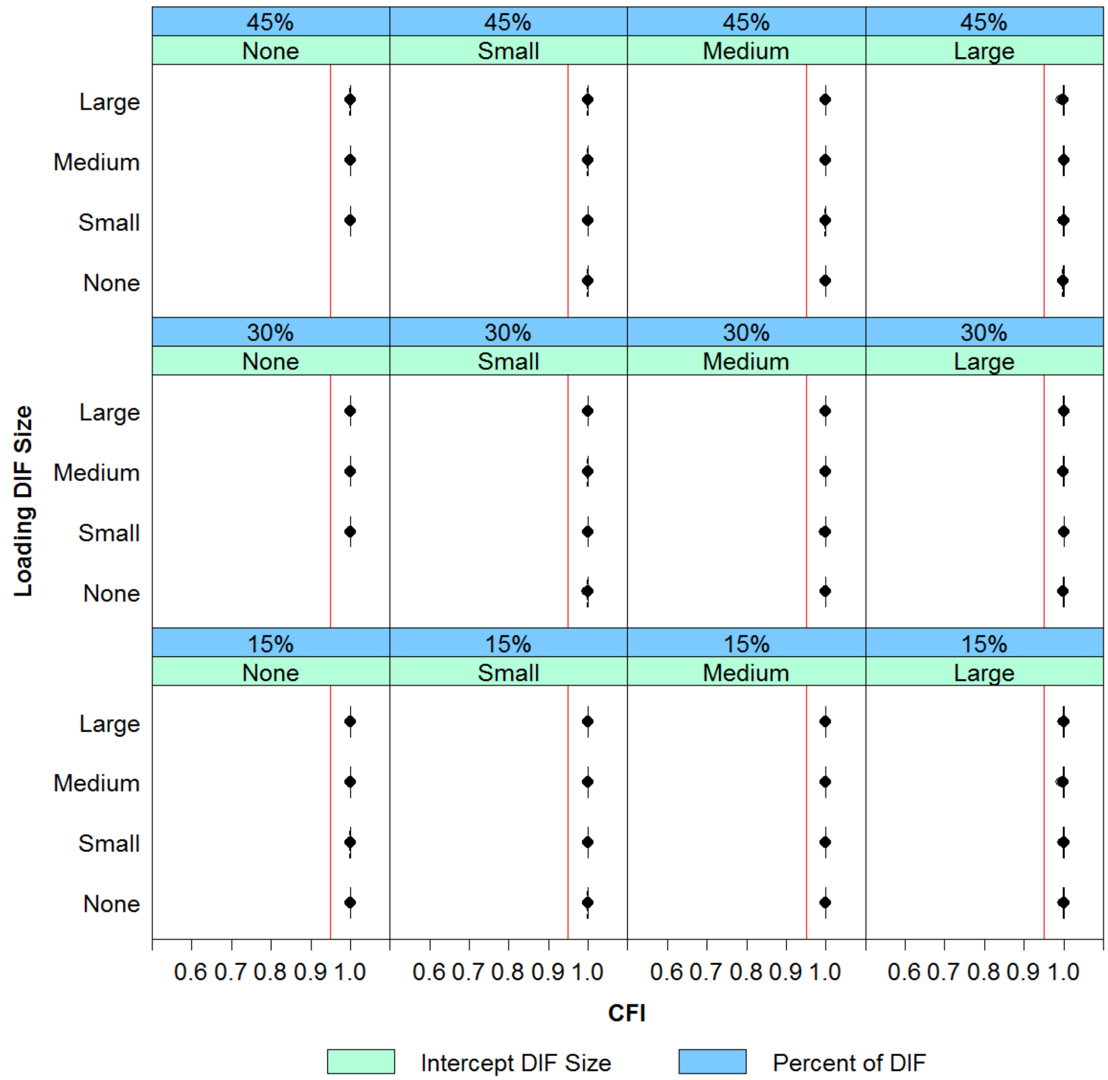

Figure 21. CFI Values for the First-Order Non-Linear Latent Growth Model.

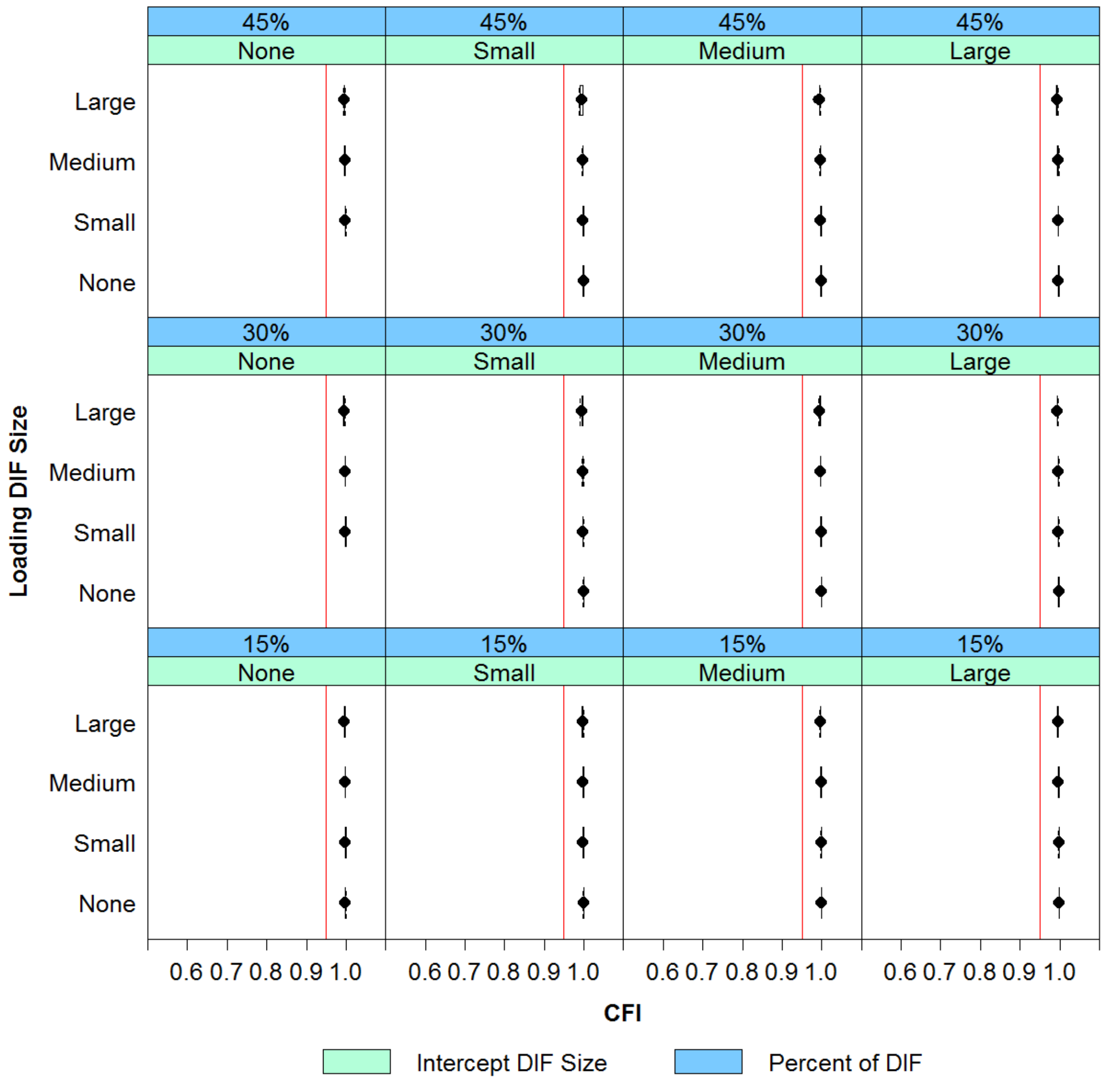

Figure 22. CFI Values for the Constrained Second-Order Non-Linear Latent Growth Model.

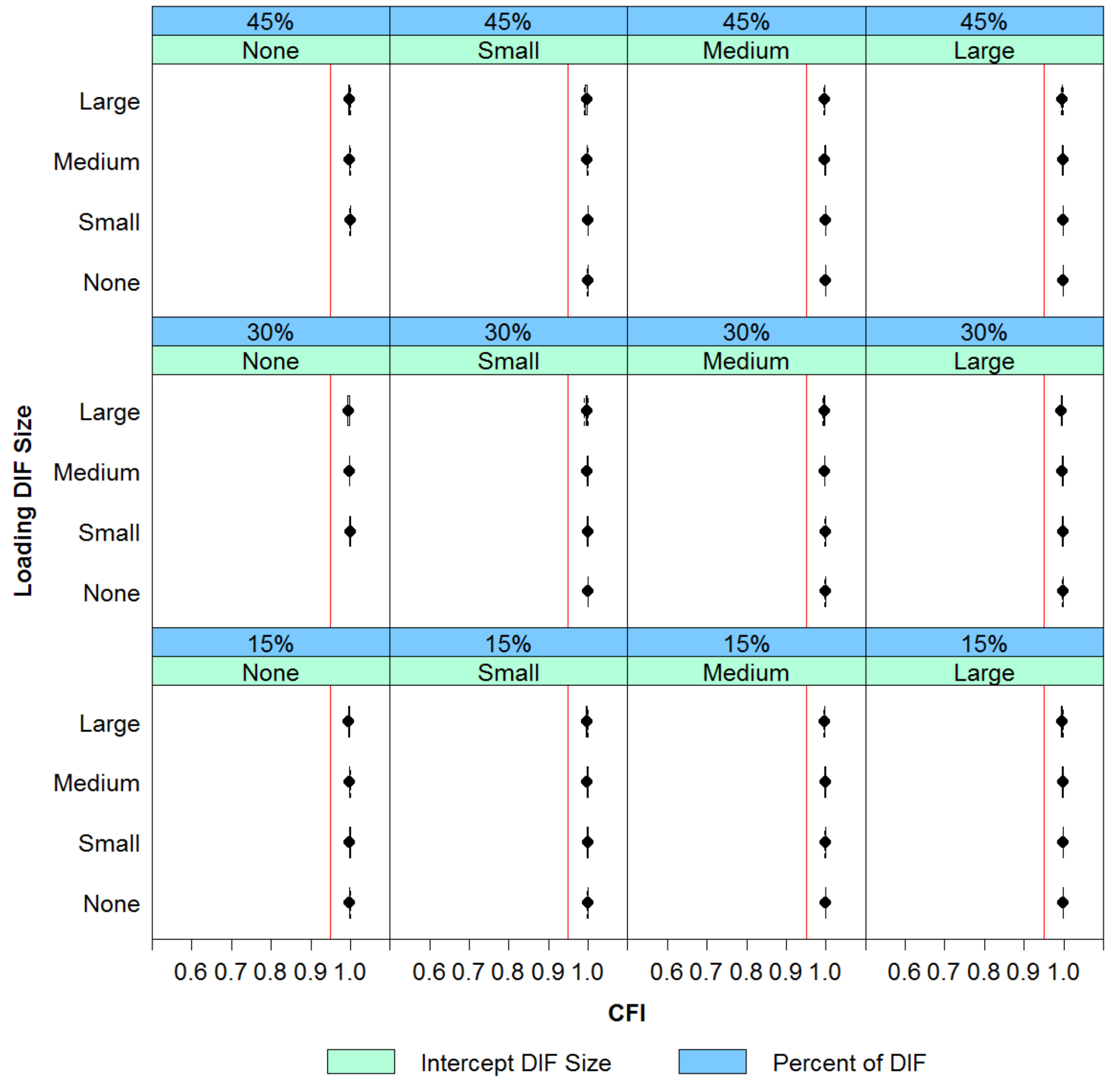

Figure 23. CFI Values for the Free Second-Order Non-Linear Latent Growth Model.

Nested Model Comparison. Three nested model comparisons were conducted to determine whether the addition of parameters to a less constrained (i.e., less parsimonious) model significantly improved model fit. The first chi-squared difference test compared the nested first-order models. The second and third chi-square difference tests compared nested models to the "true" model (i.e., the second-order free latent growth model). The results of the chi-squared difference tests are presented in Figures 24-26. The mean and standard deviation values for the chi-squared difference tests are included in Appendix D.

The first nested model comparison was between the first-order linear growth model and the first-order non-linear growth model. This chi-square difference test examined whether the addition of the non-linear slope, non-linear slope variance, and the corresponding covariances significantly improved model fit over the first-order linear model. The *p*-values for the chi-squared difference tests between the first-order linear and non-linear model are presented in Figure 24. Notably, because some of the non-linear models did not converge to a solution, there were fewer replications per condition than originally planned. The results suggested that the addition of the non-linear slope (and the associated variance and covariance parameters) often significantly improved model fit. The variability and range of the *p*-values decreased as the proportion of DIF items and the size of intercept DIF increased. Thus, as the proportion of non-invariant items and the size of intercept invariance violations increased, the chi-squared difference tests more consistently suggested that the addition of the non-linear parameters significantly improved model fit. Marginalizing across the loading DIF conditions, the chi-square difference test was significant 86% of the time for the condition with 45% of items with large intercept DIF, 91% of the time for the condition with 45% of items with medium intercept DIF, and 90% of the time for the condition with 30% of items with large intercept DIF.

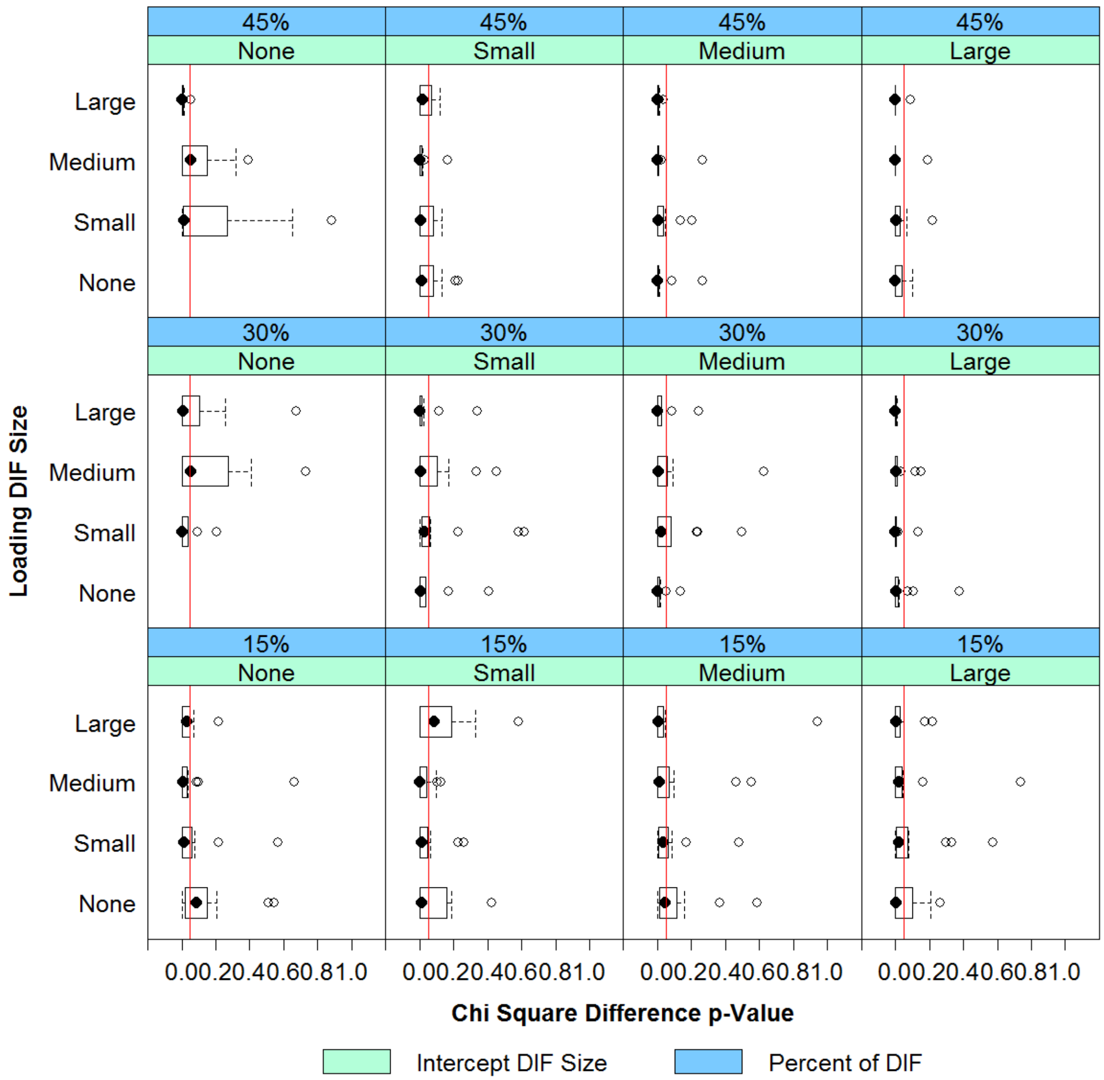

Figure 24. Chi-Squared Difference Test *p*-Value Comparing the First-Order Linear and Non-Linear Models

The second chi-squared difference test compared the second-order free latent growth model and the second-order free non-linear latent growth model. This comparison, much like the first-order comparisons, examined whether adding the non-linear item parameters (i.e., non-linear slope, non-linear variance, and the associated covariances) significantly improved model fit over the linear model. As with the first-order model comparison, many of the non-linear models did not converge and results should be interpreted with caution as there are fewer replications than intended. The *p*-values for the chi-squared difference test between the second-order free linear model and the second-order free non-linear model are presented in Figure 25. The median *p*-value was often below the 0.05 cutoff suggesting that many of the comparisons identified the non-linear model as a significant improvement on model fit. Notably, the *p*-values for the chi-squared comparisons had the least variability in the conditions with 45% of items with DIF, large intercept DIF, and large or medium loading DIF. In the condition with 45% of items with large intercept and large loading DIF, the chi-squared difference test found significant model fit improvement when the non-linear parameters were added 80% of the time.

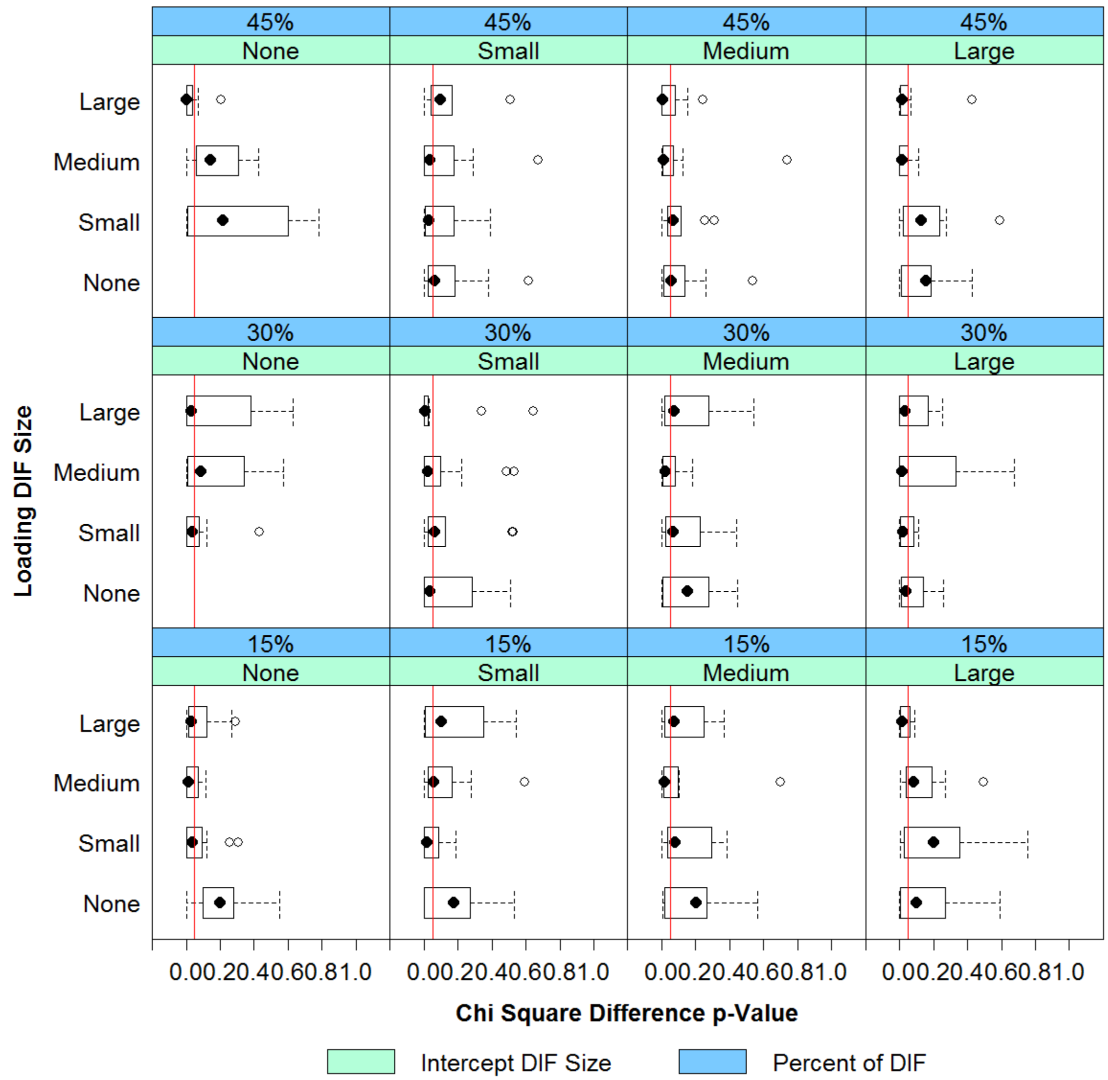

Figure 25. Chi-Squared Difference Test *p*-Value Comparing the Second-Order Linear Free and Non-Linear Free Models

The third chi-squared difference test compared the second-order free linear model with the second-order constrained linear model. This comparison examined whether freely estimating the non-invariant item parameters significantly improved model fit. The results suggested that for almost all replications in all conditions, modeling the invariance violations significantly improved model fit. The *p*-values were slightly more variable in conditions with a small proportion of items with DIF, but overall suggested that the model fit significantly better when the invariance violations were explicitly modeled.

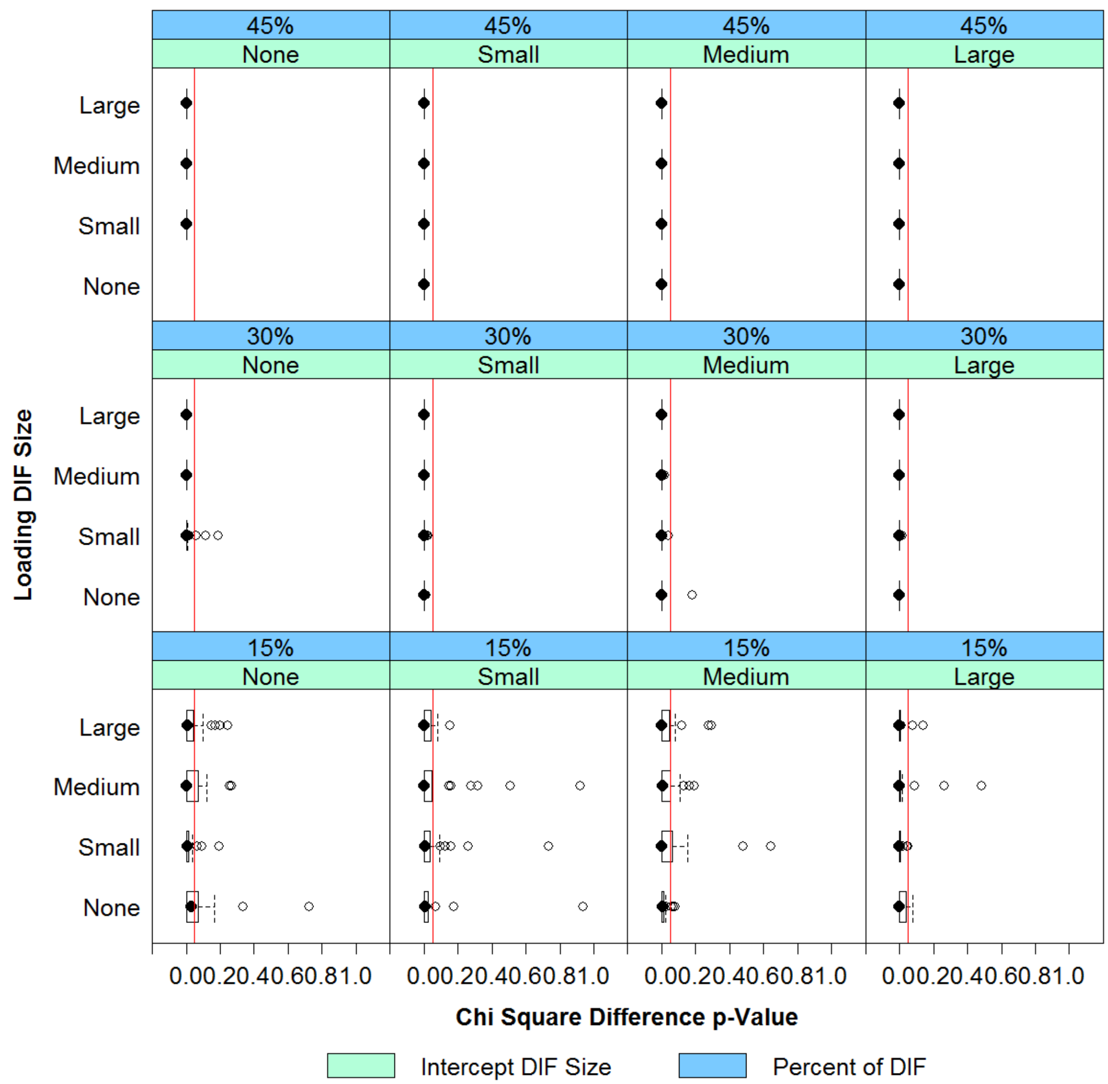

Figure 26. Chi-Squared Difference Test *p*-Value Comparing the Second-Order Linear Constrained and Linear Free Models

Parameter Recovery

Bias and RMSE were calculated to examine parameter recovery for the intercept variance, slope, and slope variance growth parameters. Bias and RMSE results are presented graphically for each of the six modeling approaches.

Intercept Variance. The relative bias estimates for the intercept variance parameter are presented in Figures 27-32. The two vertical red lines are plotted at -0.05 and 0.05 to outline the acceptable range. The mean and standard deviation values for the relative bias of the intercept variance are included in Appendix E. The relative bias in intercept variance was seemingly unrelated to the size of loading DIF, intercept DIF, or the proportion of DIF. Intercept variance parameter recovery, however, did seem to be related to the modeling approach used to model the growth. All of the linear models (LGM, 2LGMC, and 2LGMF) performed moderately well in recovering the intercept variance parameter value. The relative bias for the first-order linear model was within the acceptable range approximately 36% of the time (aggregated across all conditions). The estimates for both second-order linear models (2LGMC and 2LGMF) tended to be slightly overestimated and were often just out of the acceptable range for relative bias (across all conditions, the values were outside the range approximately 80% and 77% of the time for the 2LGMC and 2LGMF, respectively). The three non-linear models (NLGM, 2NLGMC, and 2NLGMF) performed similarly to one another in terms of intercept variance bias; however the number of iterations was small given that many of the non-linear models did not converge. The range of intercept variance relative bias was much larger for the non-linear models than the linear models. Across all conditions, 87%, 82%, and 83% of the intercept variance relative bias values fell outside of the acceptable range for the NLGM, 2NLGMC, and 2NLGMF, respectively.

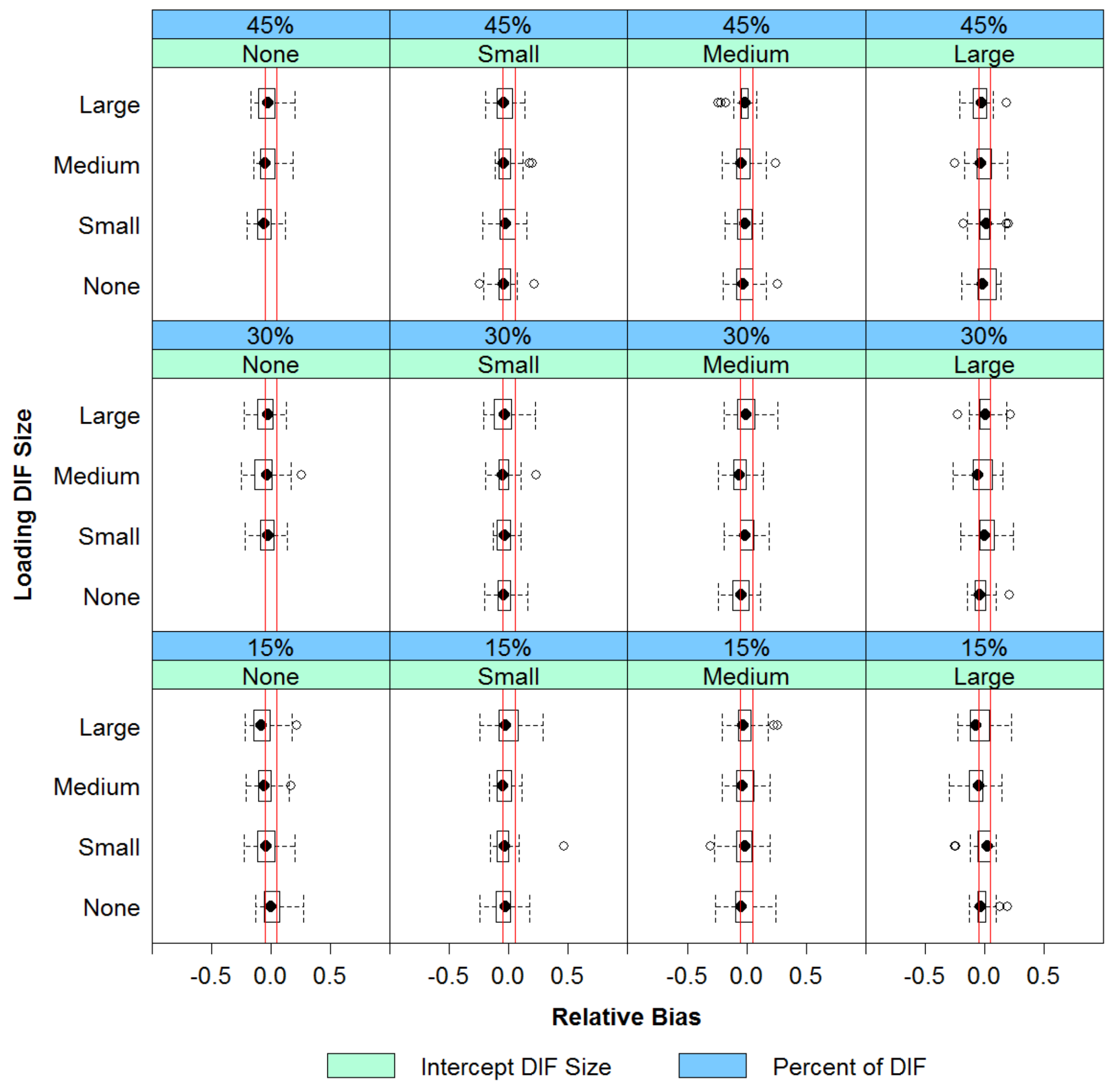

Figure 27. Intercept Variance Relative Bias for the First-Order Latent Growth Model.

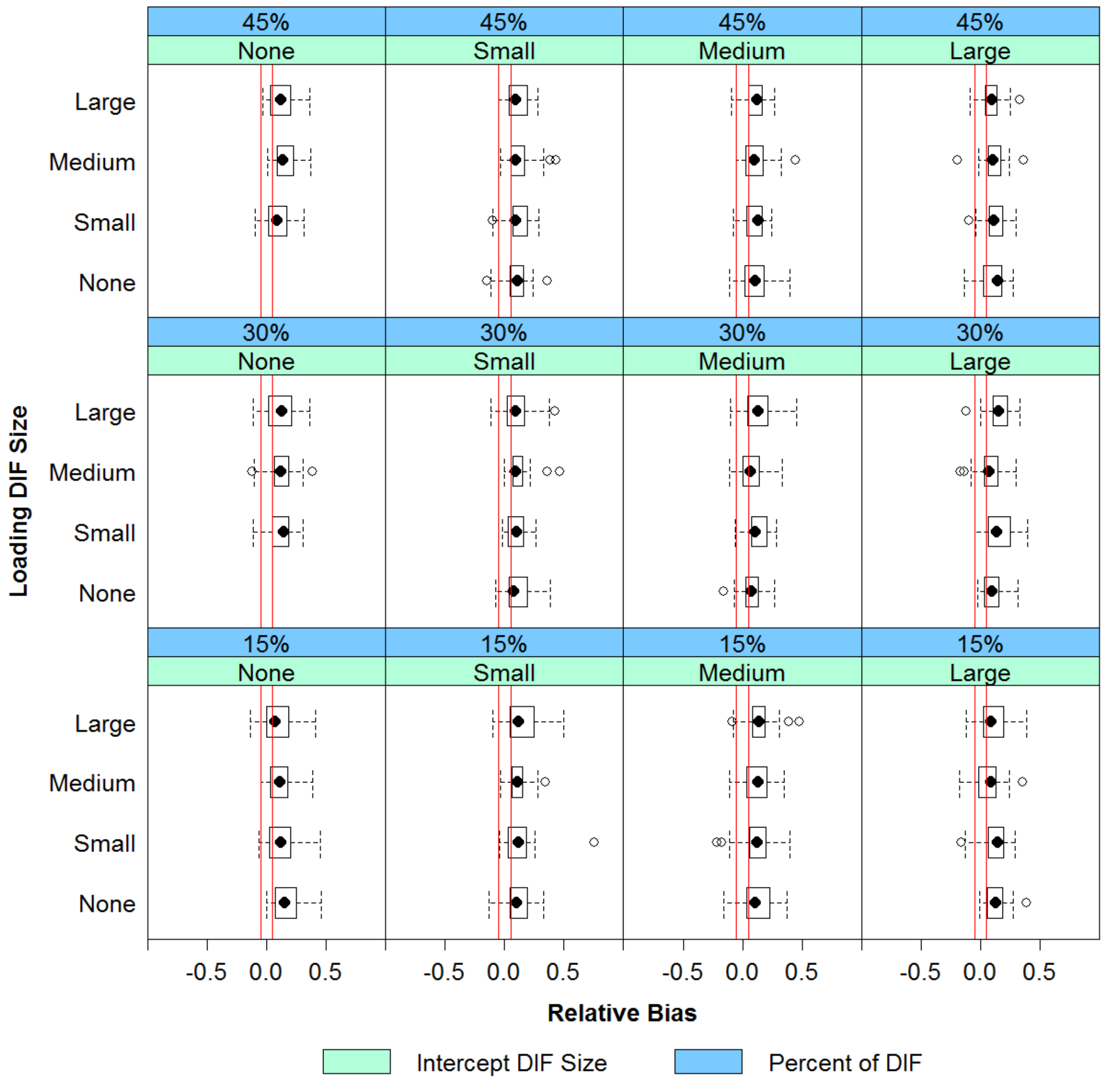

Figure 28. Intercept Variance Relative Bias for the Constrained Second-Order Latent Growth Model.

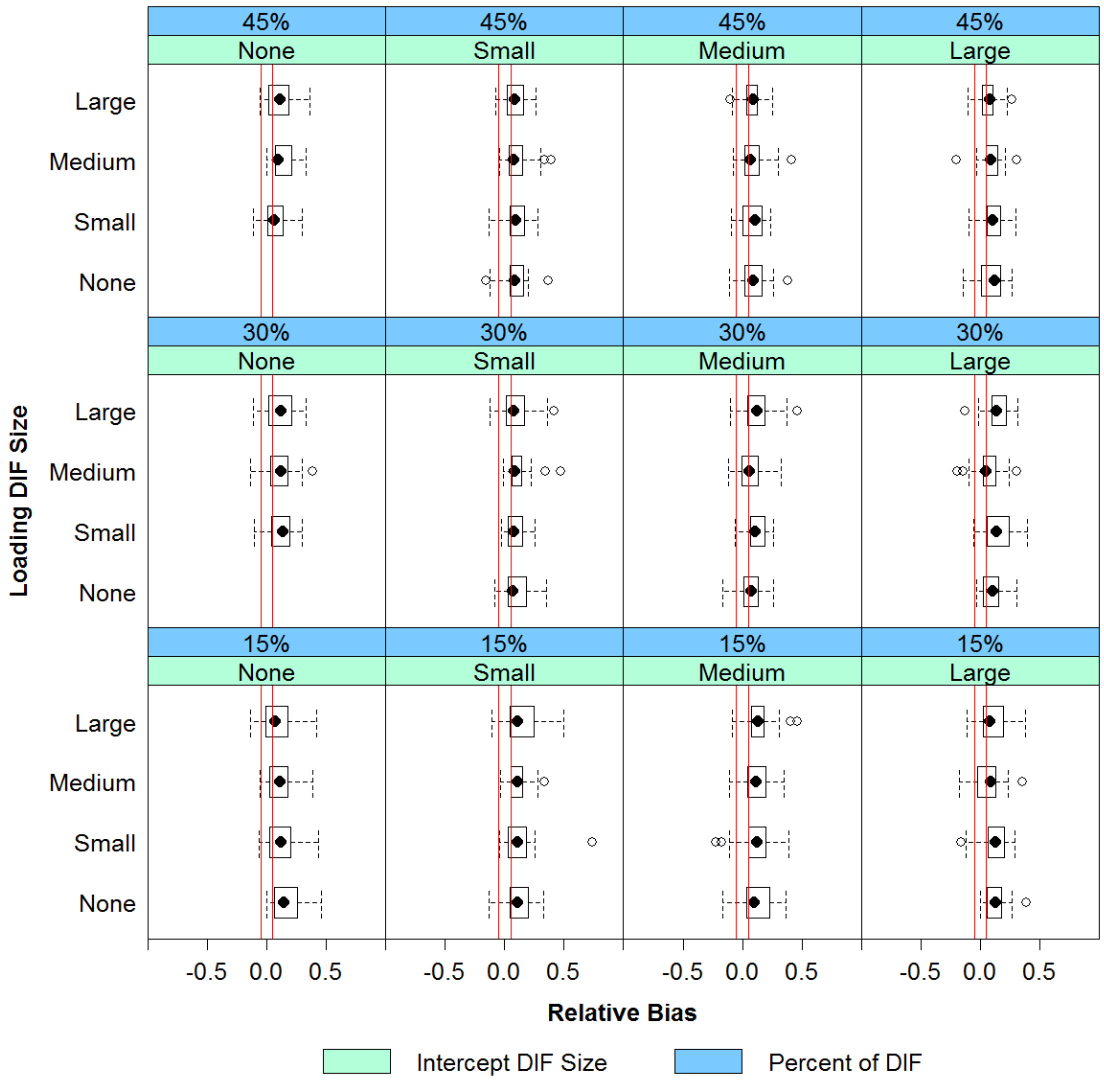

Figure 29. Intercept Variance Relative Bias for the Free Second-Order Latent Growth Model.

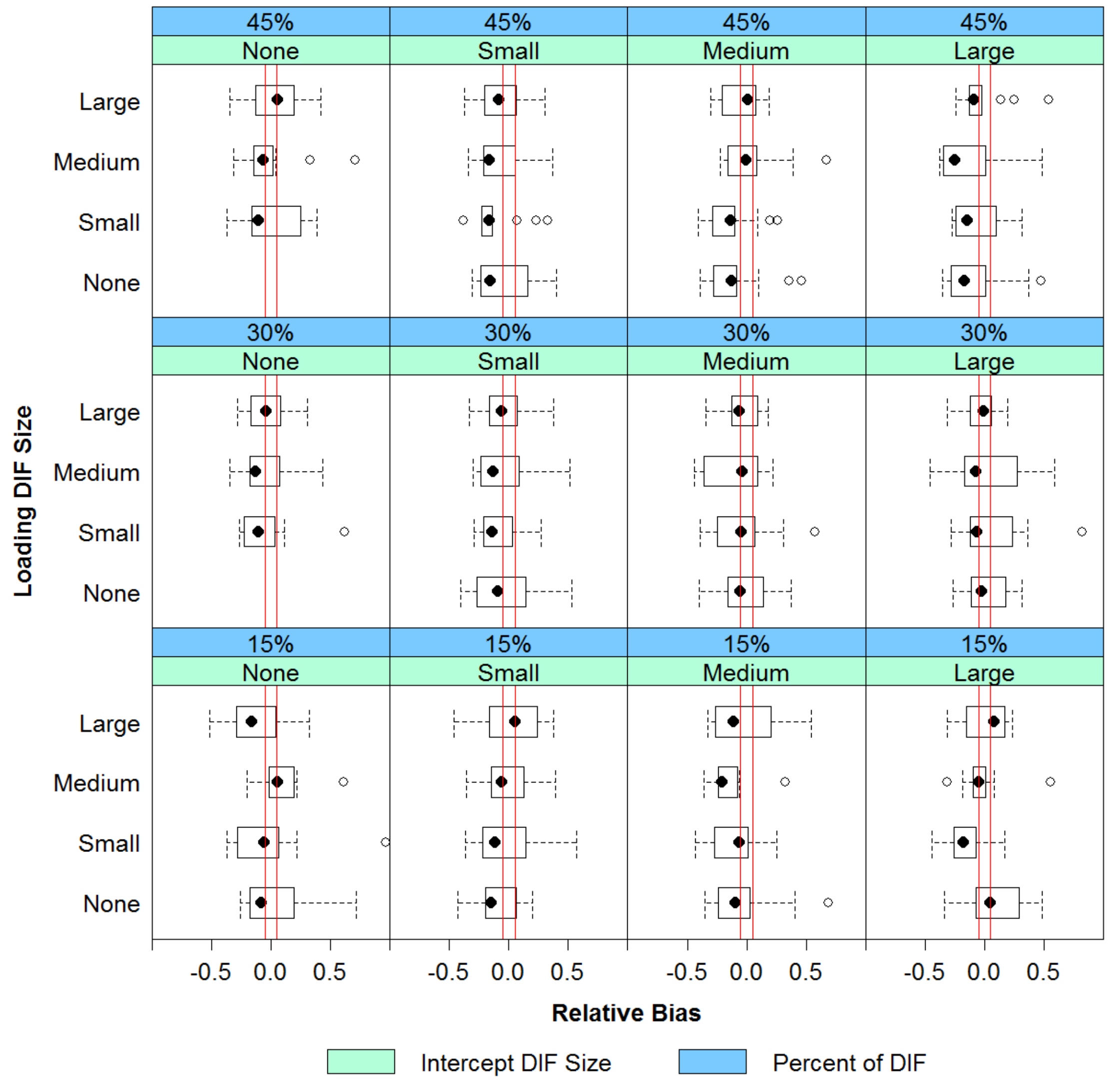

Figure 30. Intercept Variance Relative Bias for the First-Order Non-Linear Latent Growth Model.

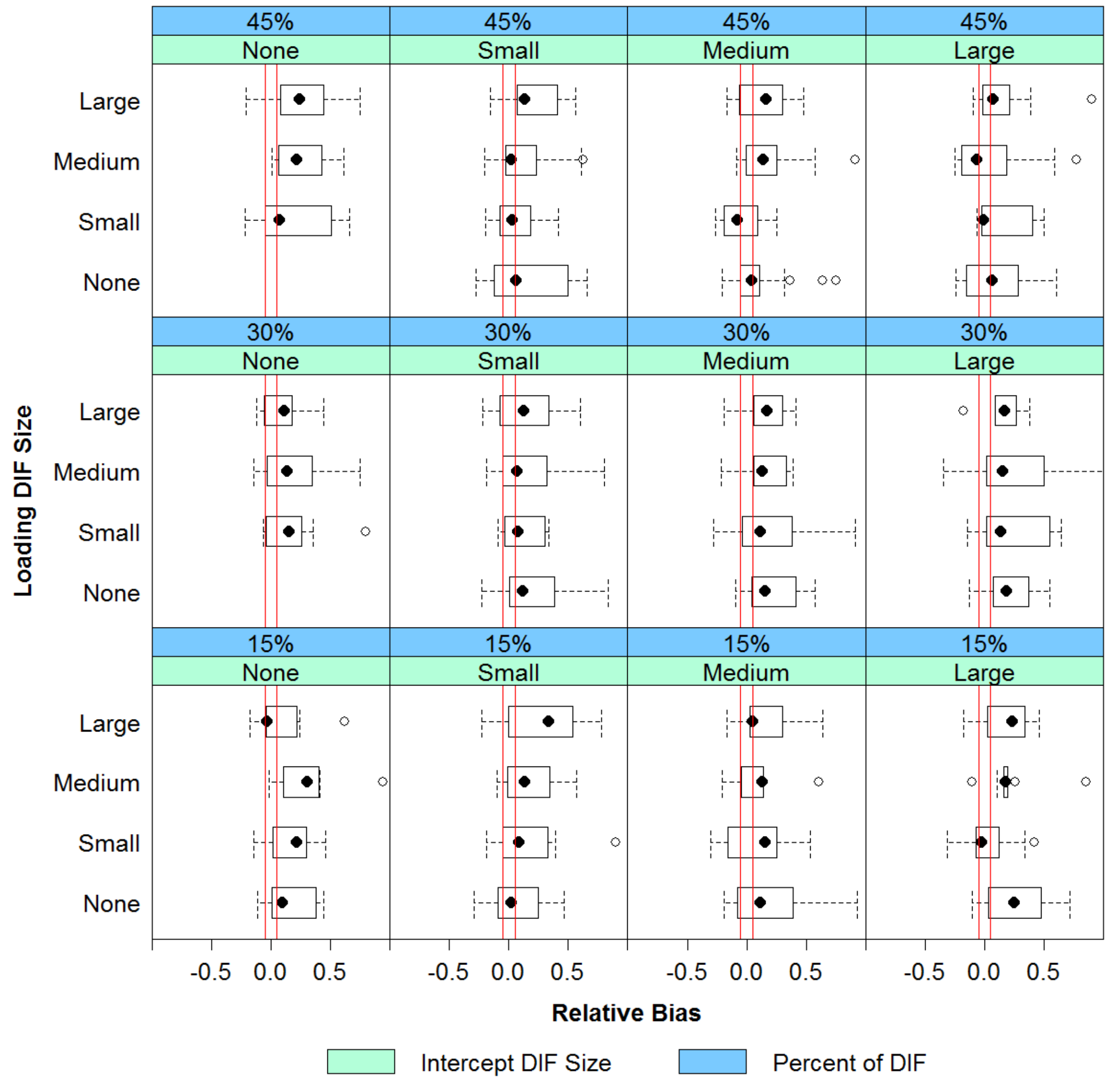

Figure 31. Intercept Variance Relative Bias for the Constrained Second-Order Non-Linear Latent Growth Model.

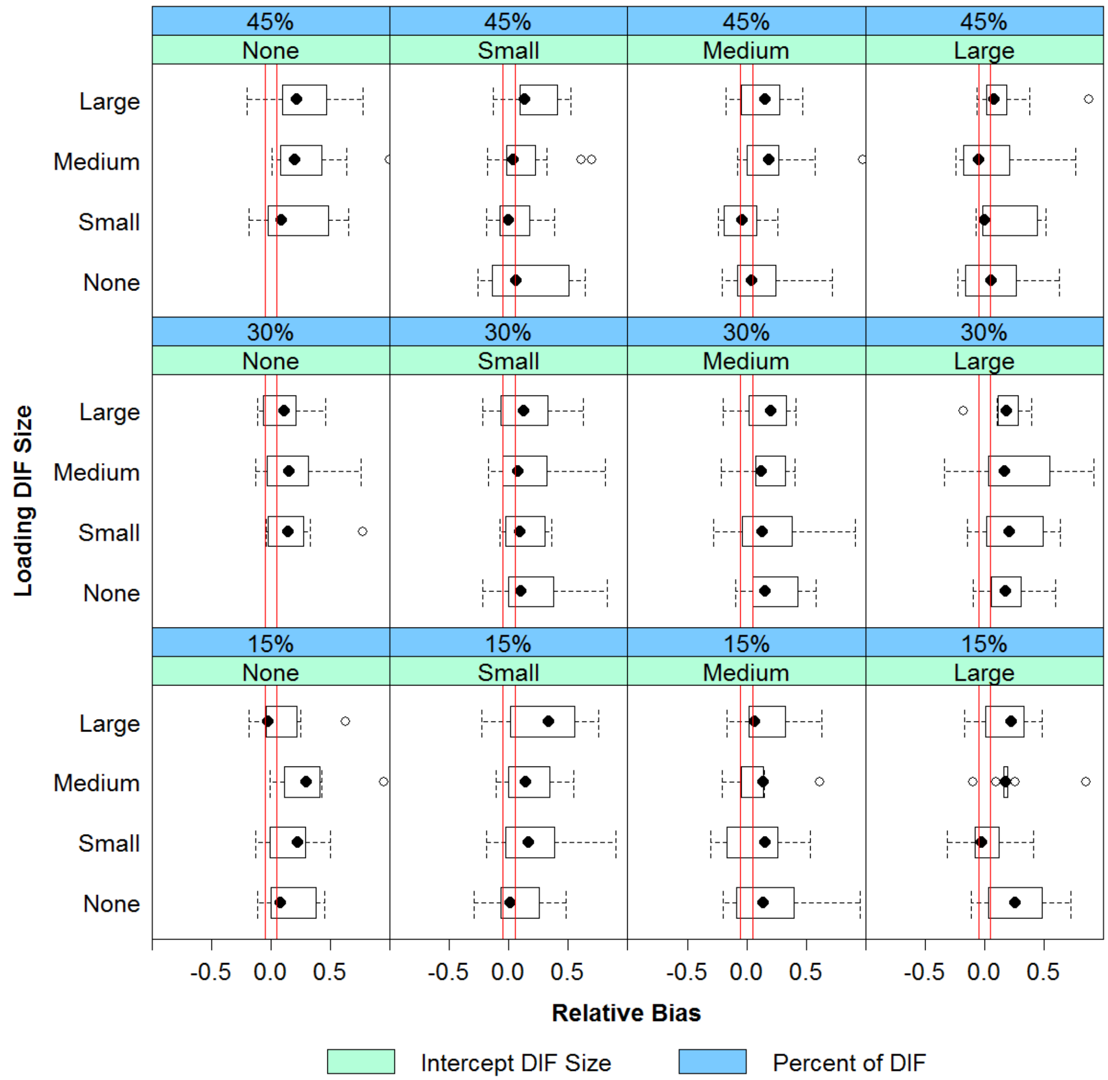

Figure 32. Intercept Variance Relative Bias for the Free Second-Order Non-Linear Latent Growth Model.

The RMSE estimates for the intercept variance parameter are presented in Figures 33-38. The mean and standard deviation values for the RMSE of the intercept variance are included in Appendix F. The RMSE values for the intercept variance parameter in the LGM, 2LGMC, and 2LGMF modeling approaches were relatively similar across all conditions varied in the study. As with the linear models, the intercept variance RMSE values for the non-linear models (NLGM, 2NLGMC, and 2NLGMF) did not depend on the conditions in the study. The values were more variable for the non-linear models.

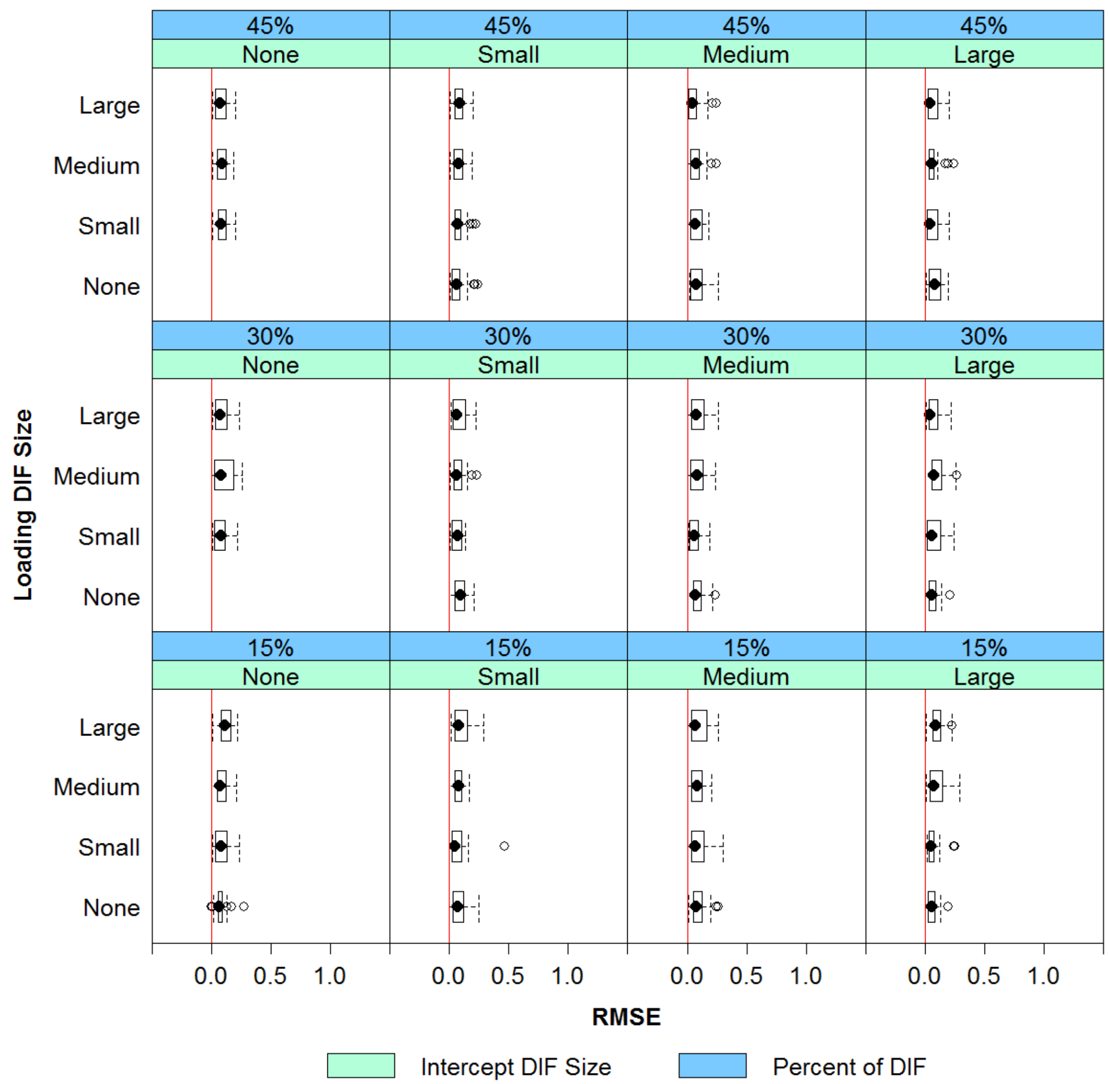

Figure 33. Intercept Variance Root Mean Square Error for the First-Order Latent Growth Model.

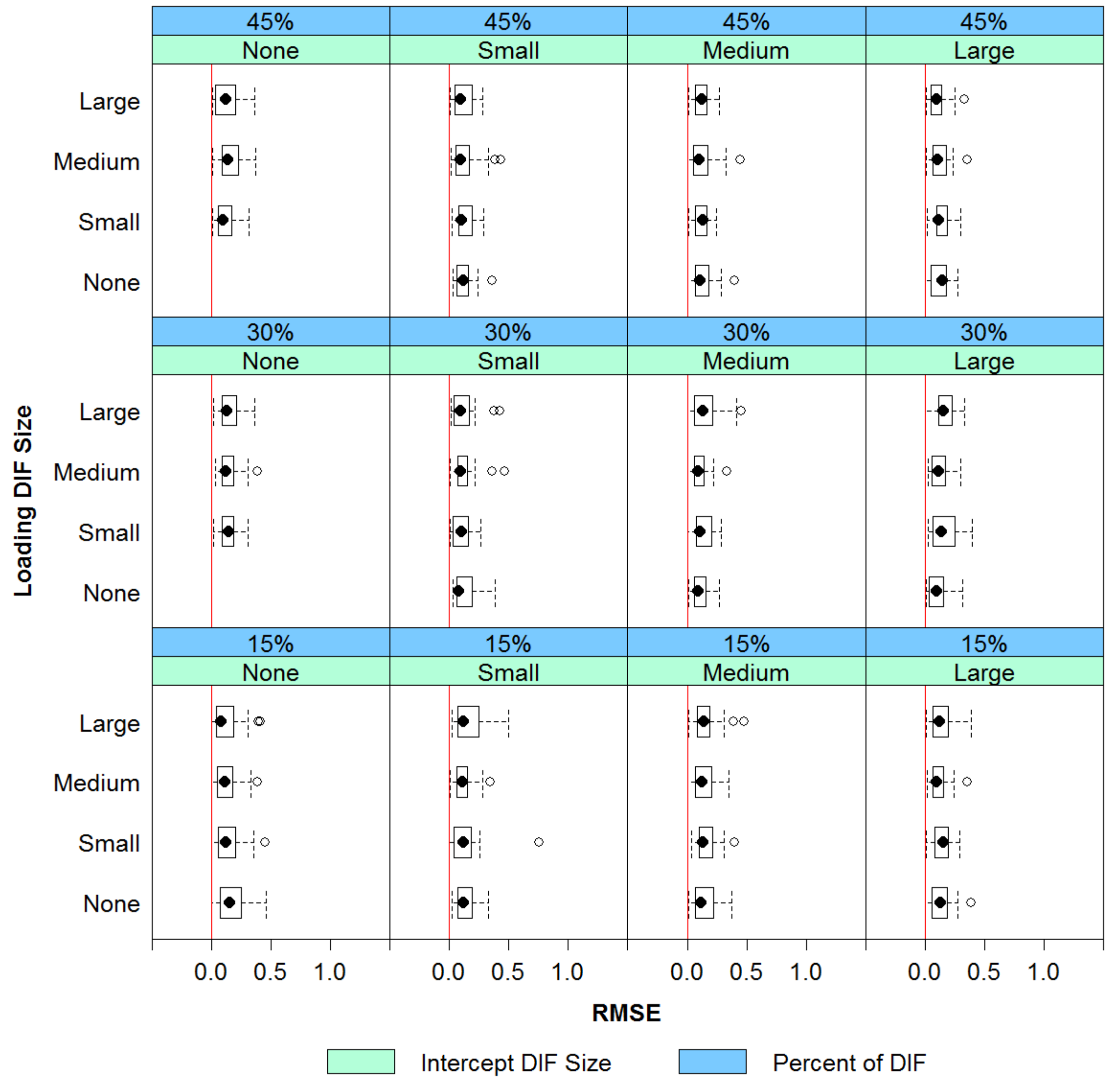

Figure 34. Intercept Variance Root Mean Square Error for the Constrained Second-Order Latent Growth Model.

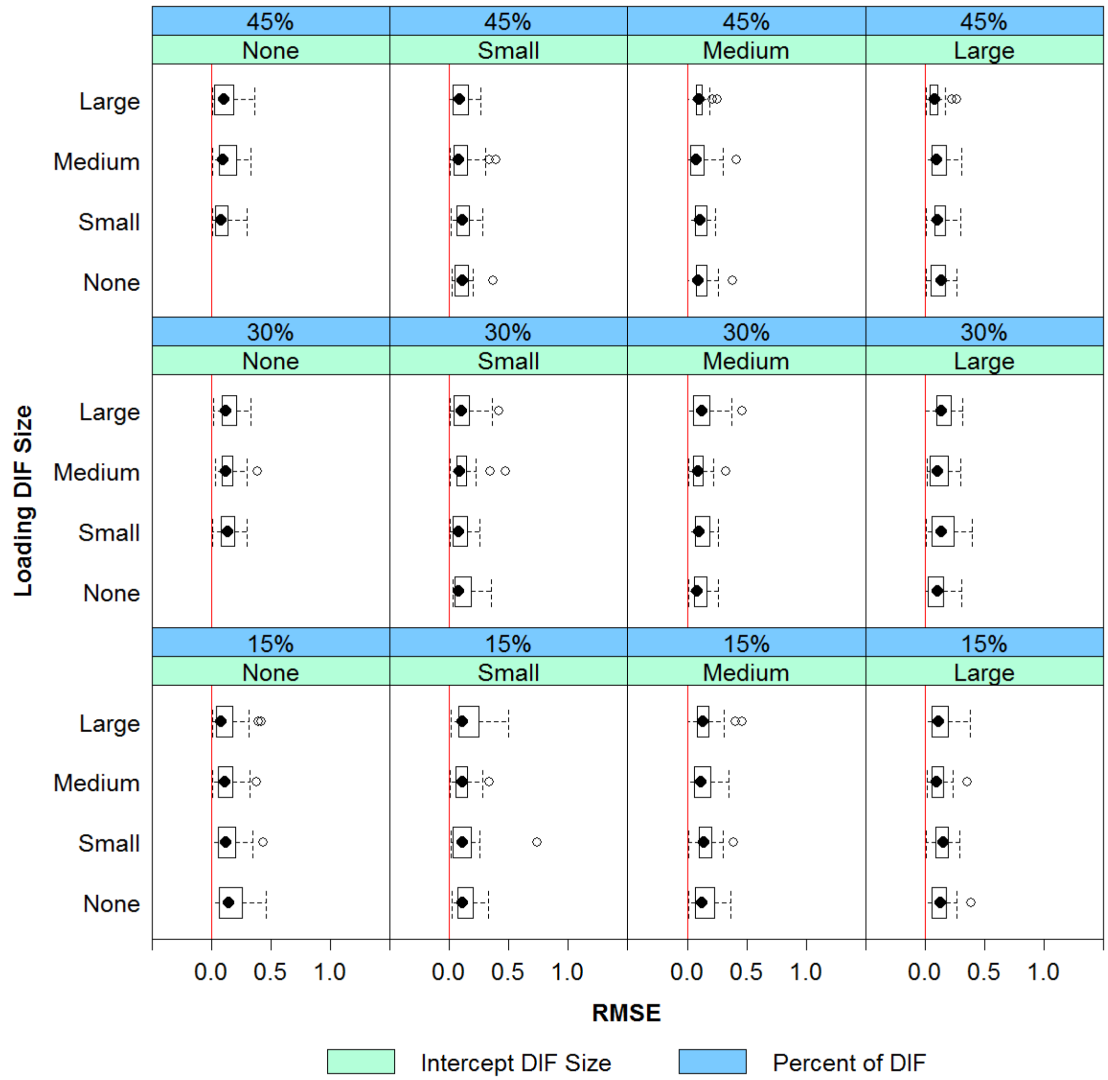

Figure 35. Intercept Variance Root Mean Square Error for the Free Second-Order Latent Growth Model.

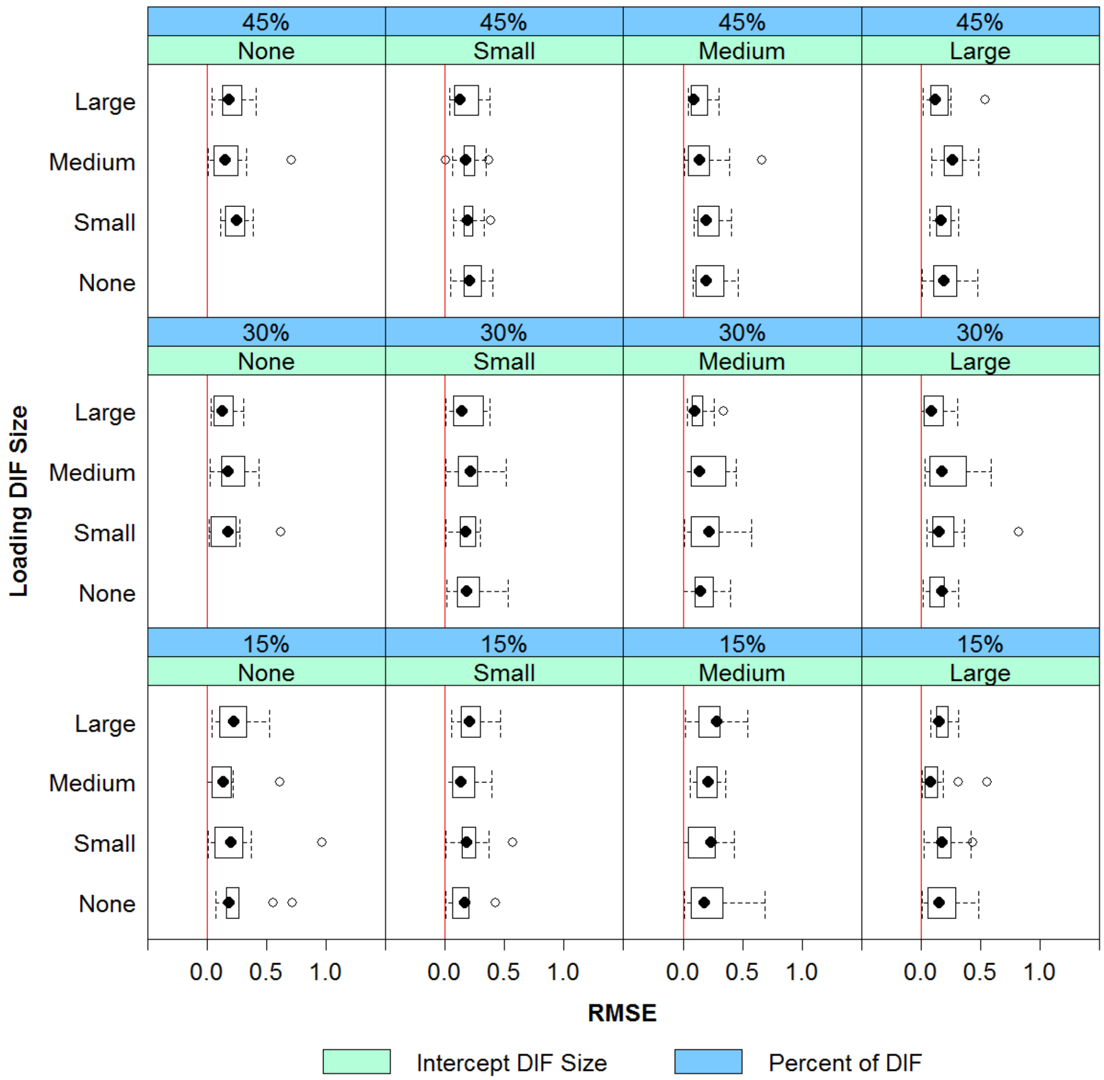

Figure 36. Intercept Variance Root Mean Square Error for the First-Order Non-Linear Latent Growth Model.

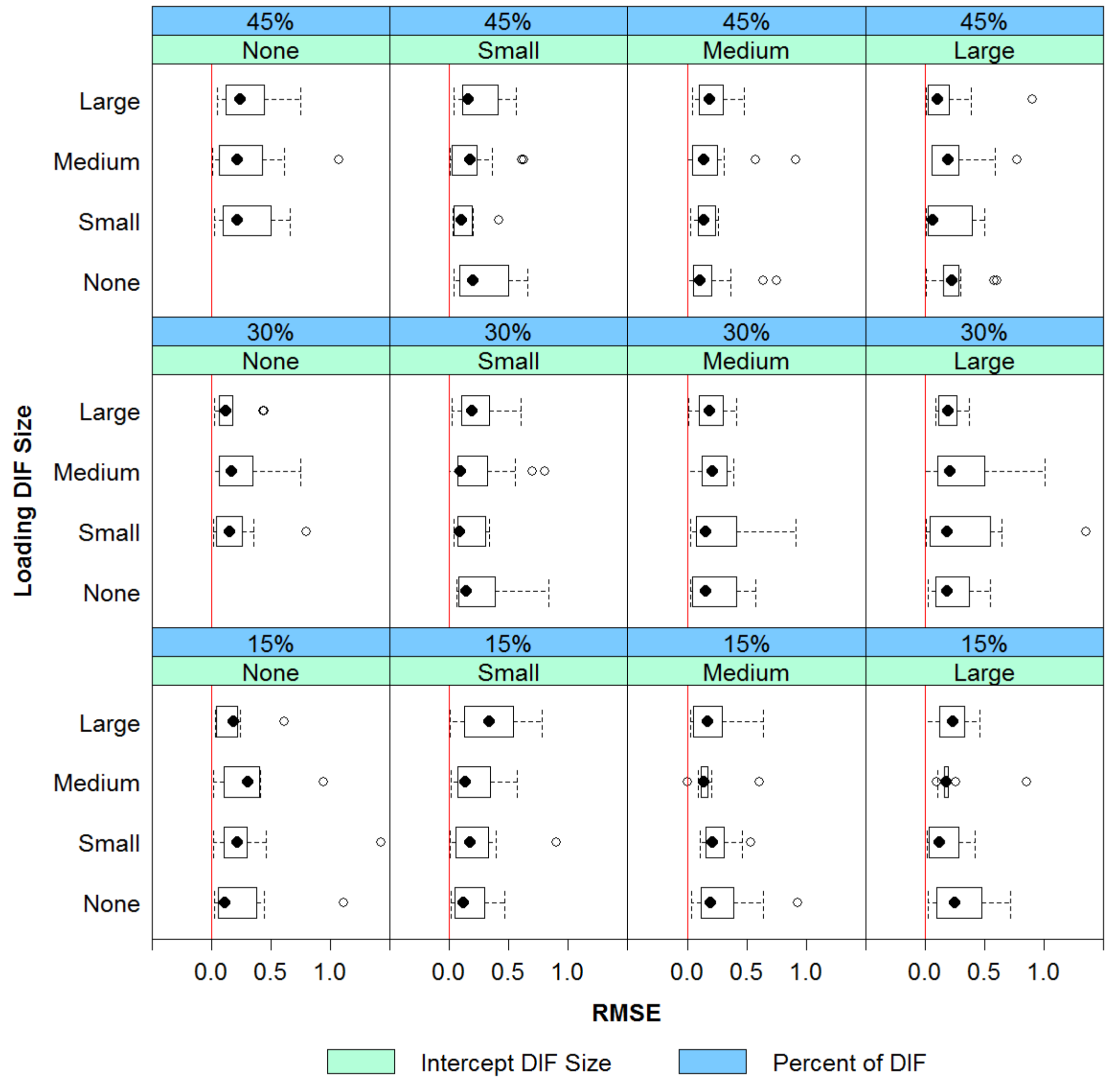

Figure 37. Intercept Variance Root Mean Square Error for the Constrained Second-Order Non-Linear Latent Growth Model.

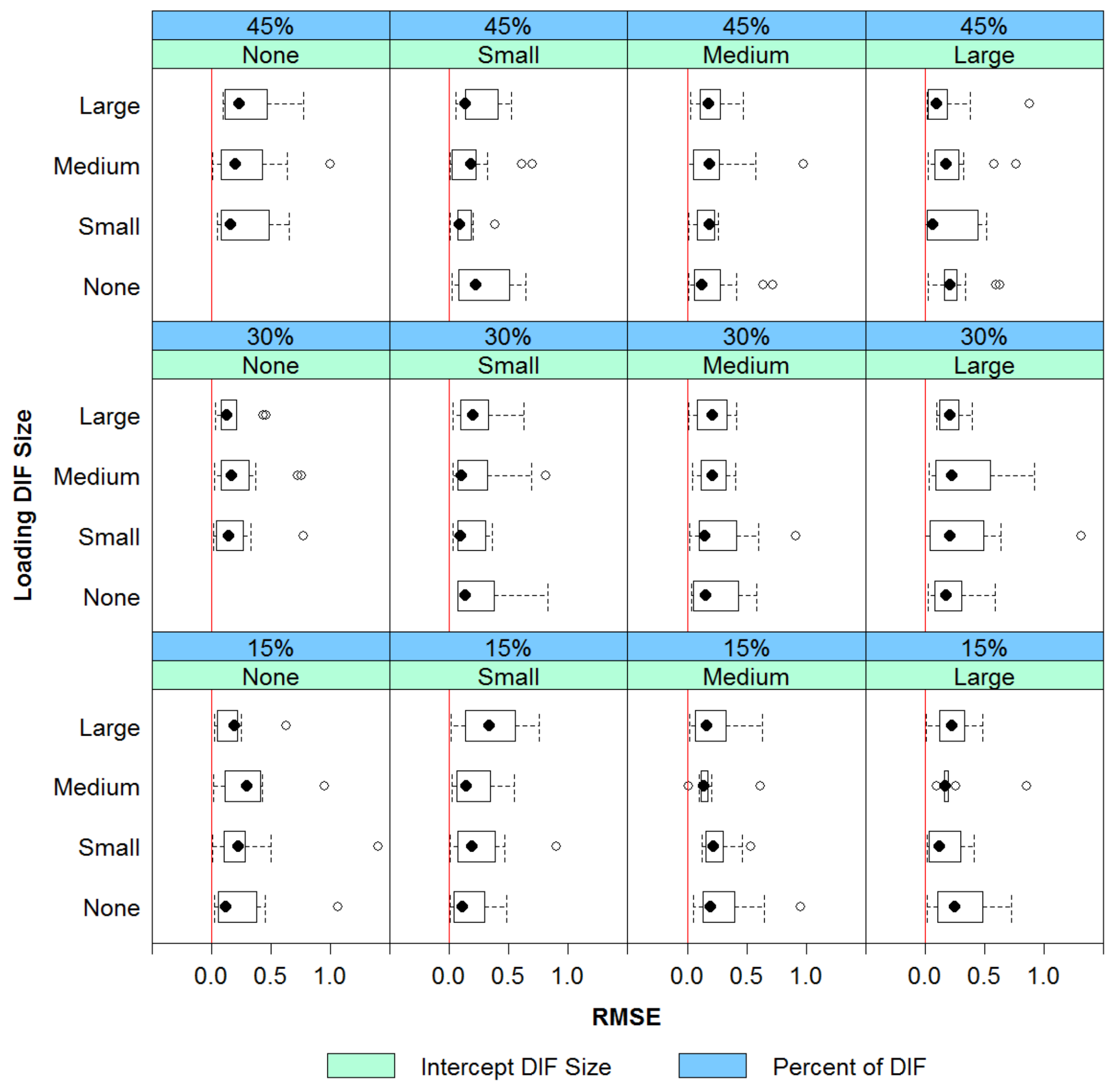

Figure 38. Intercept Variance Root Mean Square Error for the Free Second-Order Non-Linear Latent Growth Model.

Slope. The bias estimates for the slope parameter are presented in Figures 39-44. The two vertical red lines are plotted at -0.05 and 0.05 to outline the acceptable range. The mean and standard deviation values for the relative bias of the slope are included in Appendix G. The relative bias of the slope parameter was modestly related to intercept

DIF size and the percent of DIF. In the linear modeling (LGM, 2LGMC, and 2LGMF) approaches the relative bias was generally within the acceptable range when the percent of DIF was 15%. In the 30% DIF conditions, the relative bias was within acceptable conditions except for when there was large intercept DIF. In the conditions with 45% DIF, the relative bias exceeded the acceptable range when the intercept DIF was medium or large. These findings were more pronounced for the second-order modeling approaches, particularly the 2LGMF modeling approach for which 100% of the slope relative bias estimates fell outside of the acceptable range. The non-linear modeling approaches (NLGM, 2NLGMC, and 2NLGMF) should be interpreted with caution as there are fewer replications per condition. Of the replications that converged, the pattern of relative slope bias in the non-linear models was similar to the patterns in the 2LGMF modeling approach. The conditions where 30-45% of items had large intercept DIF had notably larger relative bias (100% fell outside of the acceptable range) than other conditions for all non-linear modeling approaches. The relative bias in the problematic conditions (i.e., conditions in which 30-45% of items had large intercept DIF) for the non-linear models was notably larger than the relative bias in the corresponding linear models.

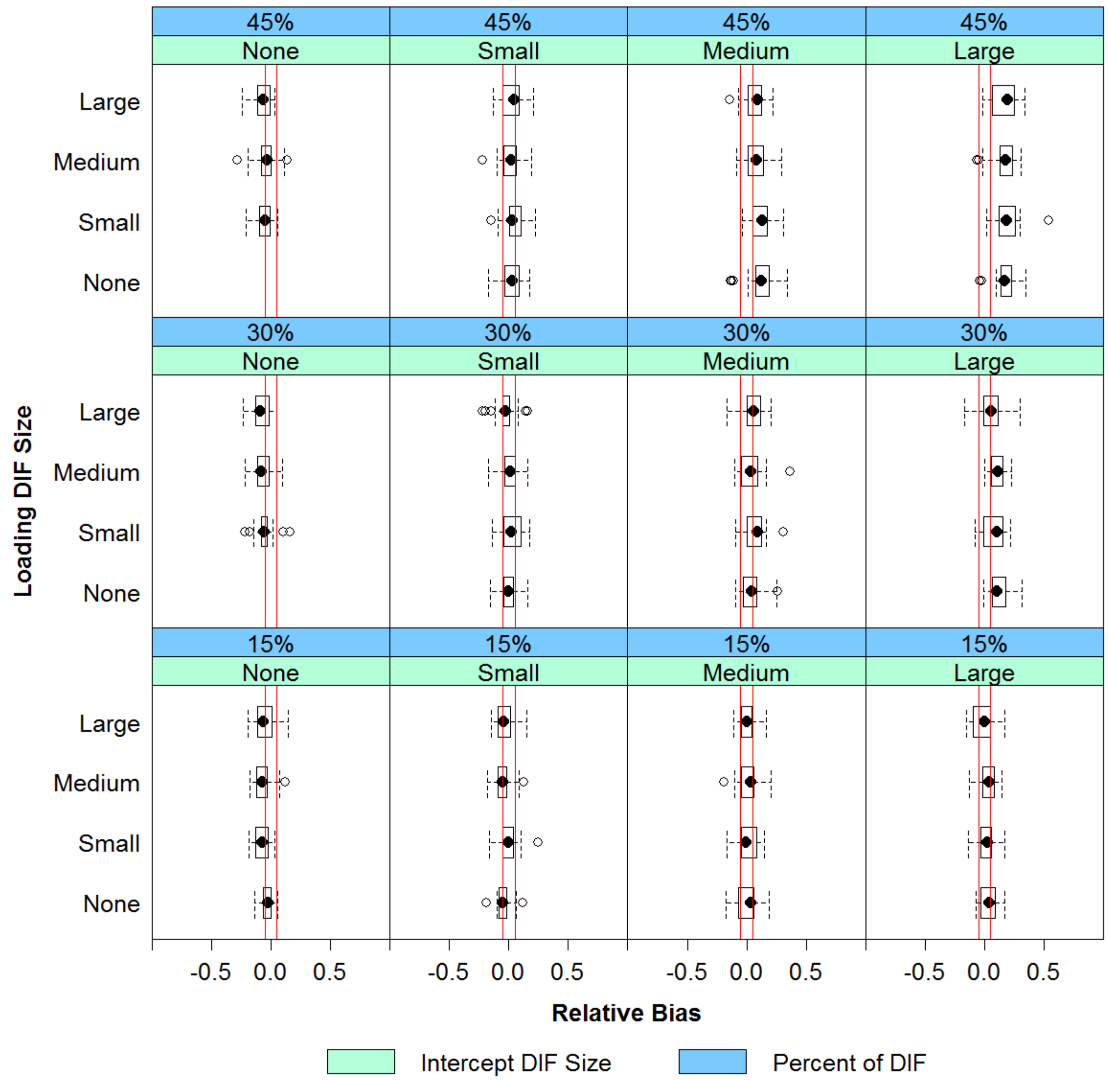

Figure 39. Slope Relative Bias for the First-Order Latent Growth Model.

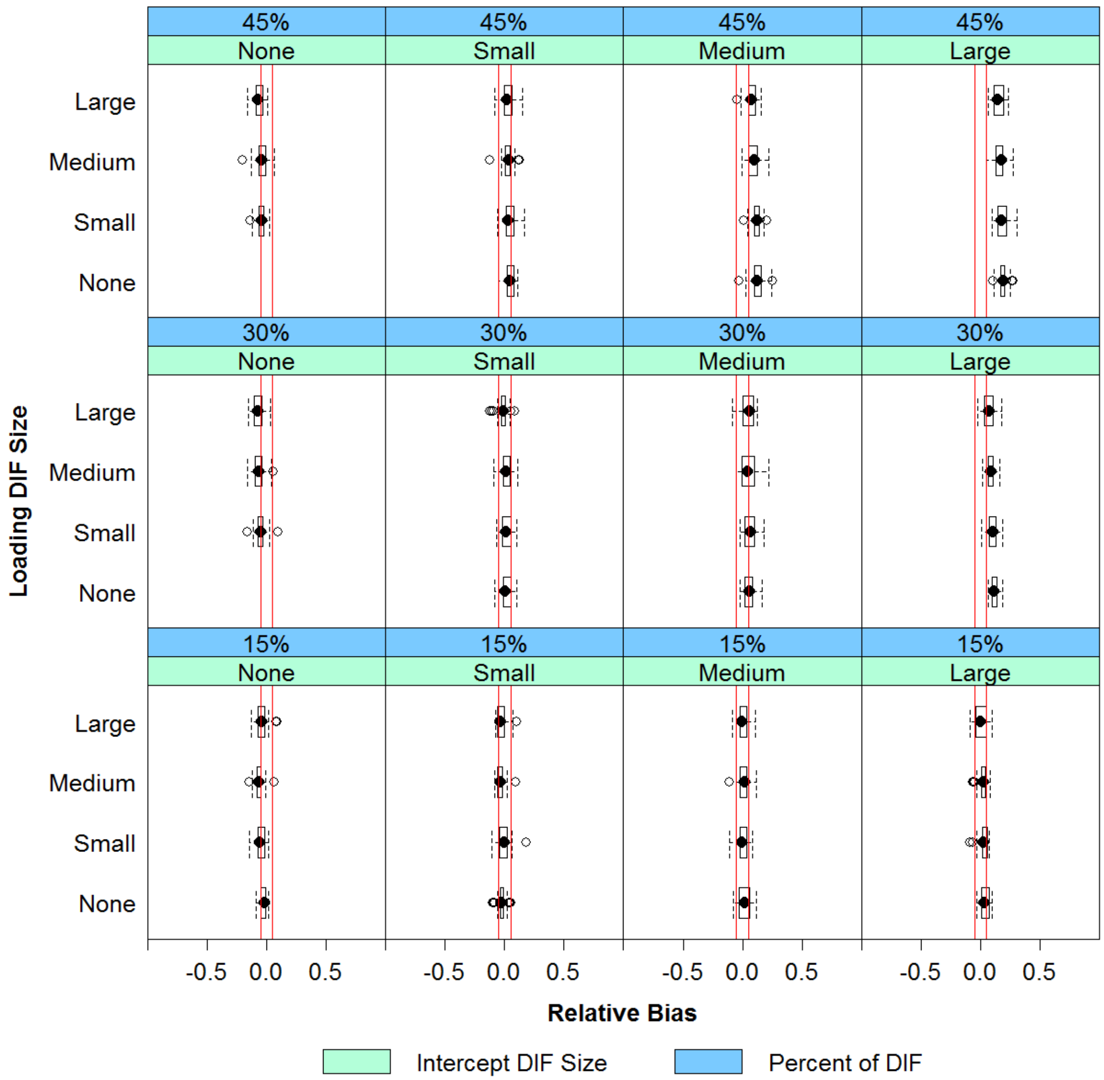

Figure 40. Slope Relative Bias for the Constrained Second-Order Latent Growth Model.

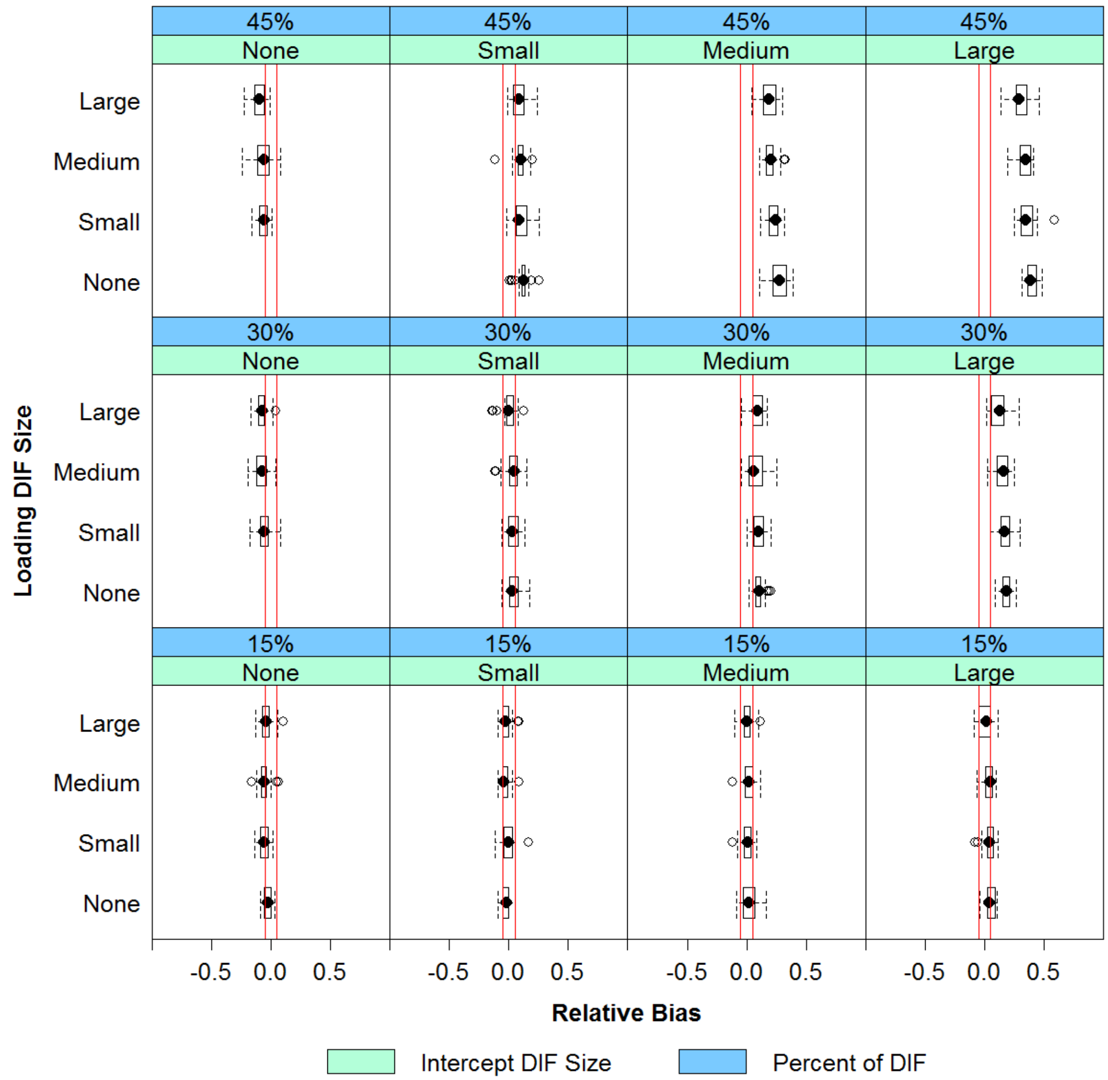

Figure 41. Slope Relative Bias for the Free Second-Order Latent Growth Model.

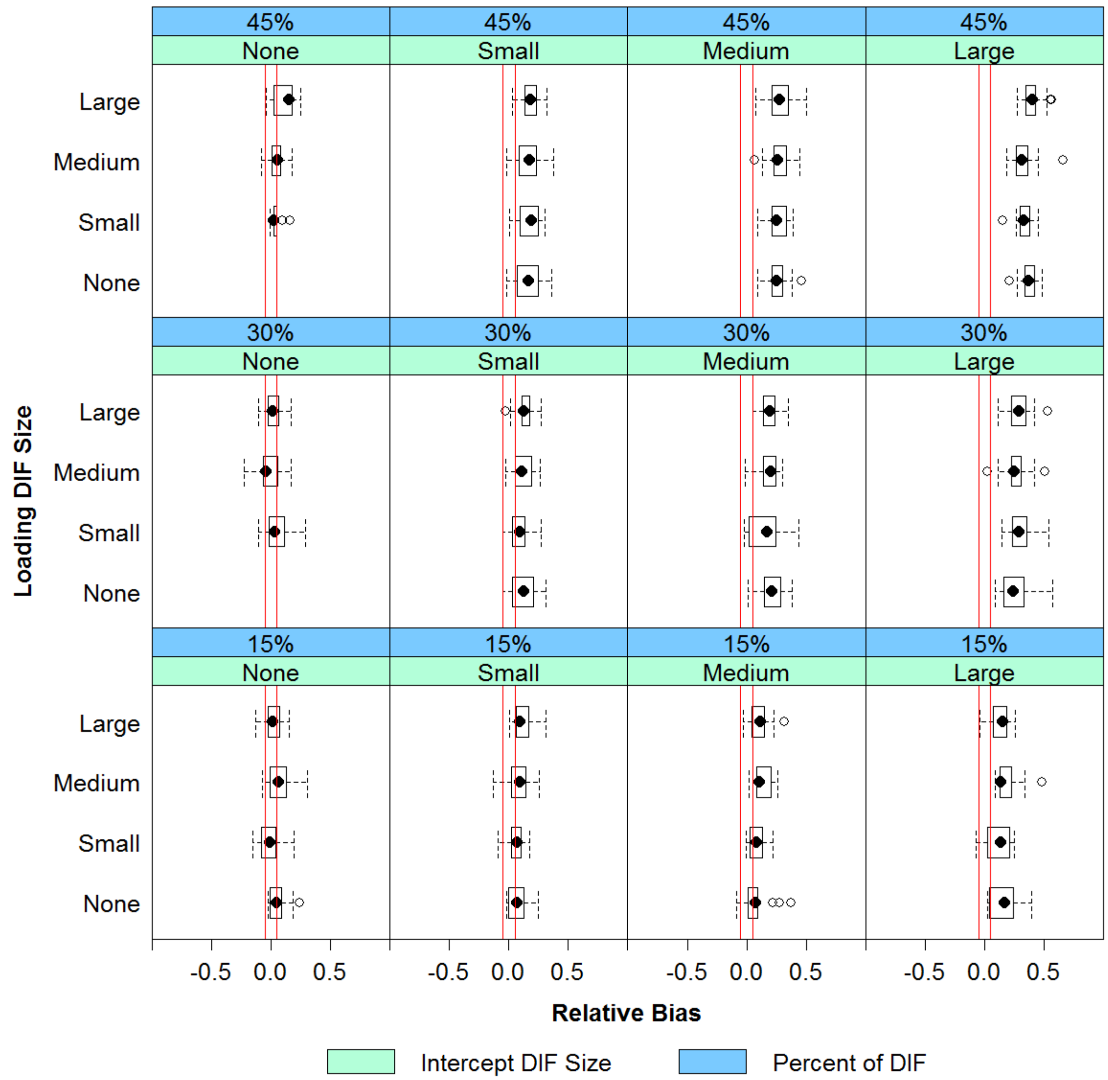

Figure 42. Slope Relative Bias for the First-Order Non-Linear Latent Growth Model.

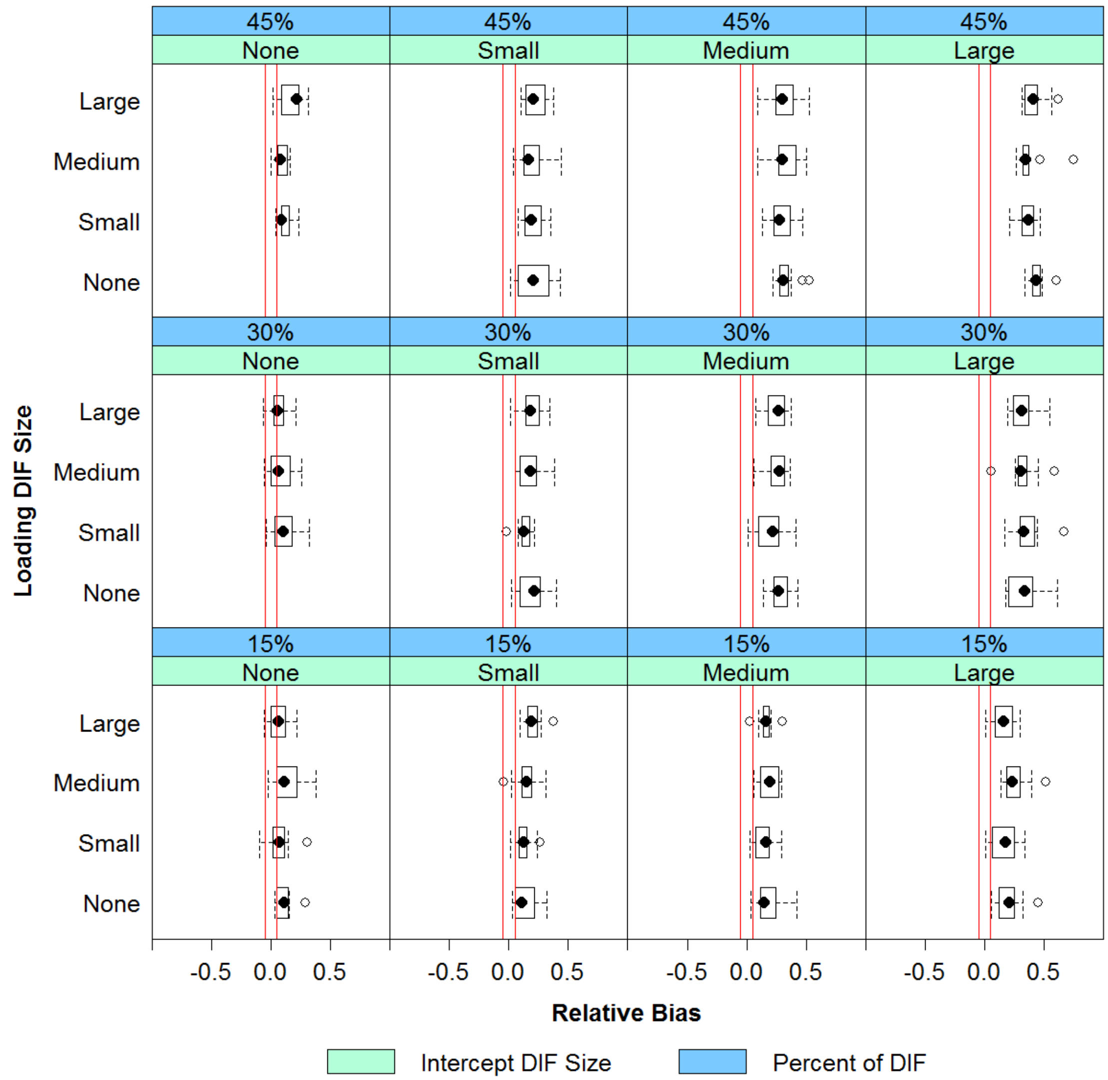

Figure 43. Slope Relative Bias for the Constrained Second-Order Non-Linear Latent Growth Model.

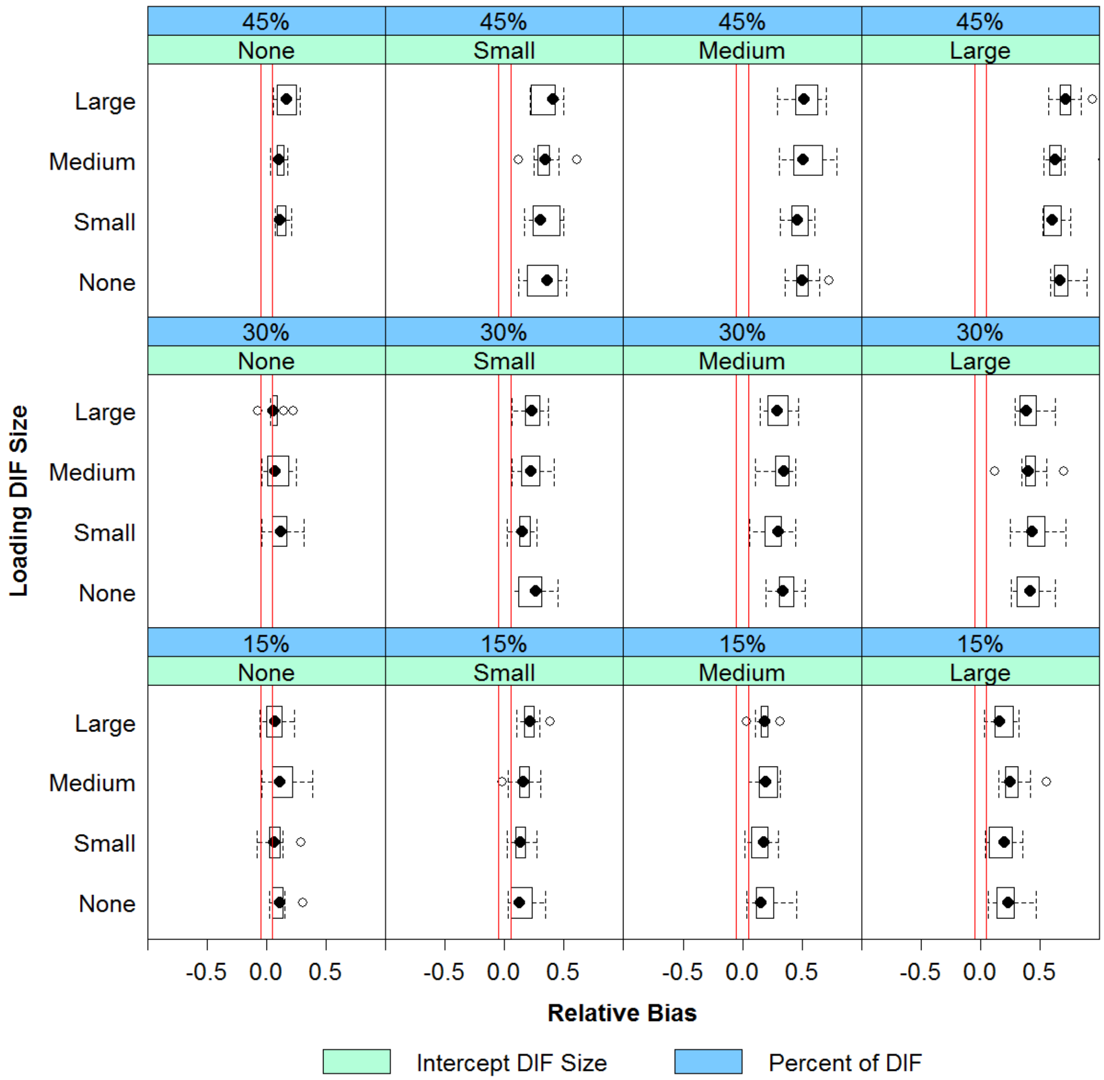

Figure 44. Slope Relative Bias for the Free Second-Order Non-Linear Latent Growth Model.

The RMSE estimates for the intercept variance parameter are presented in Figures 45-50. The mean and standard deviation values for the RMSE of the slope are included in Appendix H. Most of the slope RMSE values for the linear modeling approaches (LGM, 2LGMC, and 2LGMF) were essentially zero. In the conditions with 45% large intercept

DIF, the values were slightly larger compared to the other conditions, but were small in an absolute sense. The RMSE values for the non-linear models (NLGM, 2NLGMC, and 2NLGMF) were larger than their linear modeling counterparts and tended to increase as the proportion of DIF and the size of intercept DIF increased.

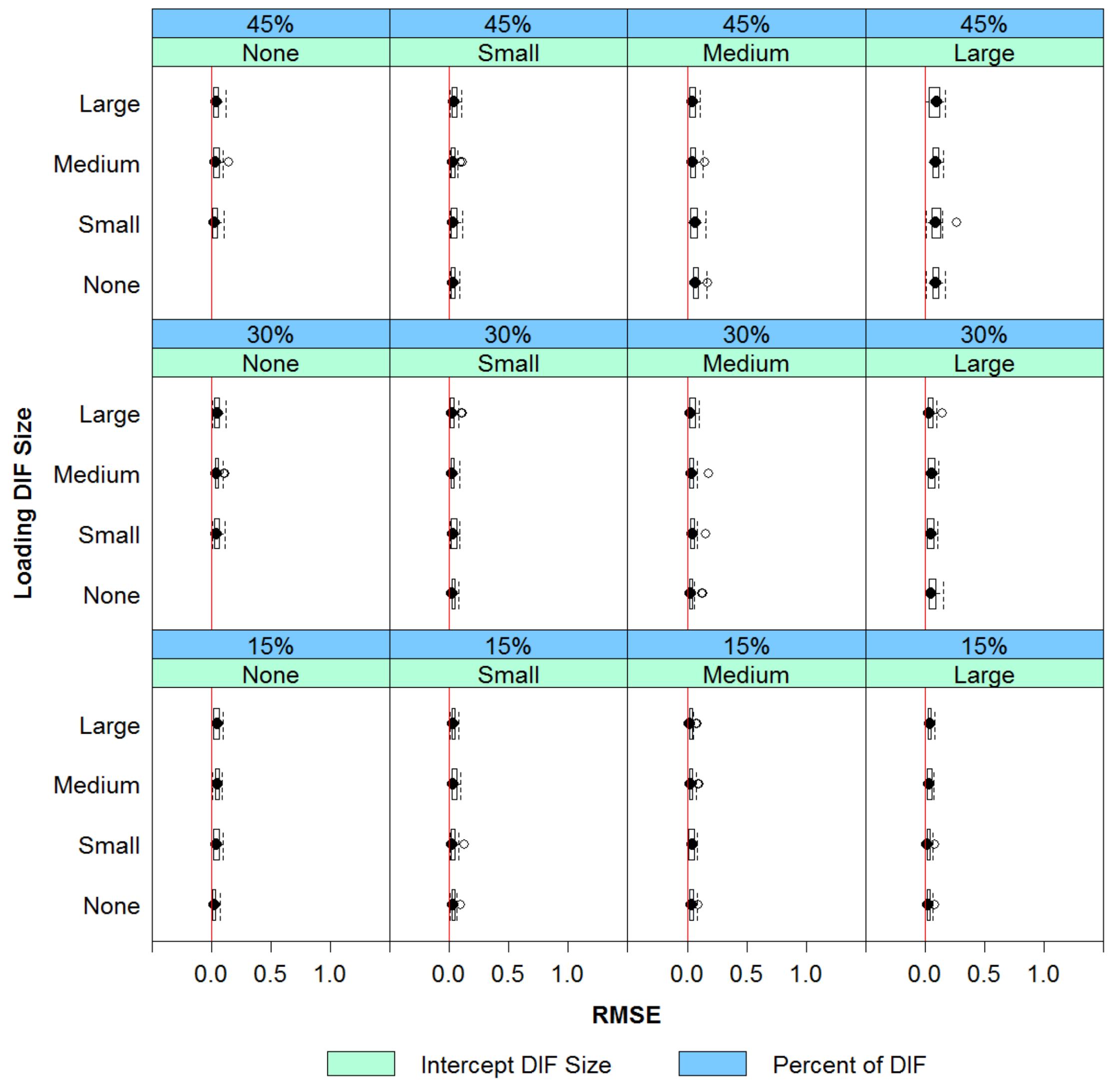

Figure 45. Slope Root Mean Square Error for the First-Order Latent Growth Model.

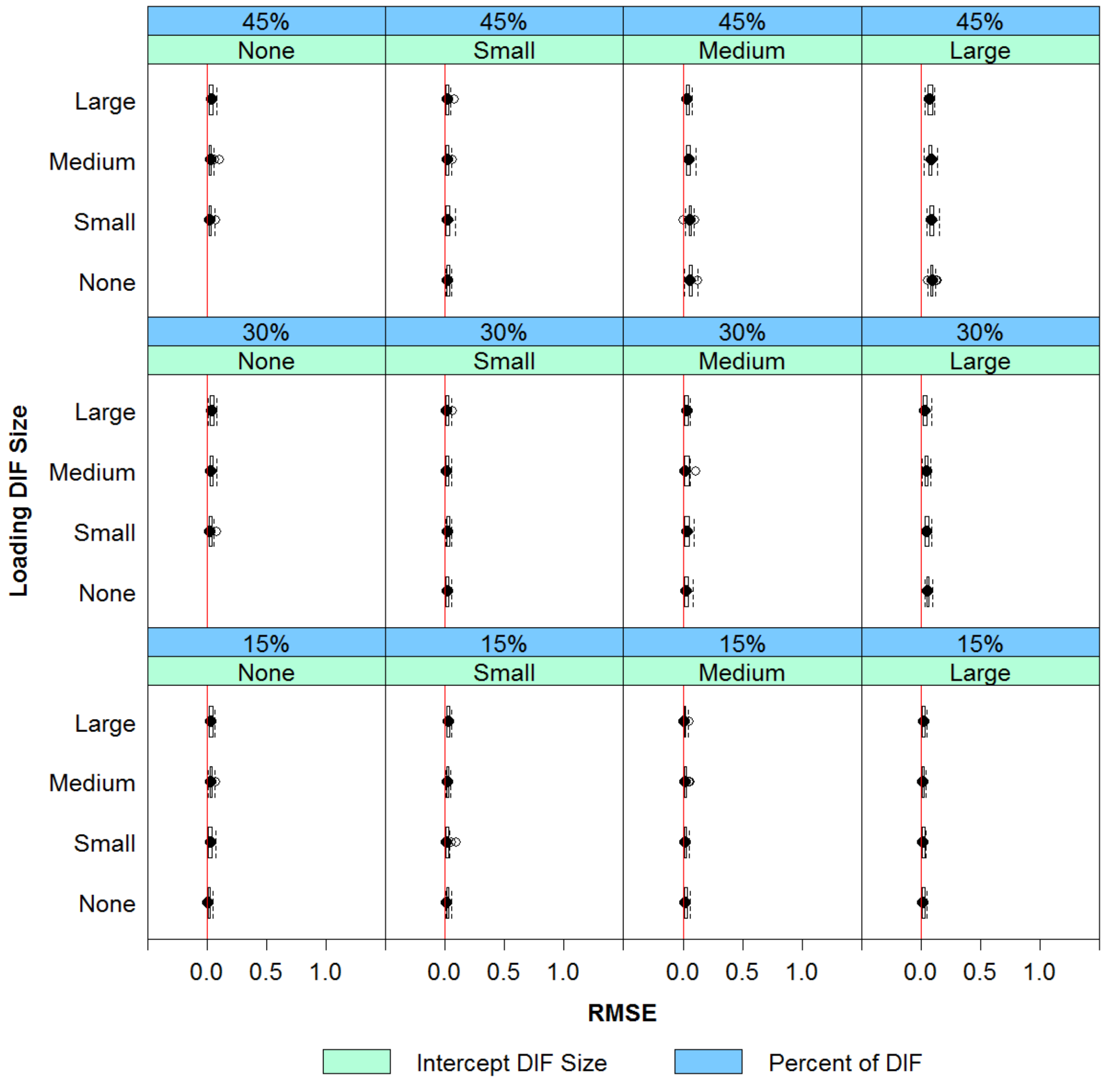

Figure 46. Slope Root Mean Square Error for the Constrained Second-Order Latent Growth Model.

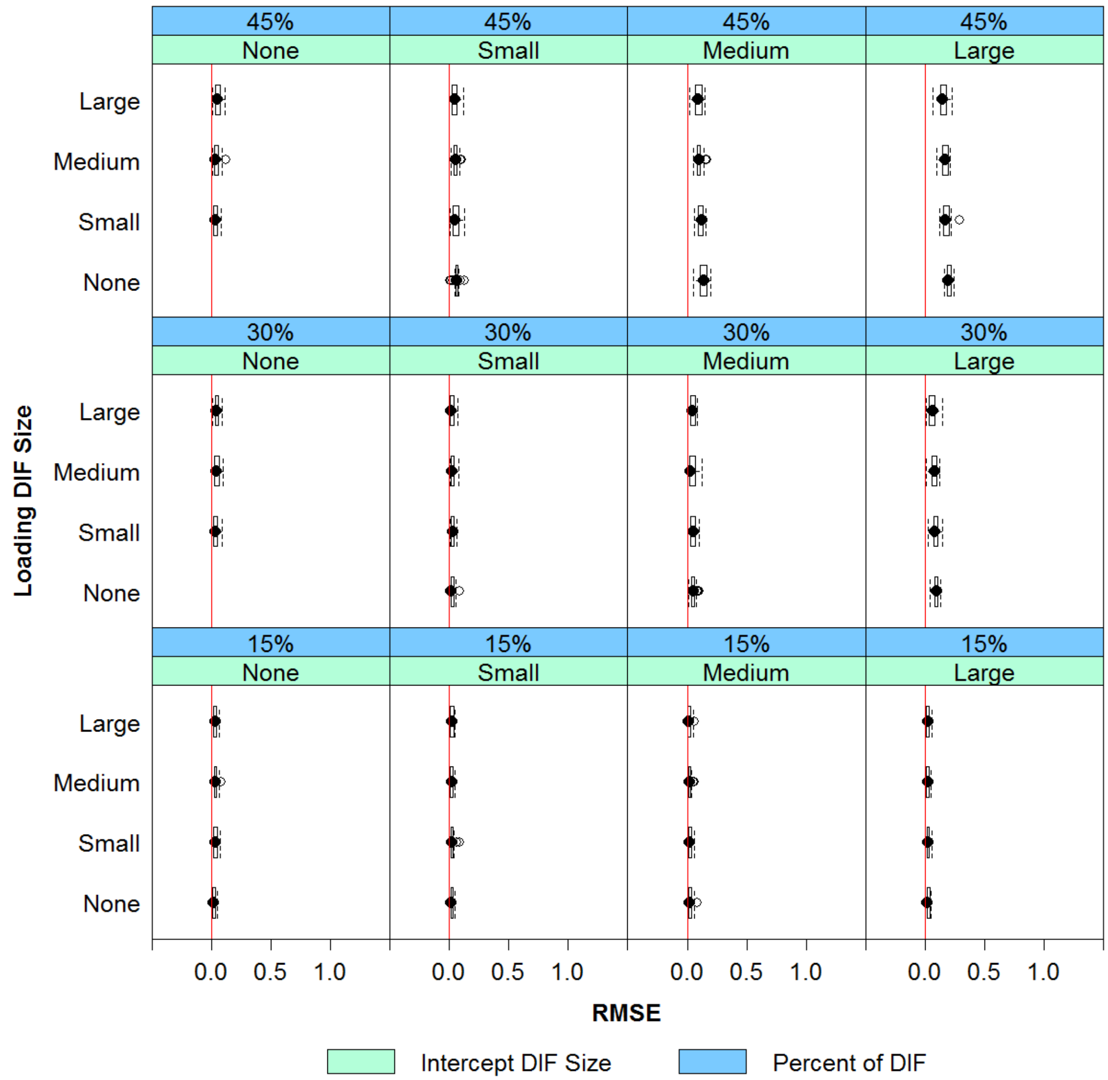

Figure 47. Slope Root Mean Square Error for the Free Second-Order Latent Growth Model.

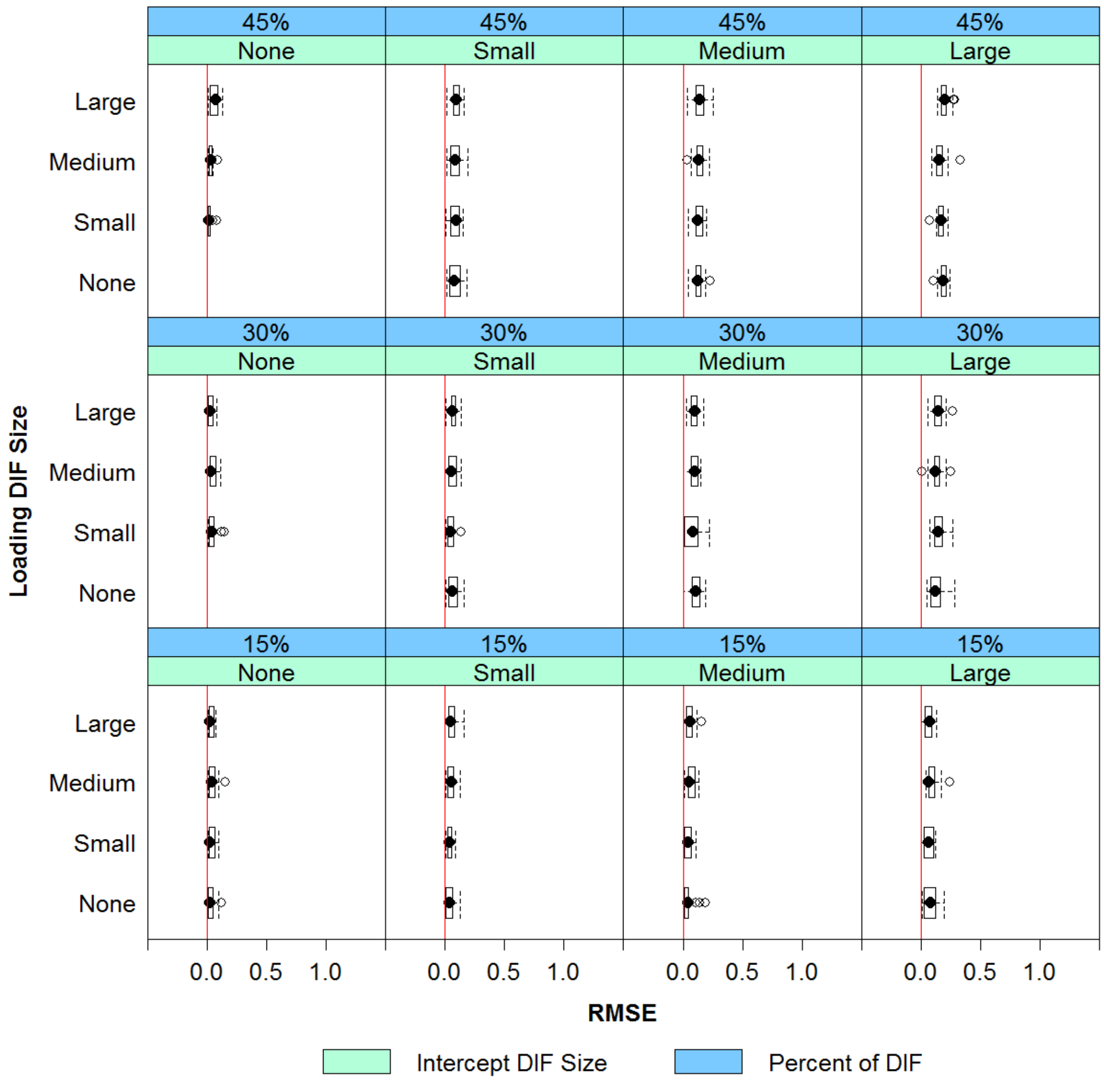

Figure 48. Slope Root Mean Square Error for the First-Order Non-Linear Latent Growth Model.

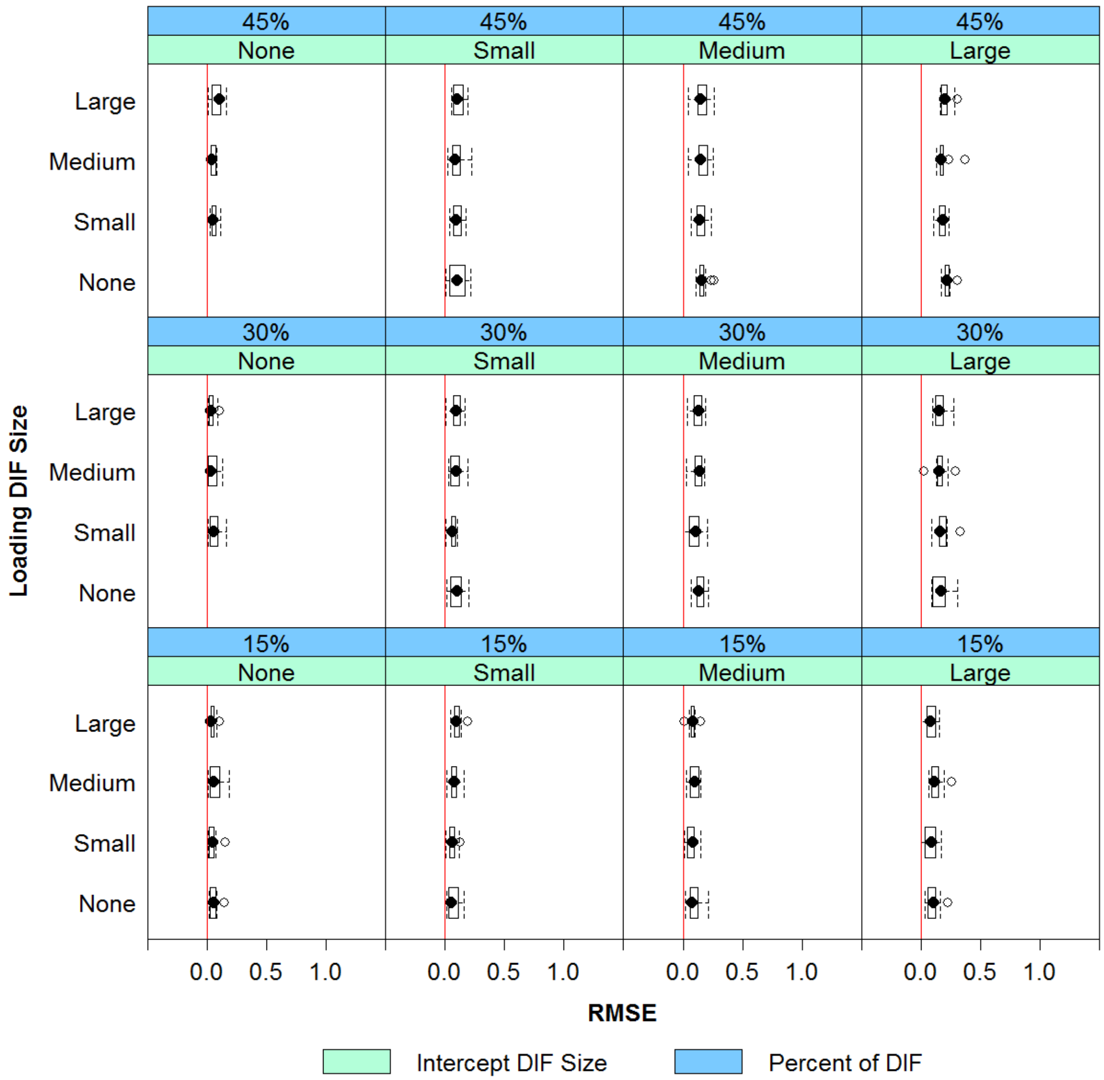

Figure 49. Slope Root Mean Square Error for the Constrained Second-Order Non-Linear Latent Growth Model.

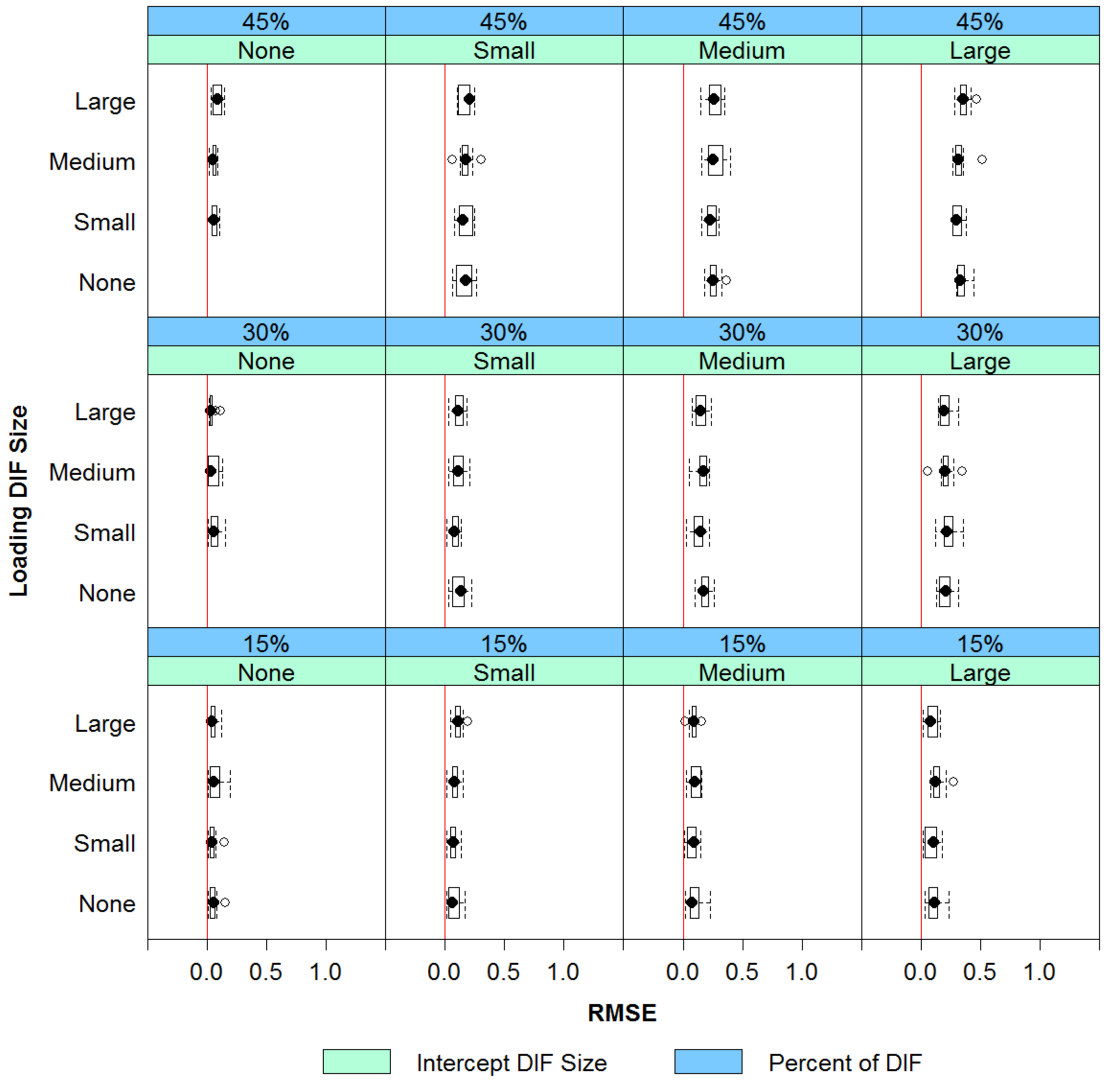

Figure 50. Slope Root Mean Square Error for the Free Second-Order Non-Linear Latent Growth Model.

Slope Variance. The bias estimates for the slope variance parameter are presented in Figures 51-56. The two vertical red lines are plotted at -0.05 and 0.05 to outline the acceptable range. Note that the range of the x-axis for the slope variance charts needed to be changed to be able to plot the residual bias values. The mean and standard deviation values for the relative bias of the slope variance are included in Appendix I. The slope variance parameter had the largest relative bias across all modeling approaches and all conditions. The relative bias for the slope variance parameter in the linear modeling approaches (LGM, 2LGMC, and 2LGM) were the smallest across all modeling approaches, but were not within the acceptable range for any of the conditions. The relative bias for the linear modeling approaches was unrelated to the conditions varied in the study. Compared to the linear models, the non-linear modeling approaches (NLGM, 2NLGMC, and 2NLGMF) had much larger and more variable relative bias estimates for the slope variance parameter. The relative bias for the non-linear modeling approaches were seemingly unrelated to the conditions varied in the study.

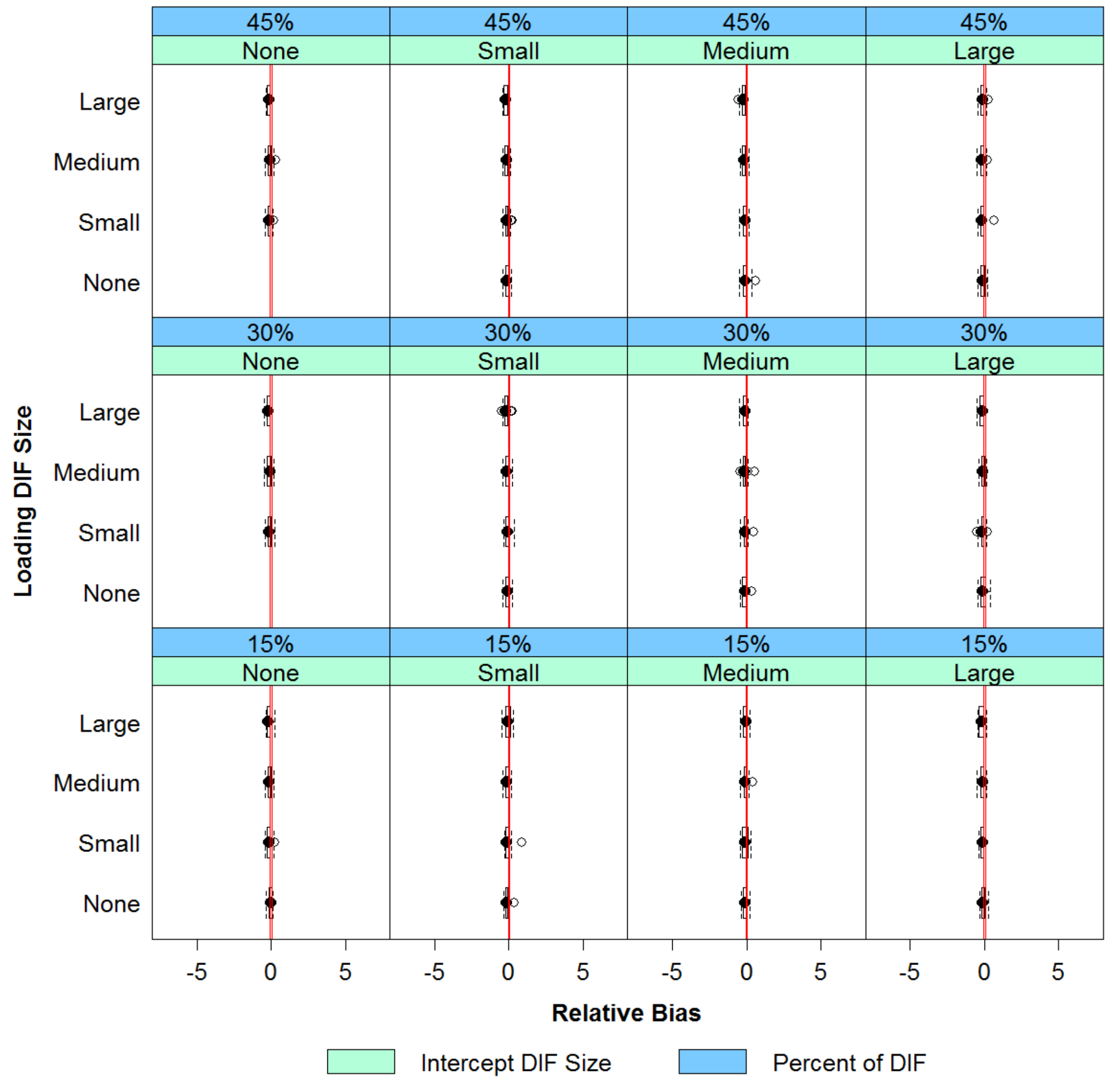

Figure 51. Slope Variance Relative Bias for the First-Order Latent Growth Model.

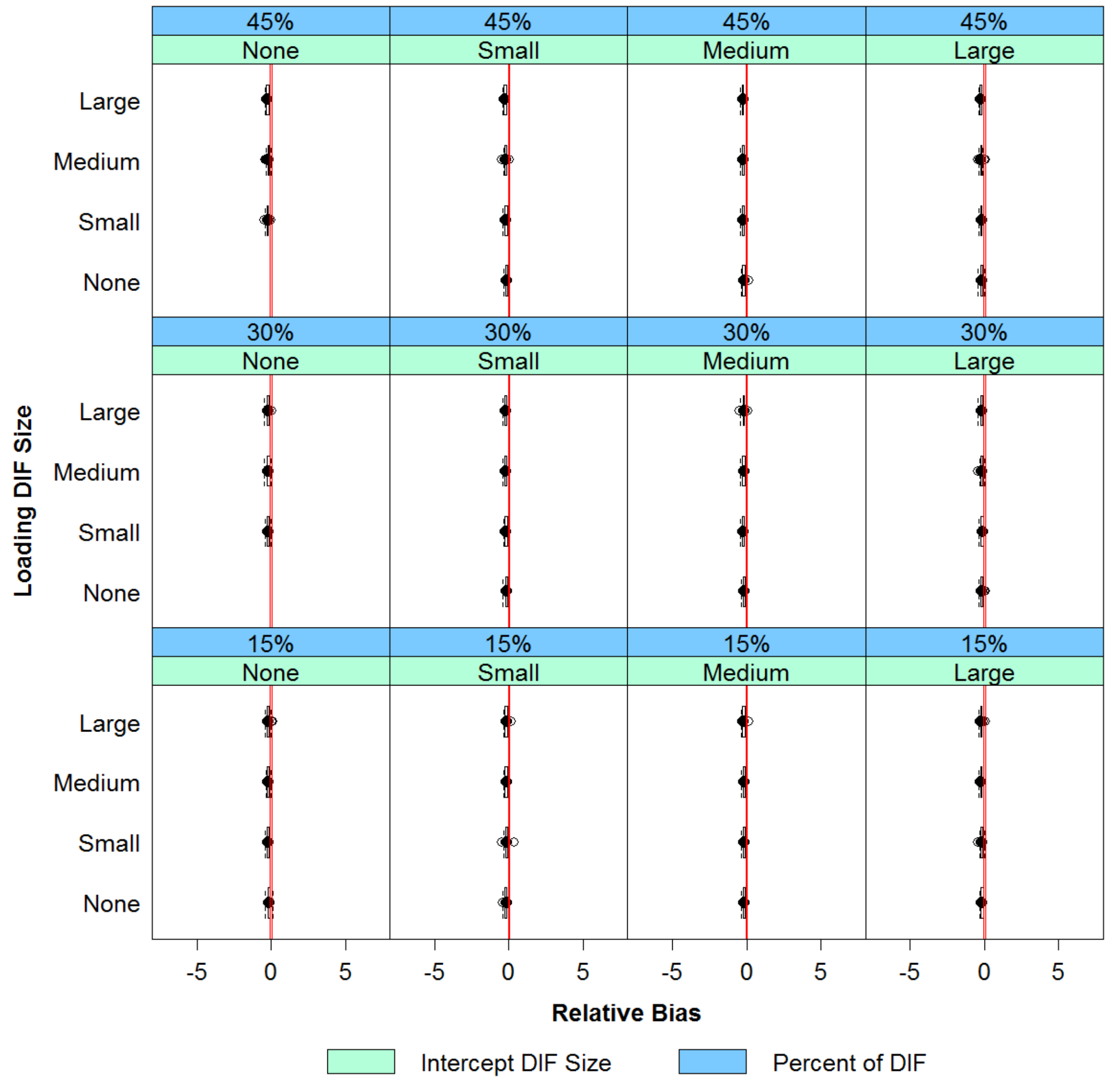

Figure 52. Slope Variance Relative Bias for the Constrained Second-Order Latent Growth Model.

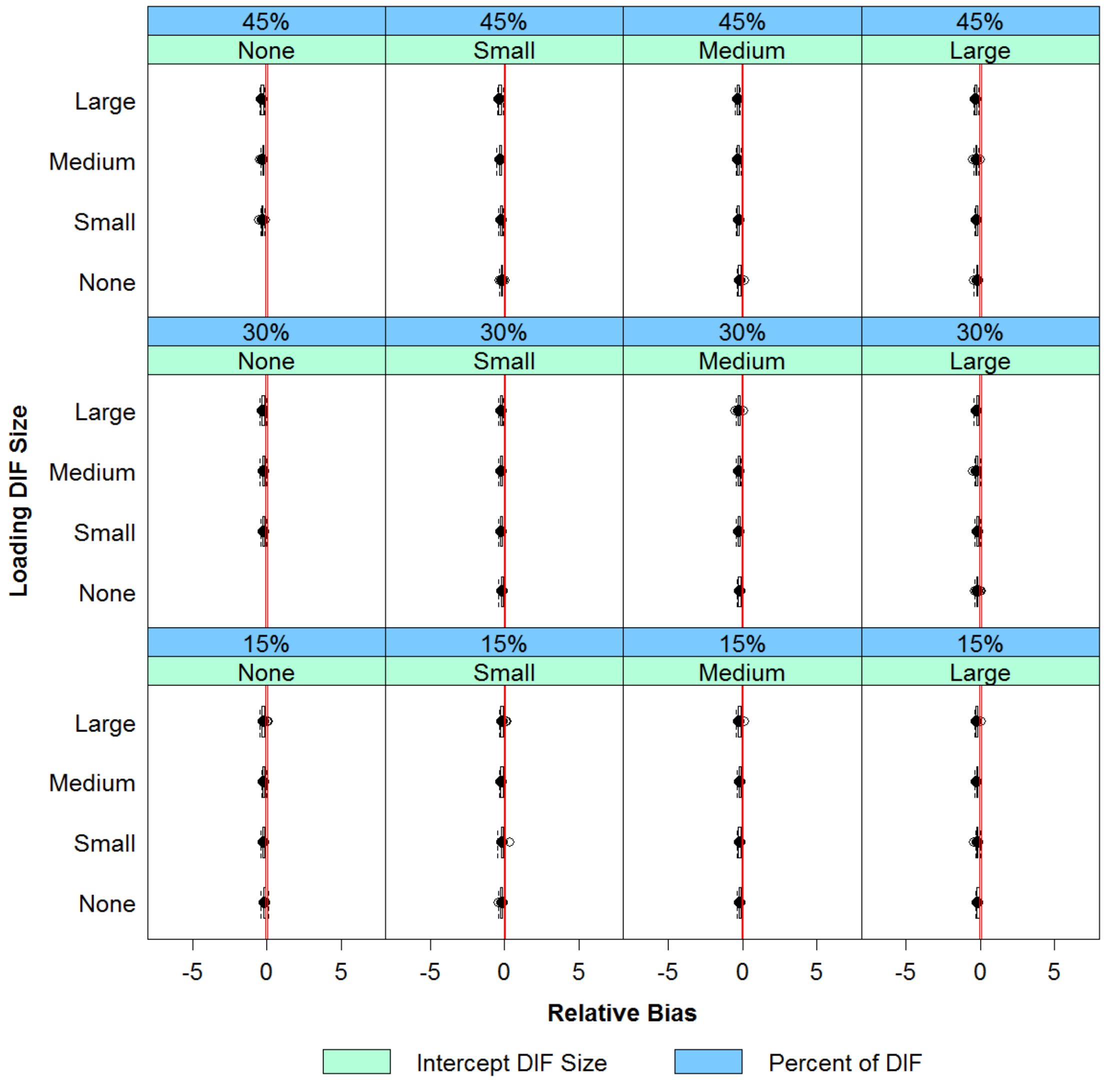

Figure 53. Slope Variance Relative Bias for the Free Second-Order Latent Growth Model.

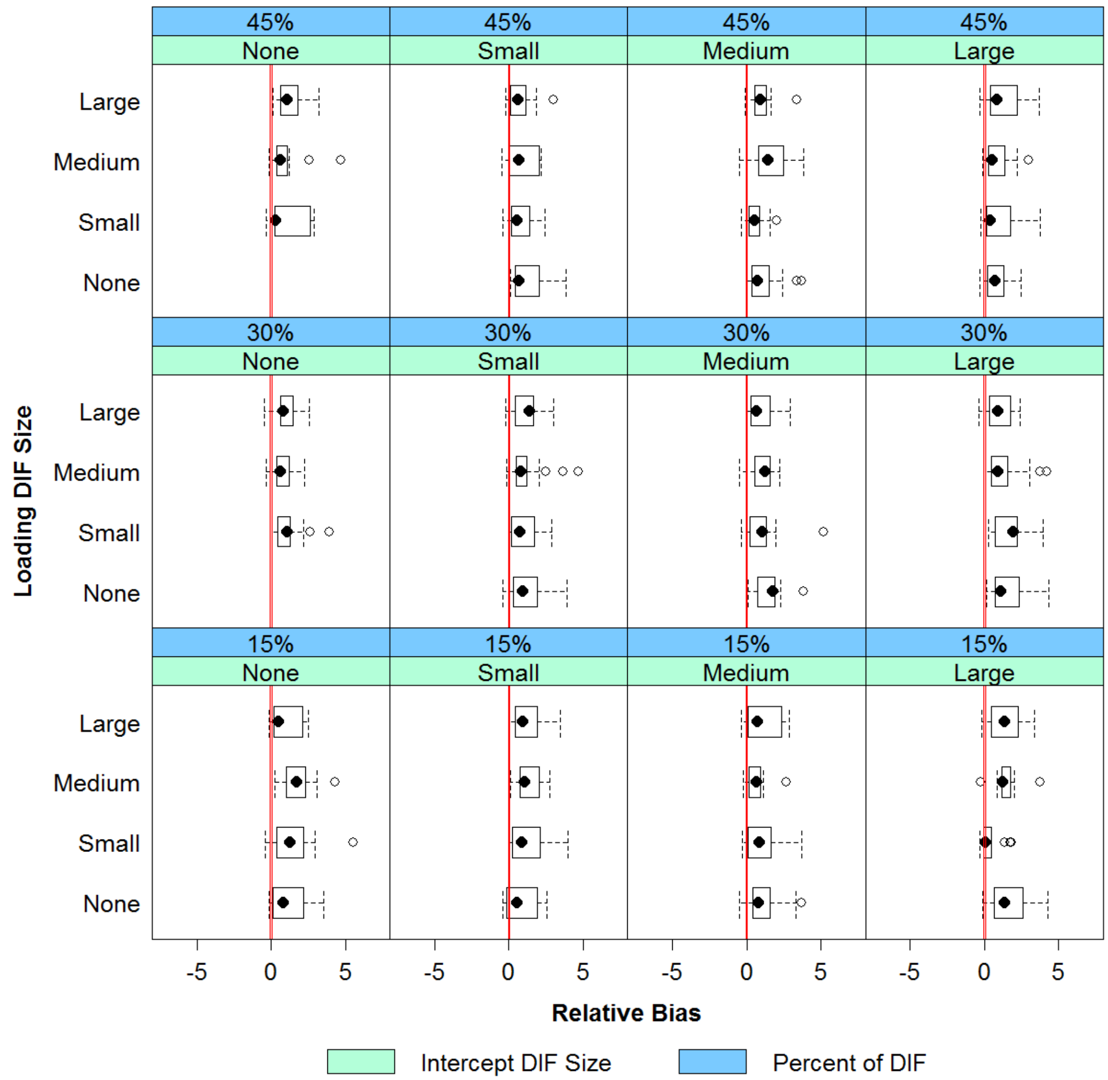

Figure 54. Slope Variance Relative Bias for the First-Order Non-Linear Latent Growth Model.

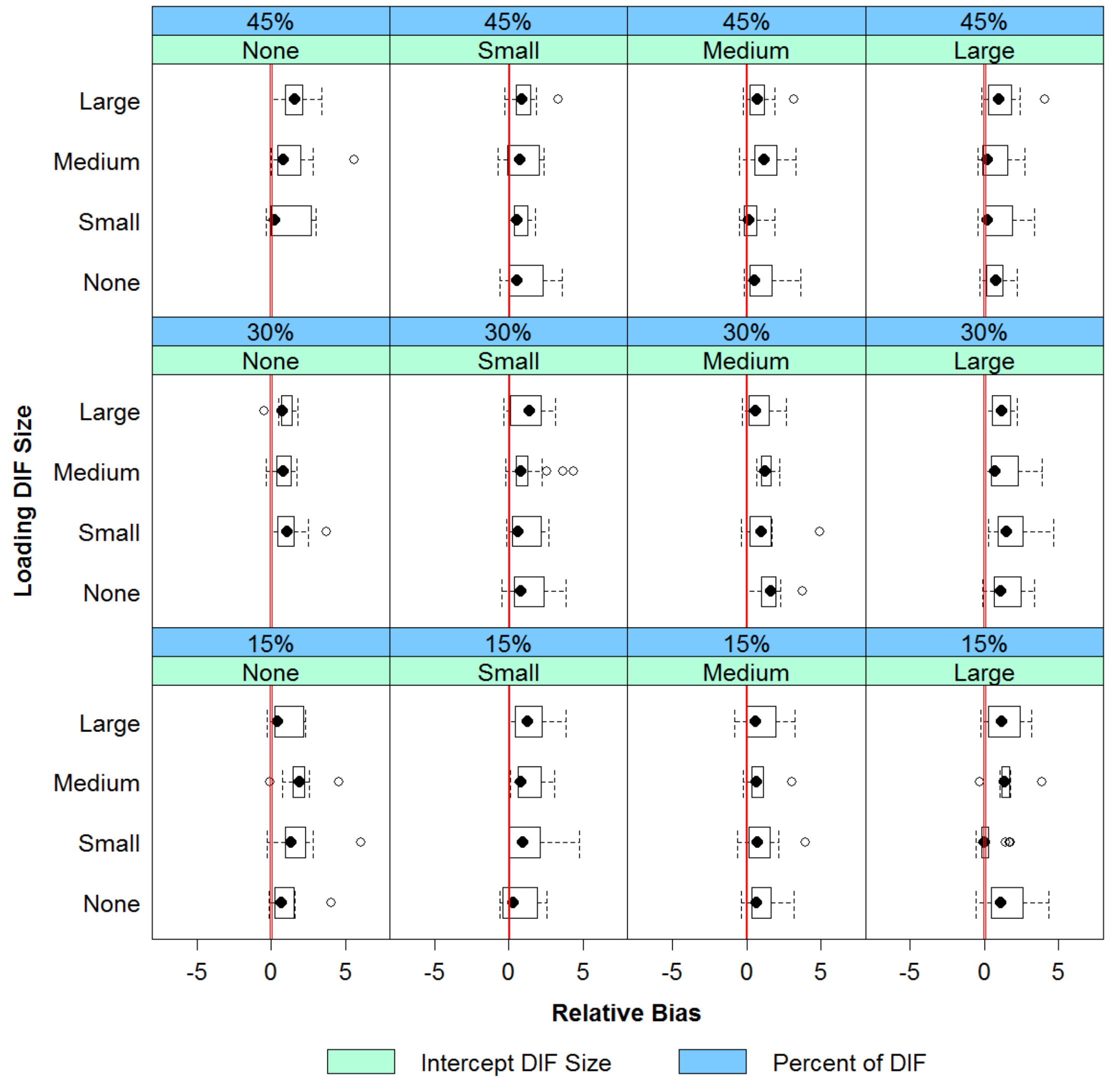

Figure 55. Slope Variance Relative Bias for the Constrained Second-Order Non-Linear Latent Growth Model.

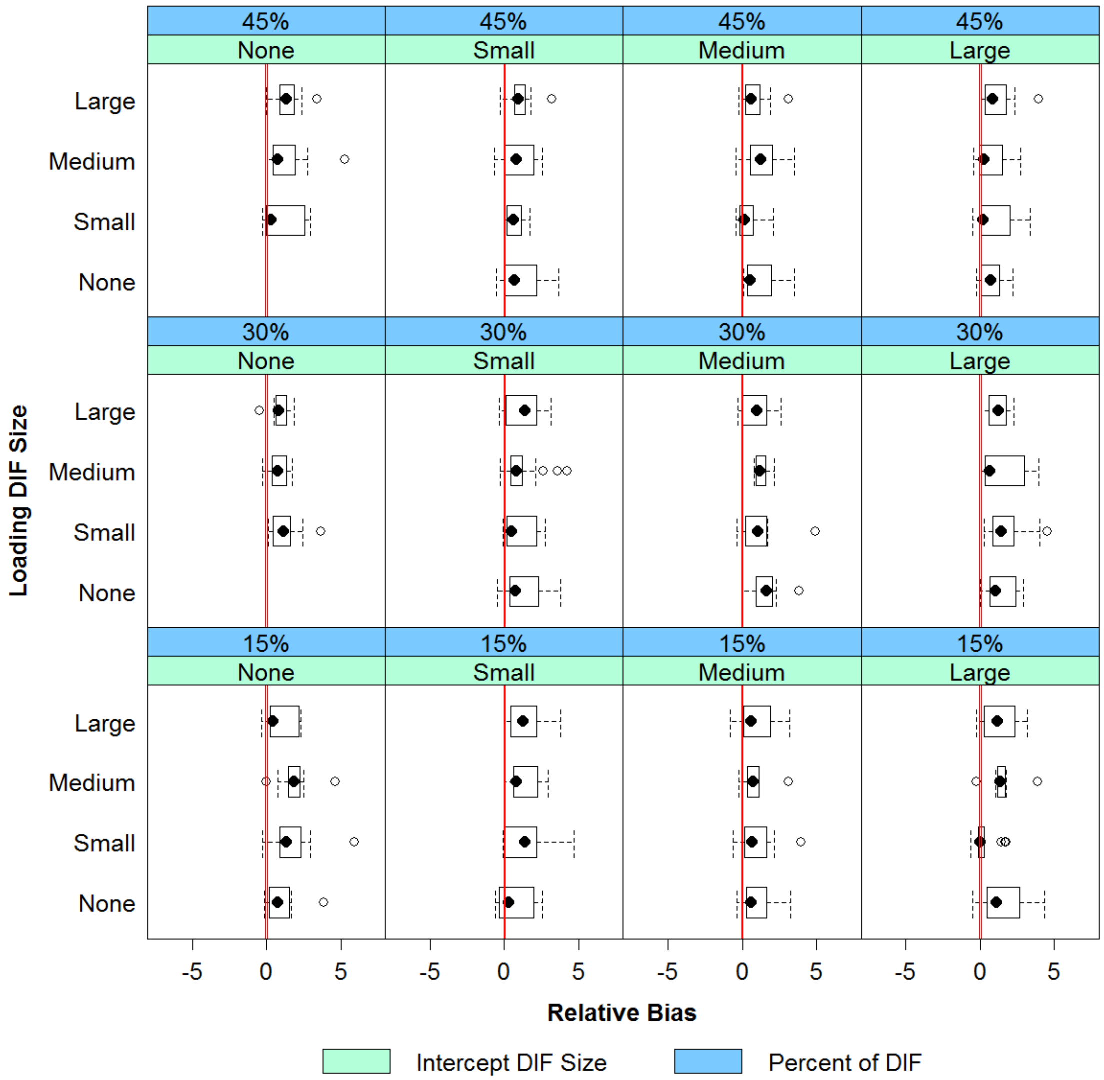

Figure 56. Slope Variance Relative Bias for the Free Second-Order Non-Linear Latent Growth Model.

The RMSE estimates for the intercept variance parameter are presented in Figures 57-62. The mean and standard deviation values for the relative bias of the slope variance are included in Appendix J. The slope variance RMSE values for the linear modeling approaches (LGM, 2LGMC, and 2LGMF) were near zero for all conditions varied in the study. The non-linear modeling approaches (NLGM, 2NLGMC, and 2NLGMF) had larger RMSE values and had more variability in the slope variance RMSE values. As with the linear models, the values were not related to the conditions varied in the study.

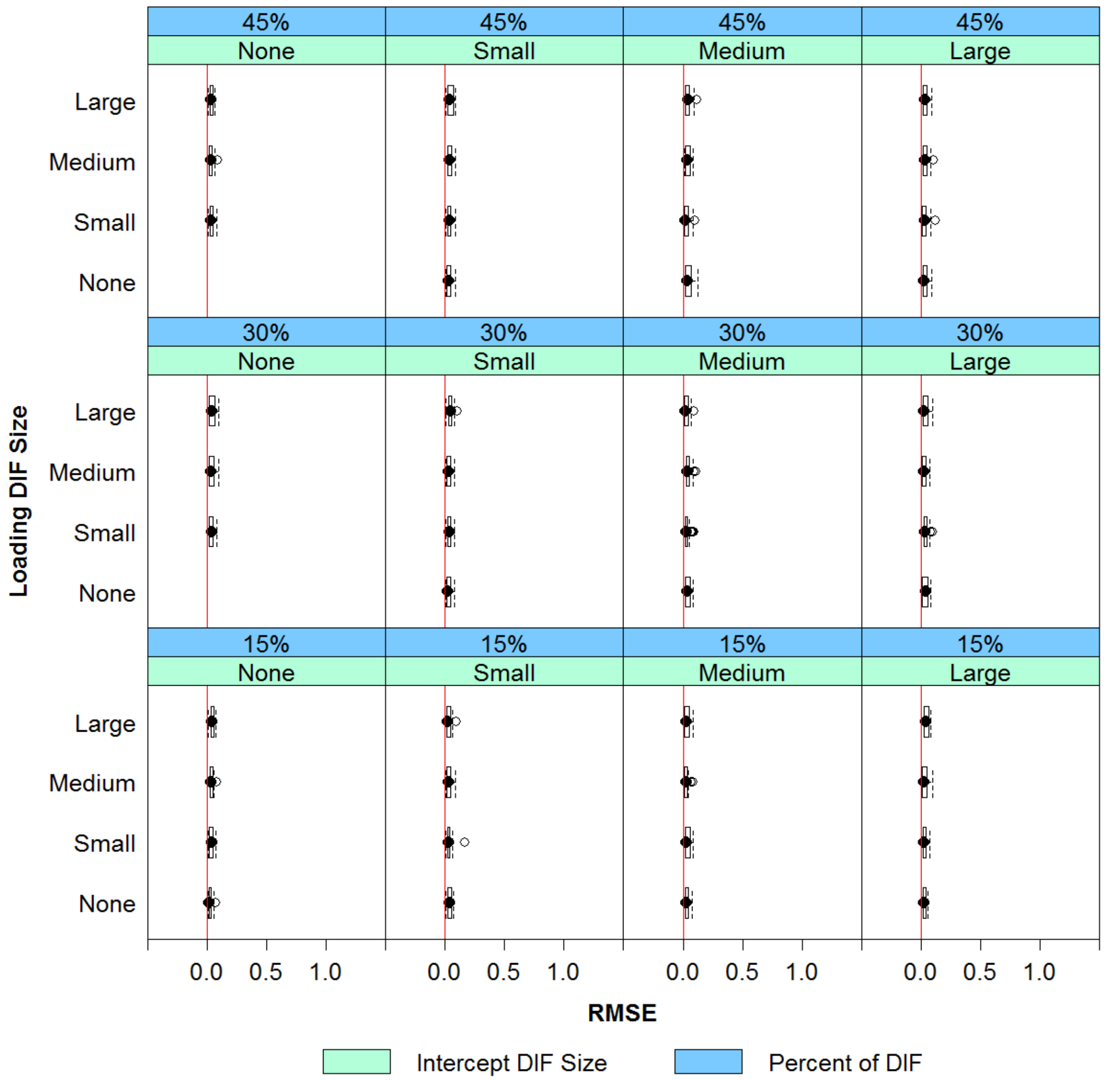

Figure 57. Slope Variance Root Mean Square Error for the First-Order Latent Growth Model.

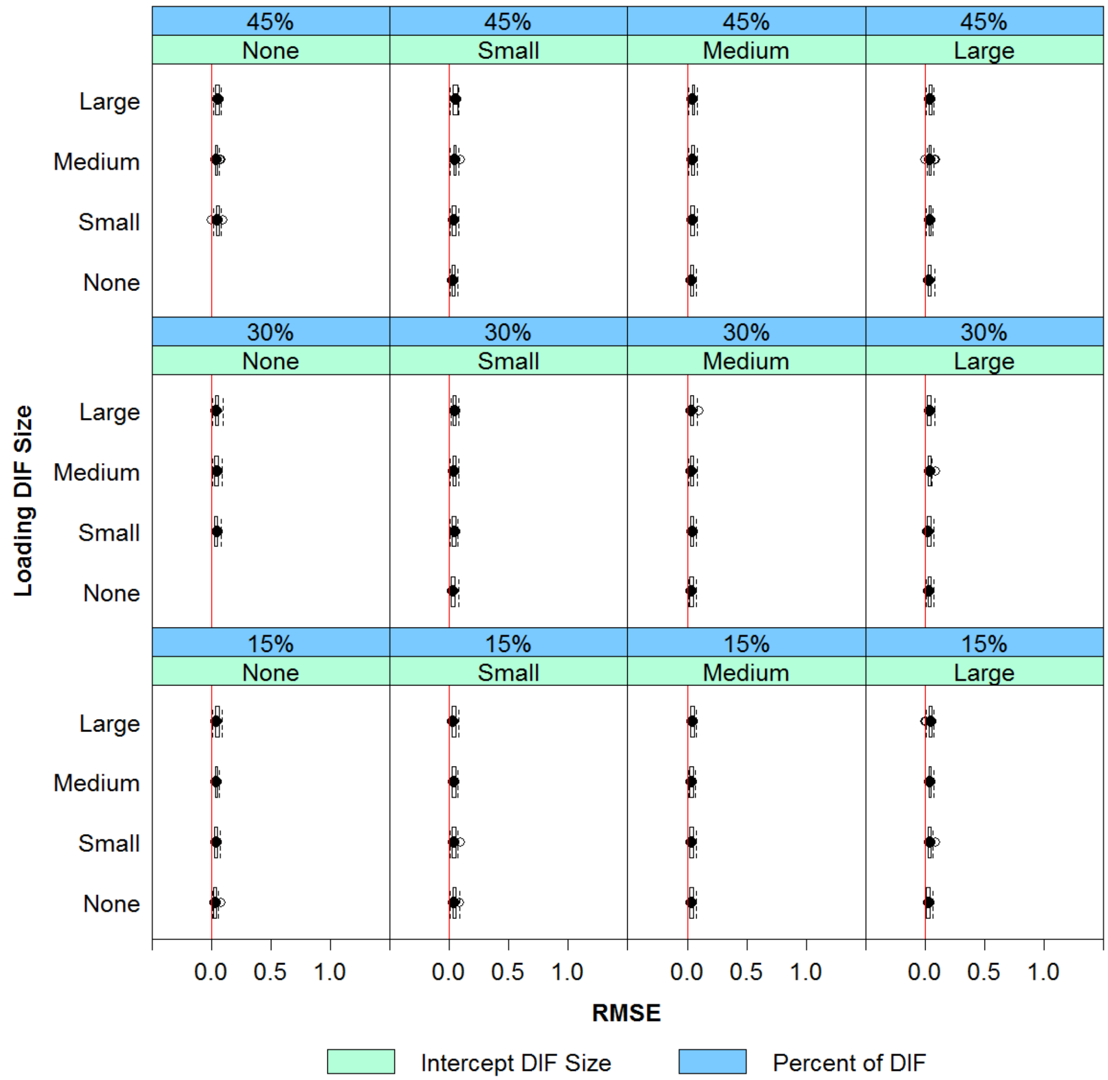

Figure 58. Slope Variance Root Mean Square Error for the Constrained Second-Order Latent Growth Model.

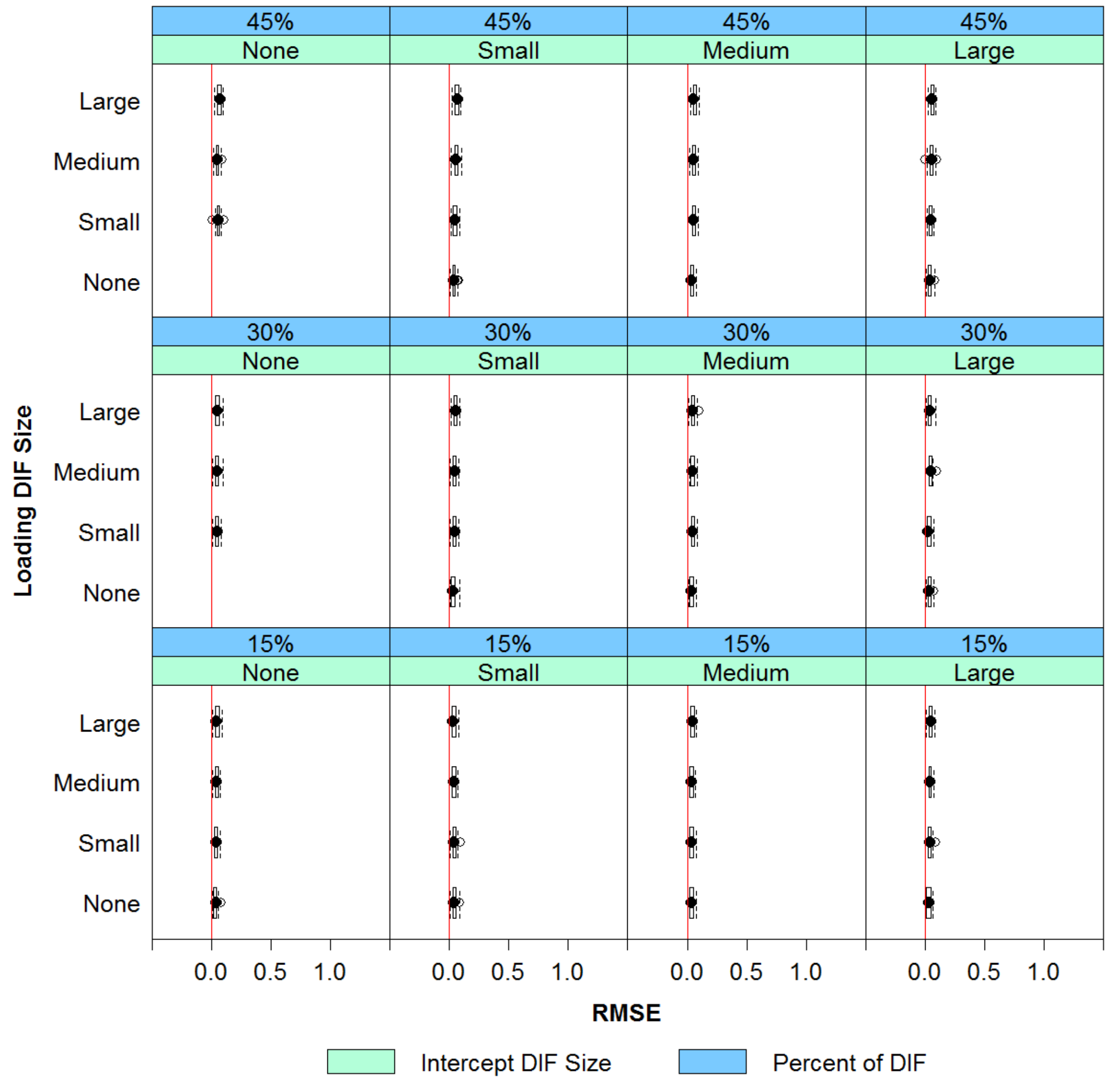

Figure 59. Slope Variance Root Mean Square Error for the Free Second-Order Latent Growth Model.

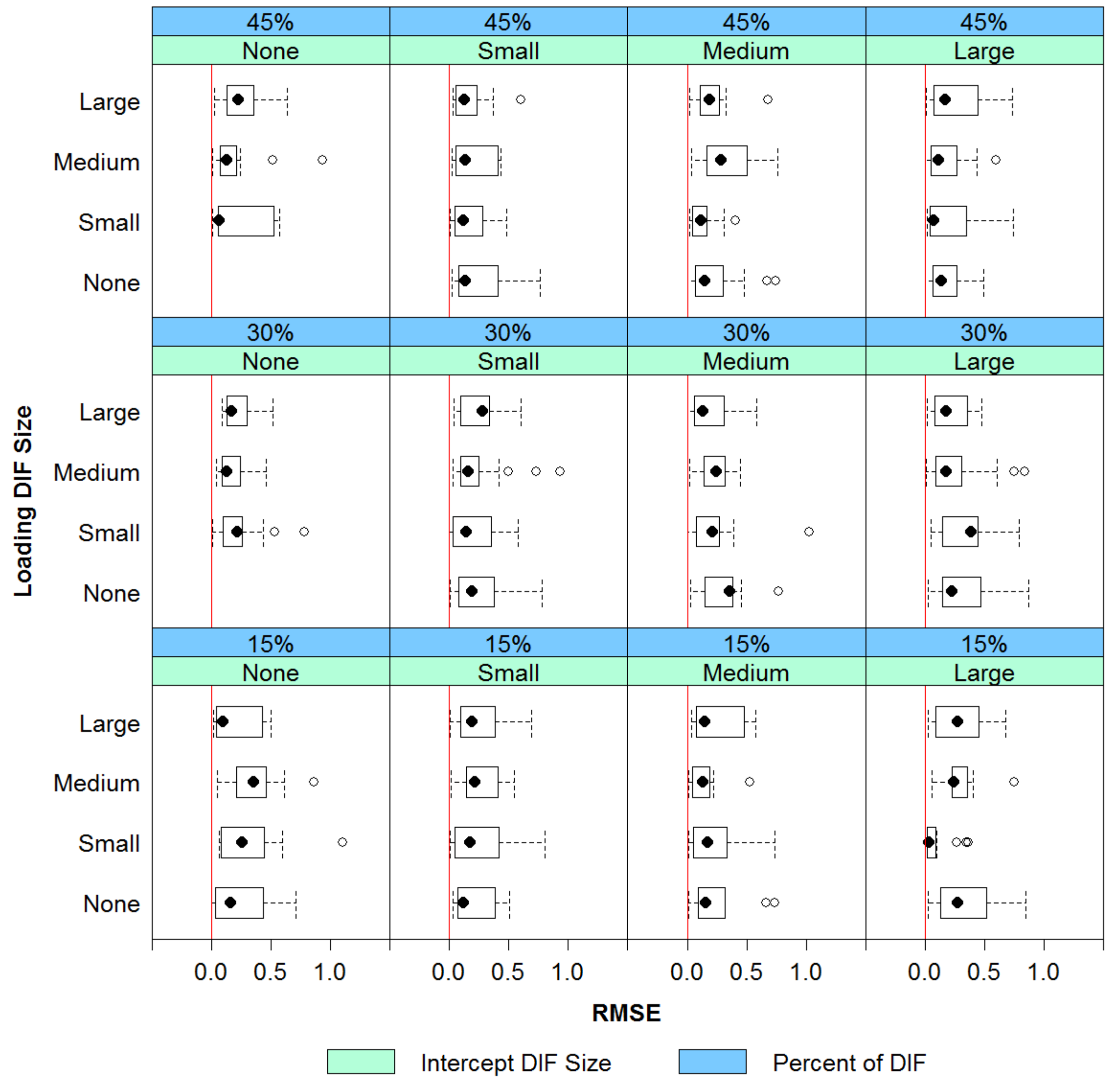

Figure 60. Slope Variance Root Mean Square Error for the First-Order Non-Linear Latent Growth Model.

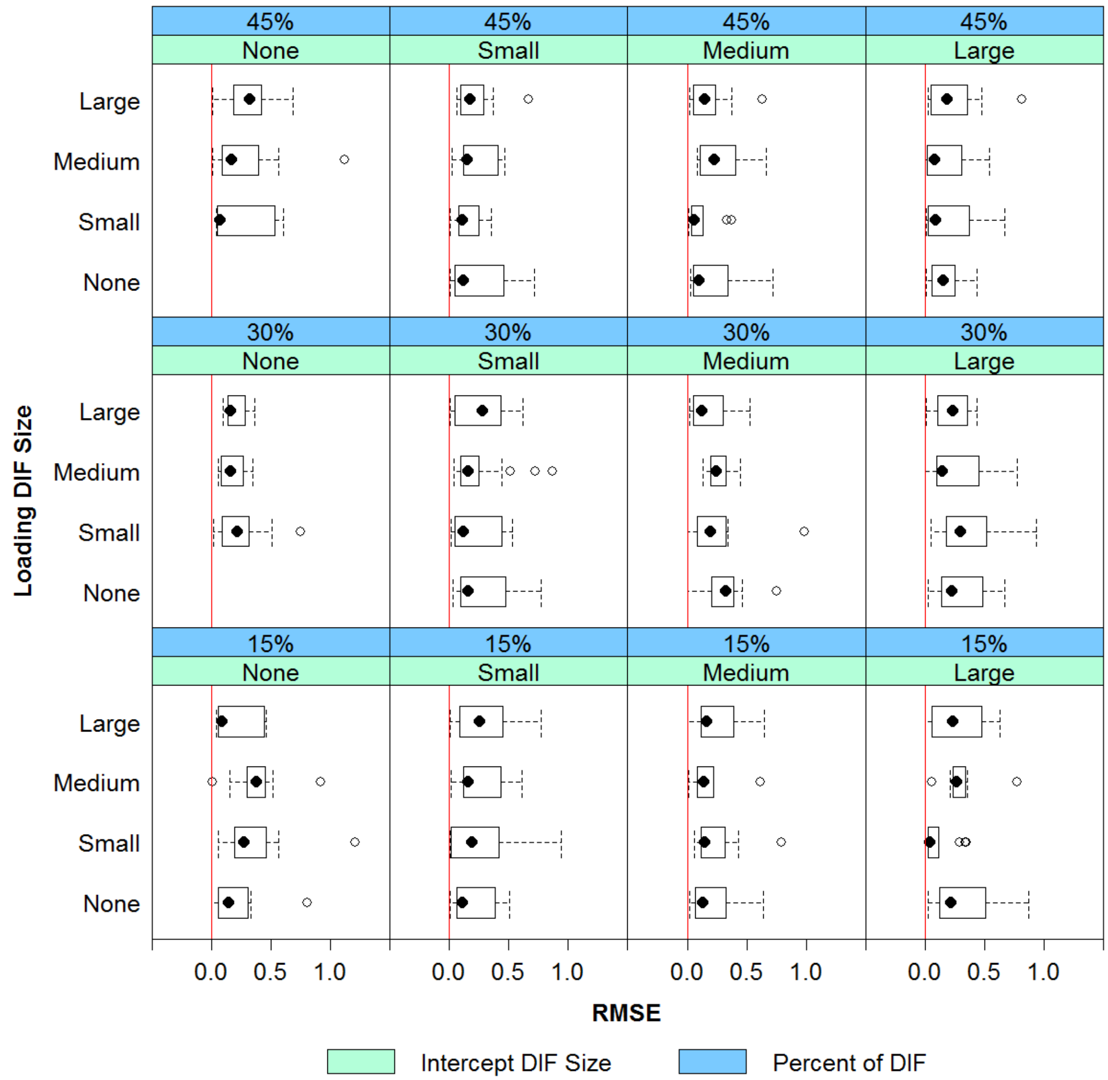

Figure 61. Slope Variance Root Mean Square Error for the Constrained Second-Order Non-Linear Latent Growth Model.

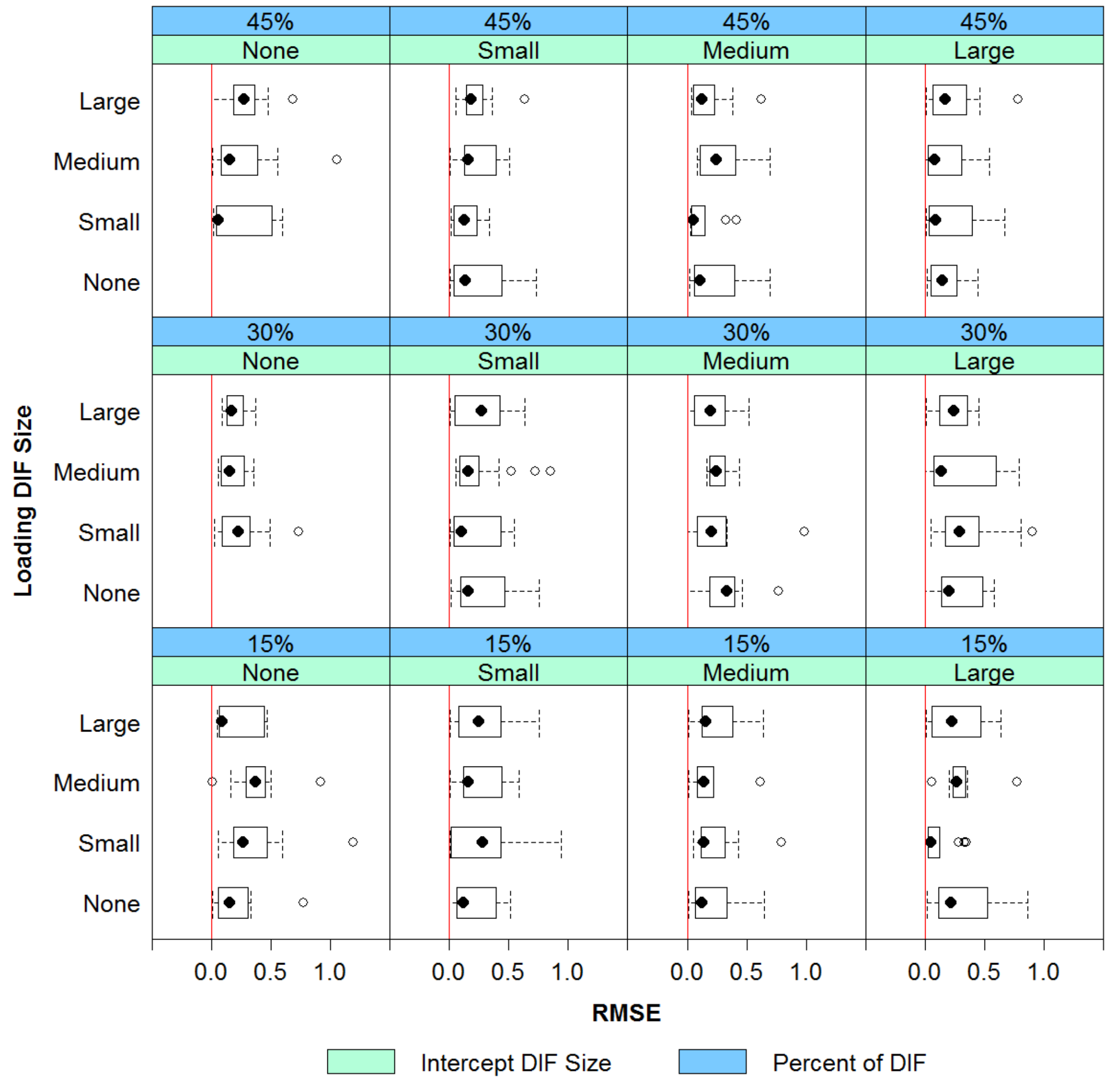

Figure 62. Slope Variance Root Mean Square Error for the Free Second-Order Non-Linear Latent Growth Model.

Incorrectly Specified Growth Parameters

There were several parameters modeled in the modeling approaches that were not included in the generating model. To investigate whether the models correctly recovered these parameters the *p*-values and estimates were examined. Ideally the parameter estimates that were not included in the generating model should be zero, thus a significant *p*-value would indicate poor parameter recovery (i.e., Type I error). For the linear models, the only parameter that was set to zero in the generating model was the covariance between the intercept and slope. For the non-linear models, the quadratic slope, quadratic slope variance, and covariances (intercept with slope, intercept with quadratic slope, and slope with quadratic slope) were all zero in the generating model.

Intercept-Slope Covariance. The *p*-values associated with the estimate of the intercept-slope covariance are presented in Figures 63-68. A red line is drawn at 0.05 to represent the critical value with which the *p*-value is often compared to determine statistical significance. The mean and standard deviation values for the *p*-value of the intercept-slope covariance are included in Appendix K. The *p*-values for the first-order linear model were often non-significant in the conditions with a small proportion of DIF and no intercept DIF, but still had a larger Type I error than typically acceptable (41% for the conditions with no intercept DIF and a small number of items with all sizes of loading DIF). In the conditions with 30% of items with DIF and small, medium, or large intercept DIF often had significant *p*-values associated with the covariance between the intercept and slope. In the conditions with 45% of items with DIF, the variability in *p*-values decreased as the size of the intercept DIF increased. The Type I error rate for the condition with 45% of items with large intercept DIF, aggregated across all loading DIF sizes, was 73%. The *p*-values for the second-order linear models (2LGMC and 2LGMF) suggested that the estimate for the intercept-slope covariance was often statistically significant (81% across all conditions in the 2LGMC and 83% across all conditions in the

2LGMF). The *p*-values for the second-order linear models were slightly less variable when the size of intercept DIF was large and when there was a large percent of DIF items. The *p*-values for non-linear models (NLGM, 2NLGMC, and 2NLGMF) suggested that the estimate for the intercept-slope parameter was typically not significant. The Type I error rate for the covariance between the intercept and slope was 0.4%, 1%, and 1% across all conditions for the NLGM, 2NLGMC, and 2NLGMF, respectively. Thus, the non-linear models did a markedly better job recovering the intercept-slope covariance parameter than the linear models.

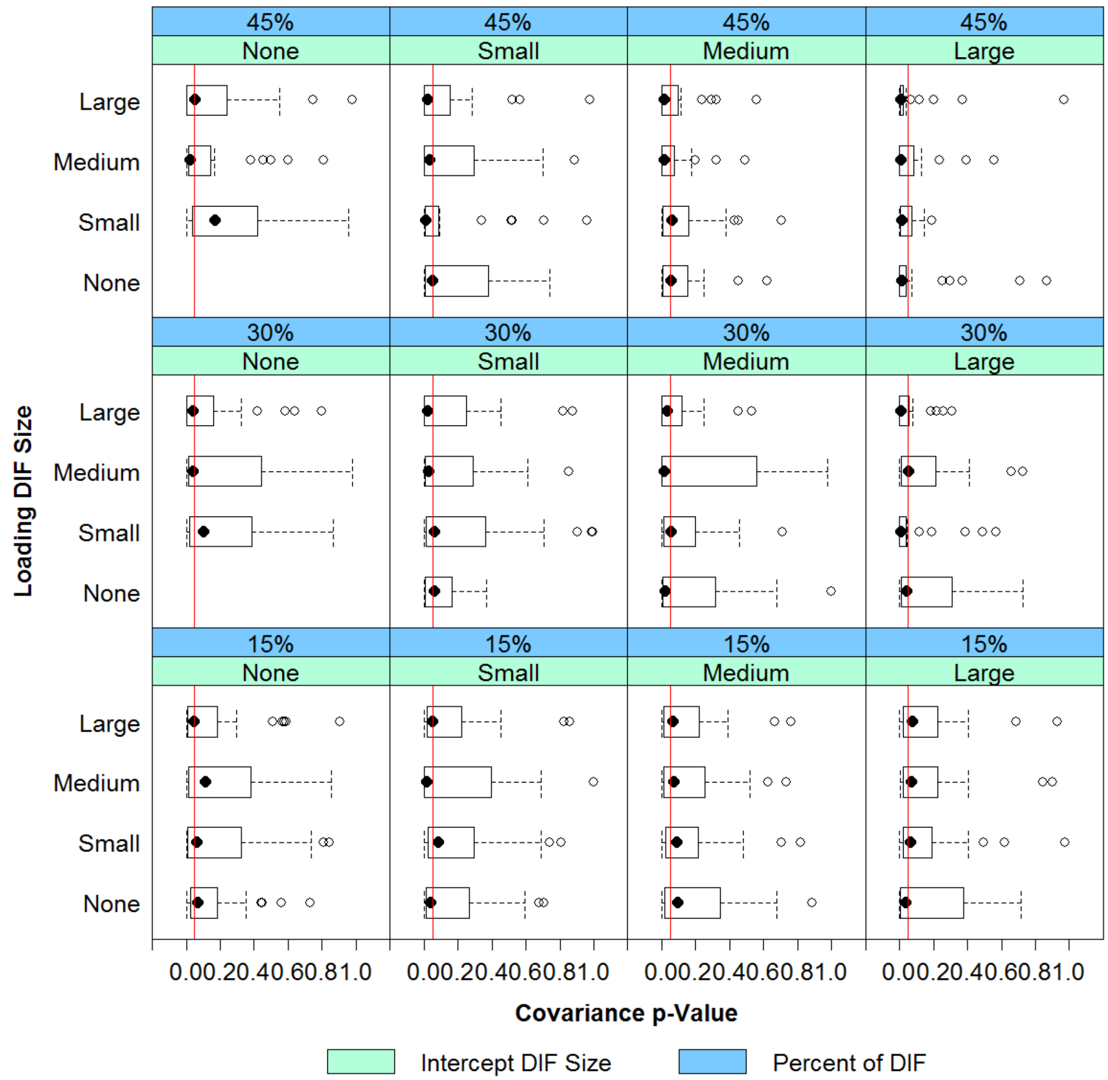

Figure 63. Intercept-Slope Covariance *p*-Value for the First-Order Latent Growth Model.

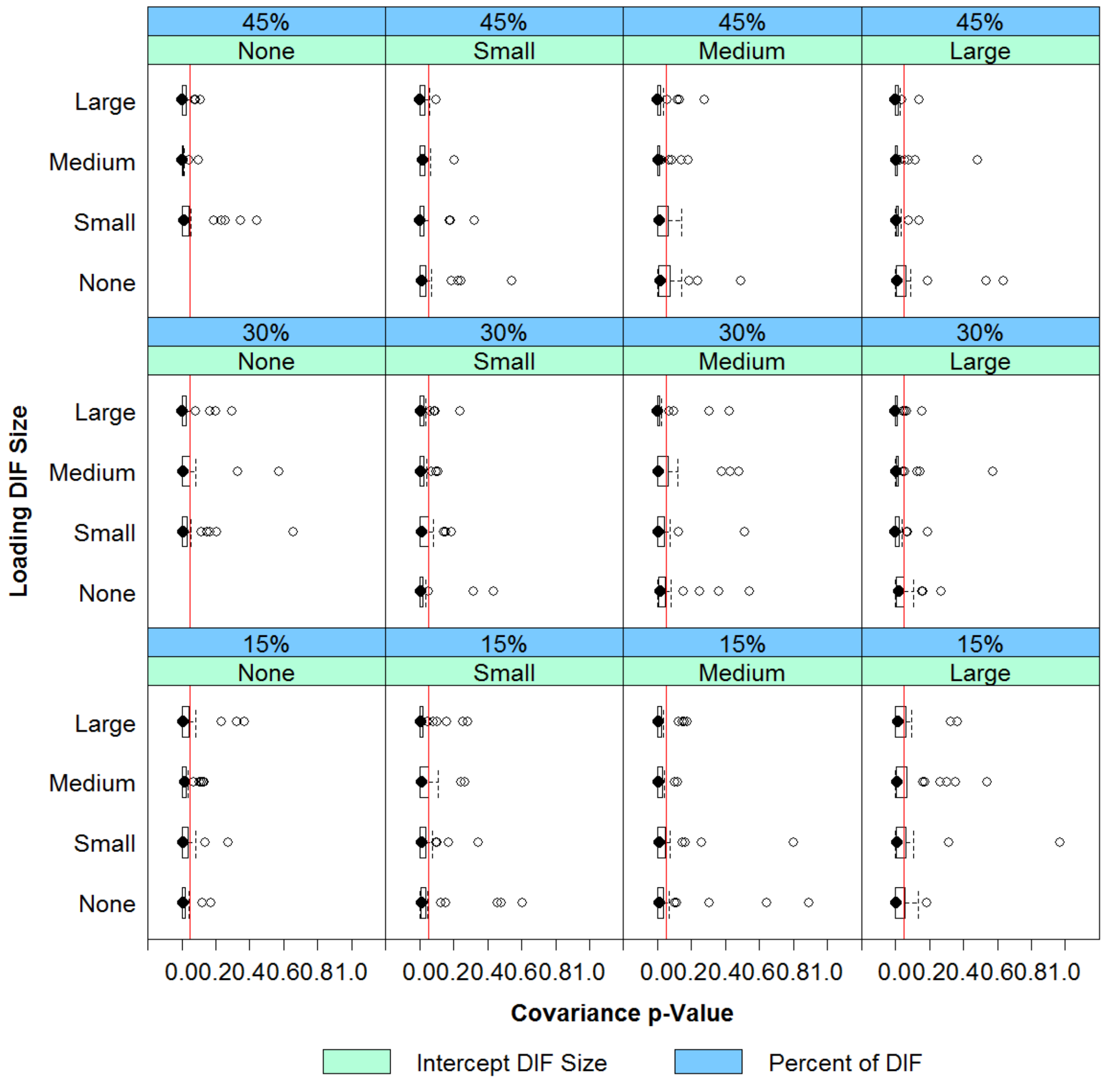

Figure 64. Intercept-Slope Covariance *p*-Value for the Constrained Second-Order Latent Growth Model.

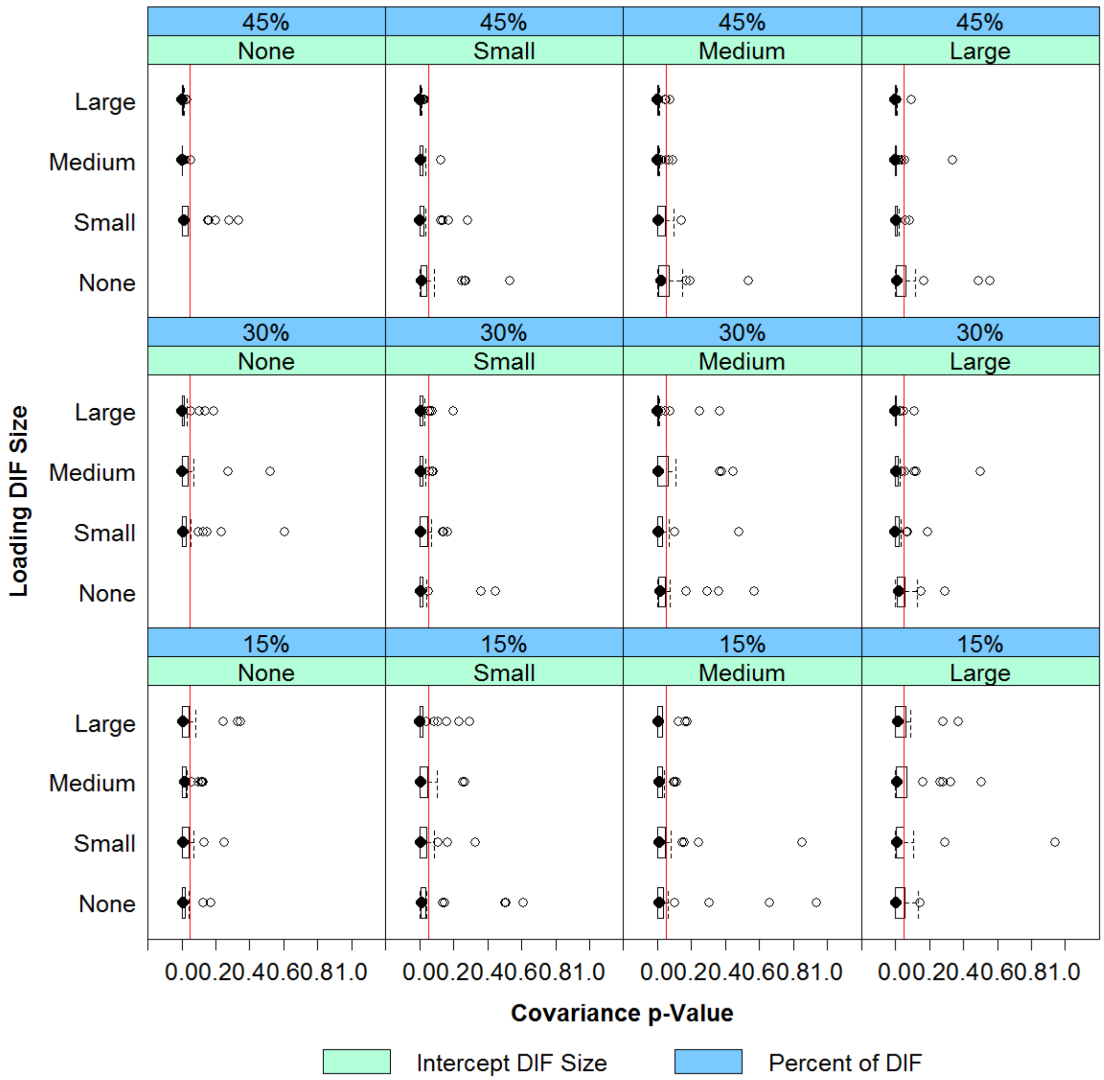

Figure 65. Intercept-Slope Covariance *p*-Value for the Free Second-Order Latent Growth Model.

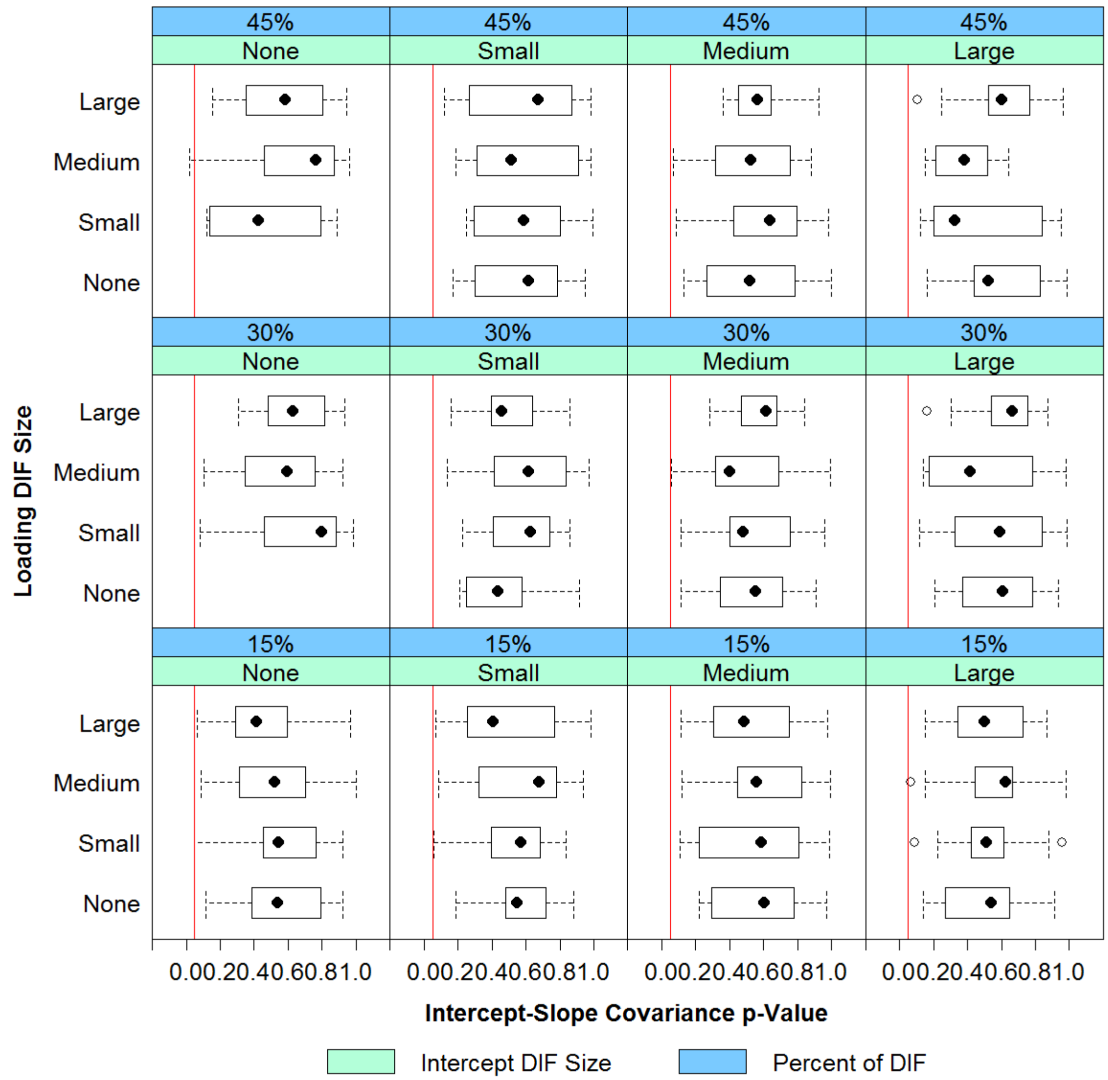

Figure 66. Intercept-Slope Covariance *p*-Value for the First-Order Non-Linear Latent Growth Model.

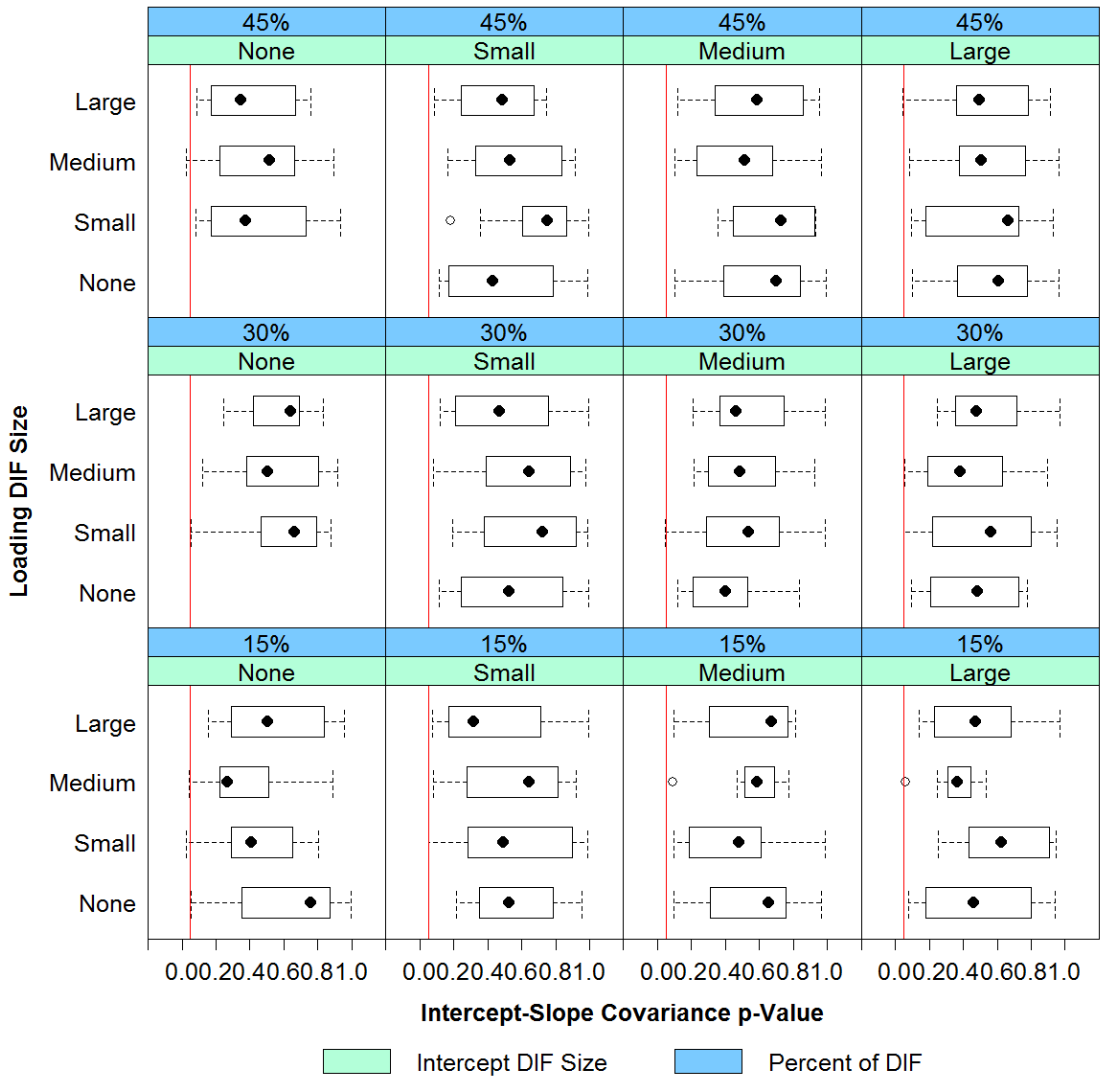

Figure 67. Intercept-Slope Covariance *p*-Value for the Constrained Second-Order Non-Linear Latent Growth Model.

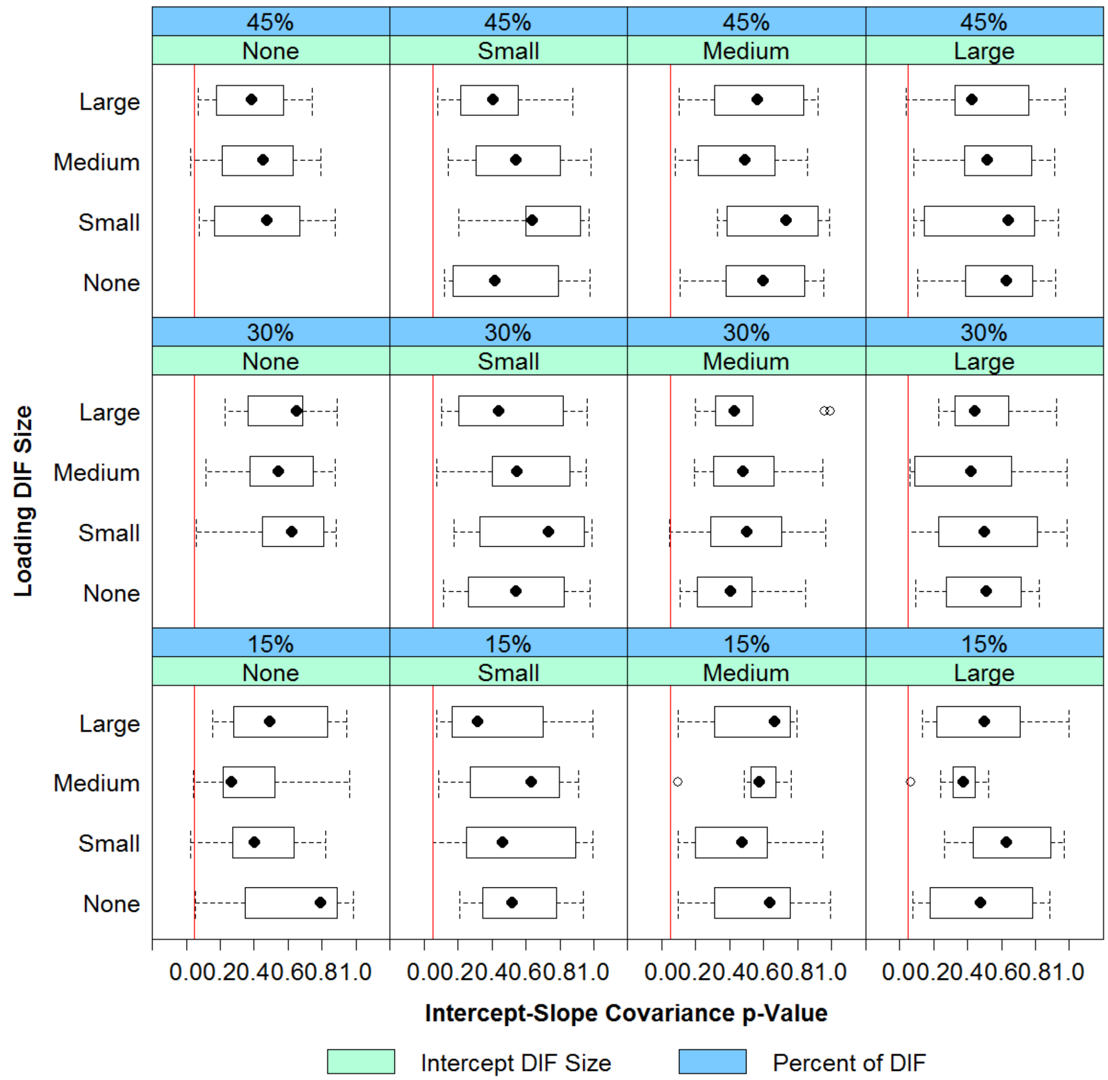

Figure 68. Intercept-Slope Covariance *p*-Value for the Free Second-Order Non-Linear Latent Growth Model.

The estimated values for the intercept-slope covariance parameter are presented in Figures 69-74. The mean and standard deviation values for the intercept-slope covariance estimates are included in Appendix L. Although the *p*-values for the linear models often suggested that the covariance between the intercept and slope was often statistically significant, the estimated values were near zero for all conditions. The linear model results suggested that the covariance parameter was estimated as statistically, but not practically, significant. The median values for the first-order non-linear model was also near zero, but were more variable than in the linear conditions. The median values for the second-order non-linear models (2NLGMC and 2NLGMF) were slightly larger in magnitude than the linear models, and were notably more variable. For the linear and non-linear models, the covariance between the intercept and slope was often estimated as negative.

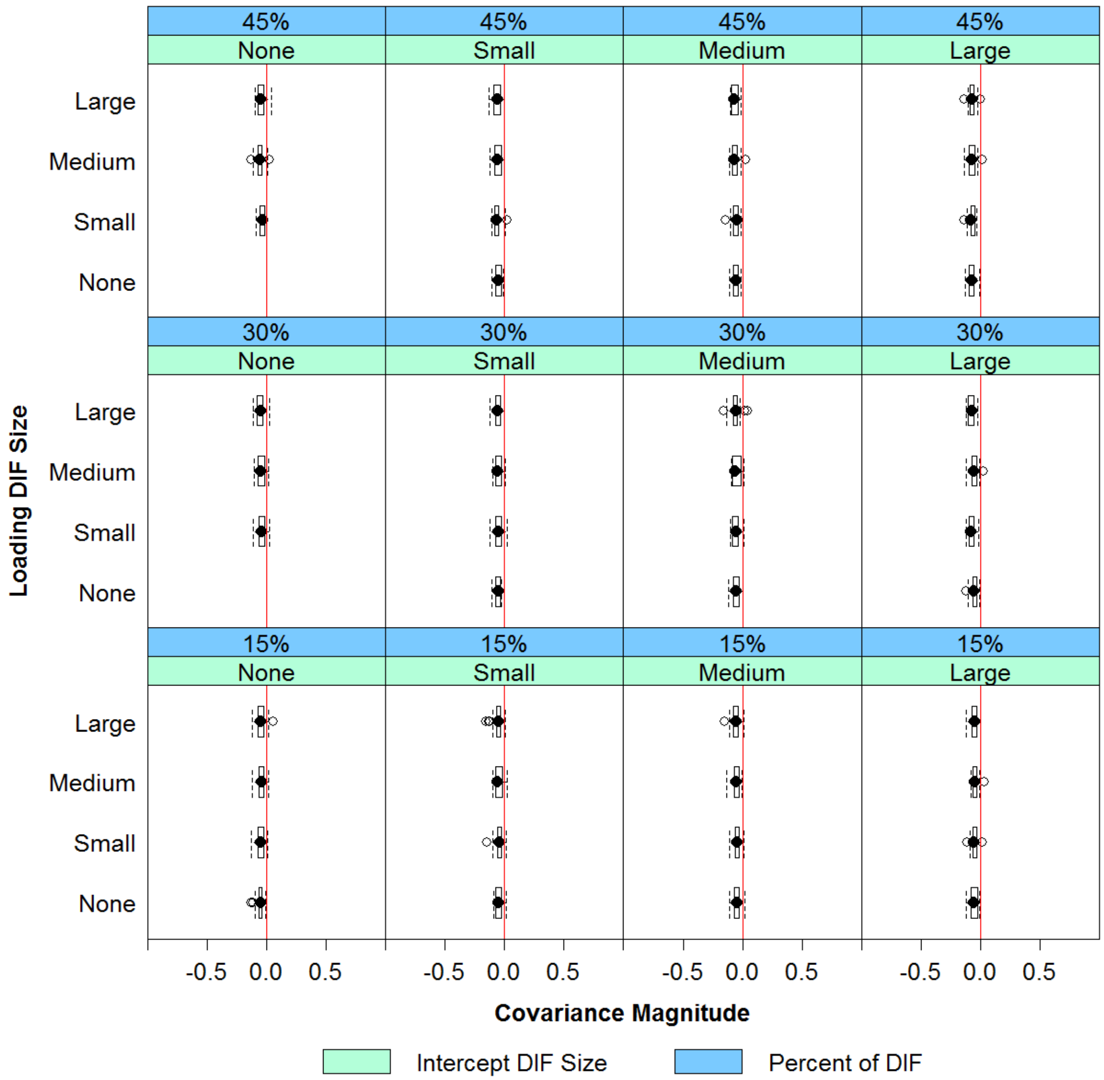

Figure 69. Intercept-Slope Covariance Estimates for the First-Order Latent Growth Model.

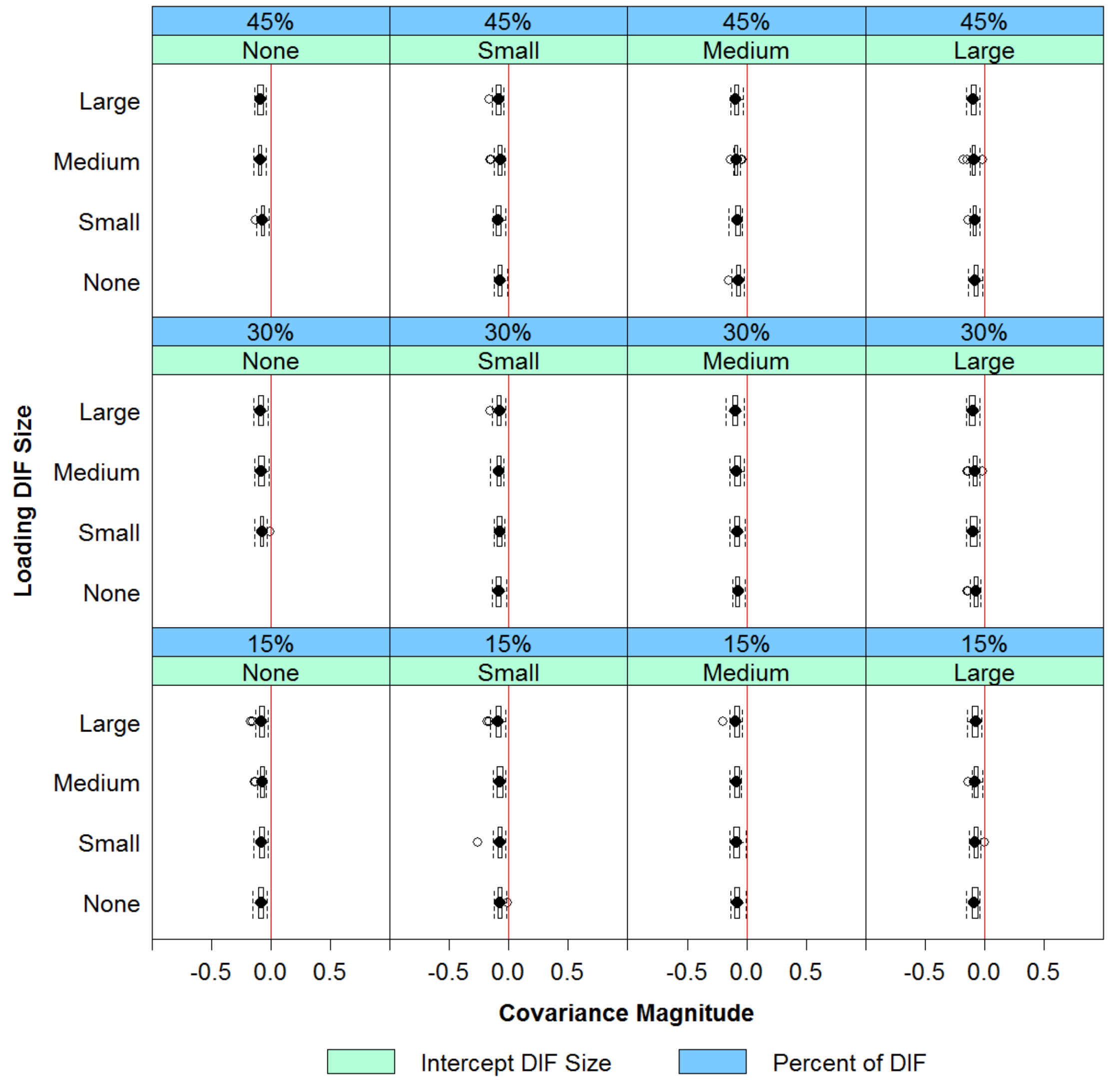

Figure 70. Intercept-Slope Covariance Estimates for the Constrained Second-Order Latent Growth Model.

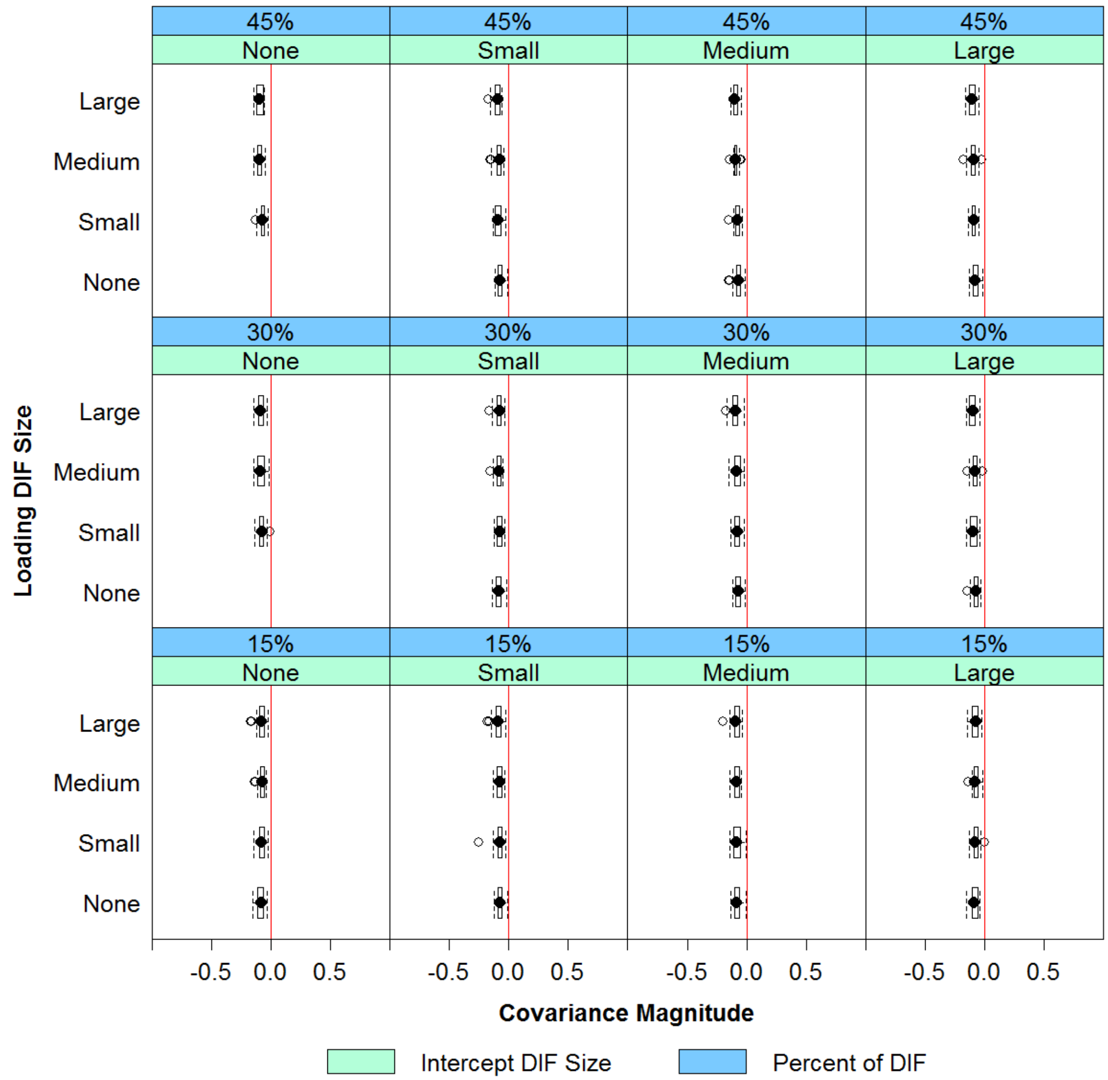

Figure 71. Intercept-Slope Covariance Estimates for the Free Second-Order Latent Growth Model.

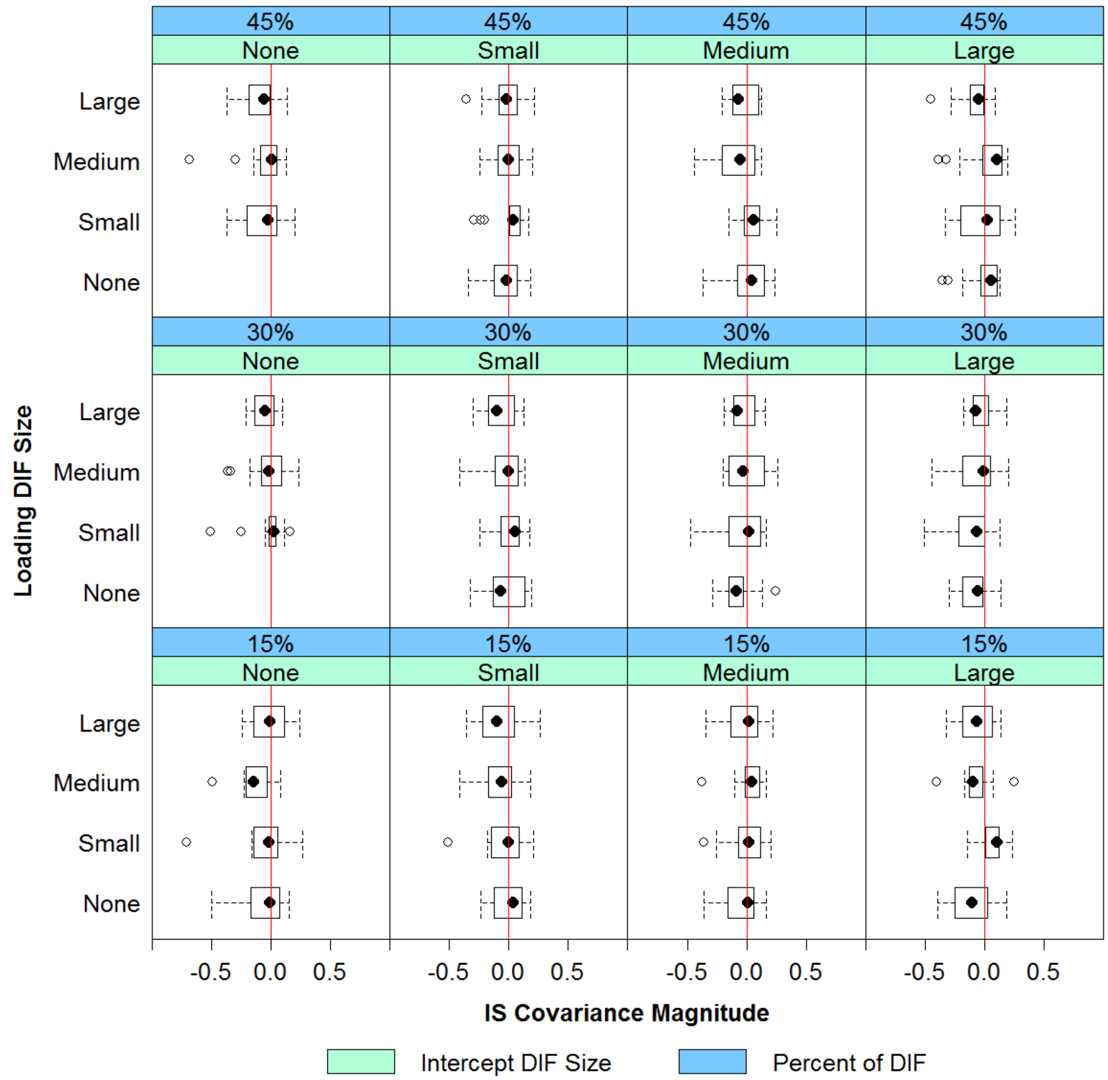

Figure 72. Intercept-Slope Covariance Estimates for the First-Order Non-Linear Latent Growth Model.

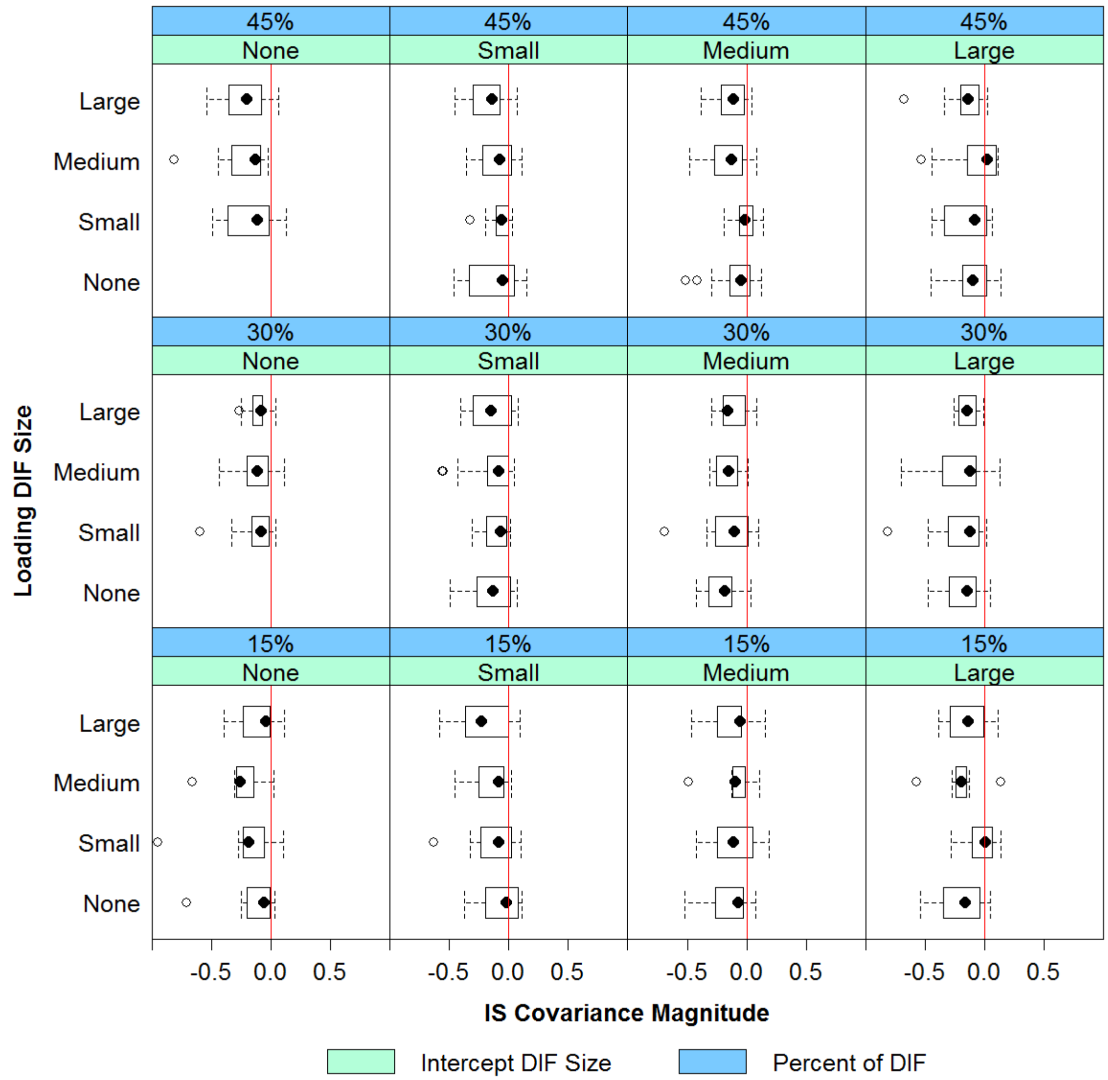

Figure 73. Intercept-Slope Covariance Estimates for the Constrained Second-Order Non-Linear Latent Growth Model.

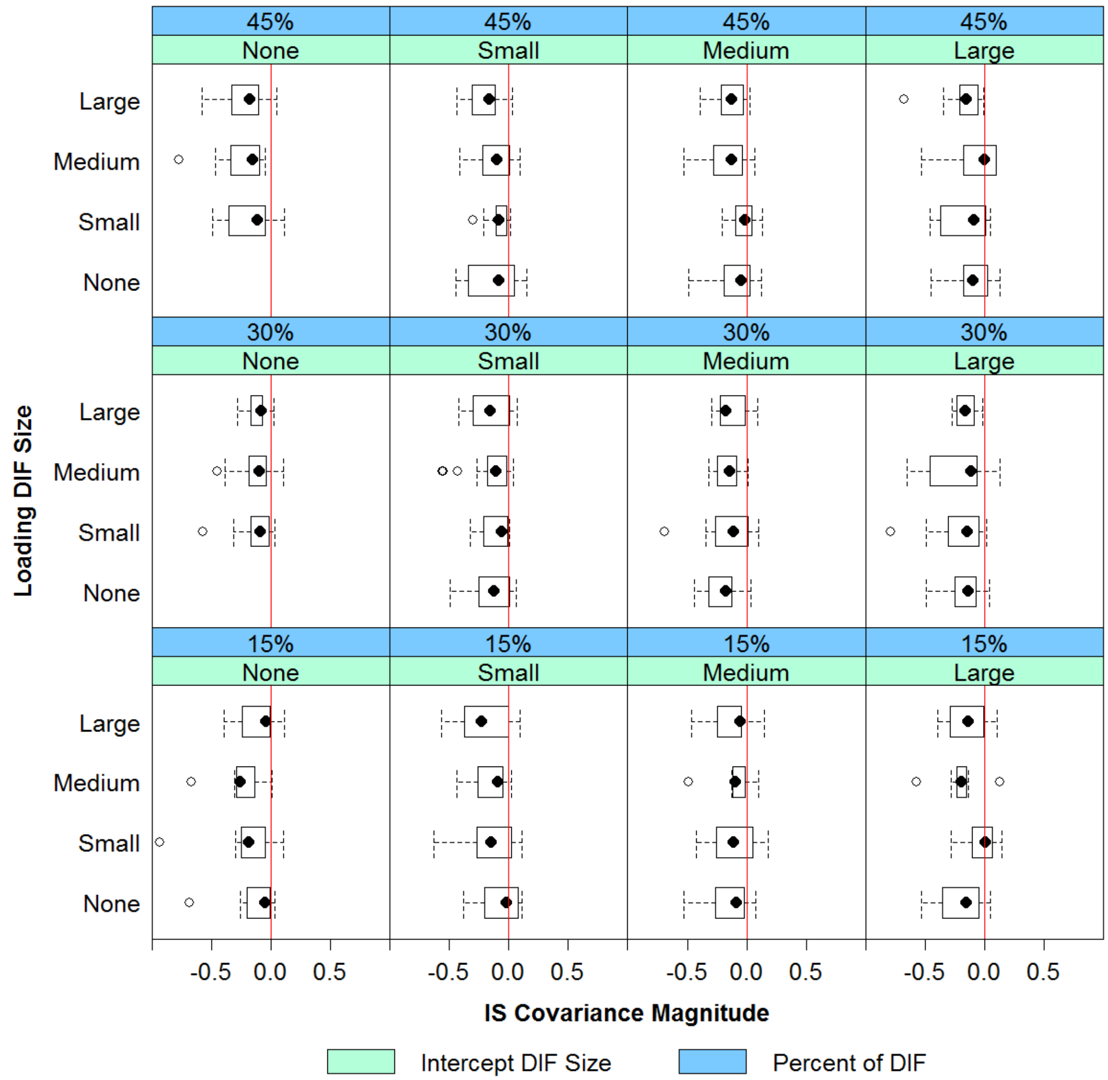

Figure 74. Intercept-Slope Covariance Estimates for the Free Second-Order Non-Linear Latent Growth Model.

Intercept-Quadratic Slope Covariance. The *p*-values associated with the estimate of the intercept-quadratic slope covariance are presented in Figures 75-77. A red line is drawn at 0.05 to represent the critical value with which the *p*-value is often compared to determine statistical significance. The mean and standard deviation values for the *p*-values of the intercept-quadratic slope covariance are included in Appendix M. For non-linear models, the *p*-values suggested that the intercept-quadratic slope estimate was almost always non-significant (2%, 0.4%, and 0.4% Type I error rate for the NLGM, 2NLGMC, and 2NLGMF, respectively). Thus, the non-linear models did an adequate job recovering the non-significant estimate for the intercept-quadratic slope covariance parameter.

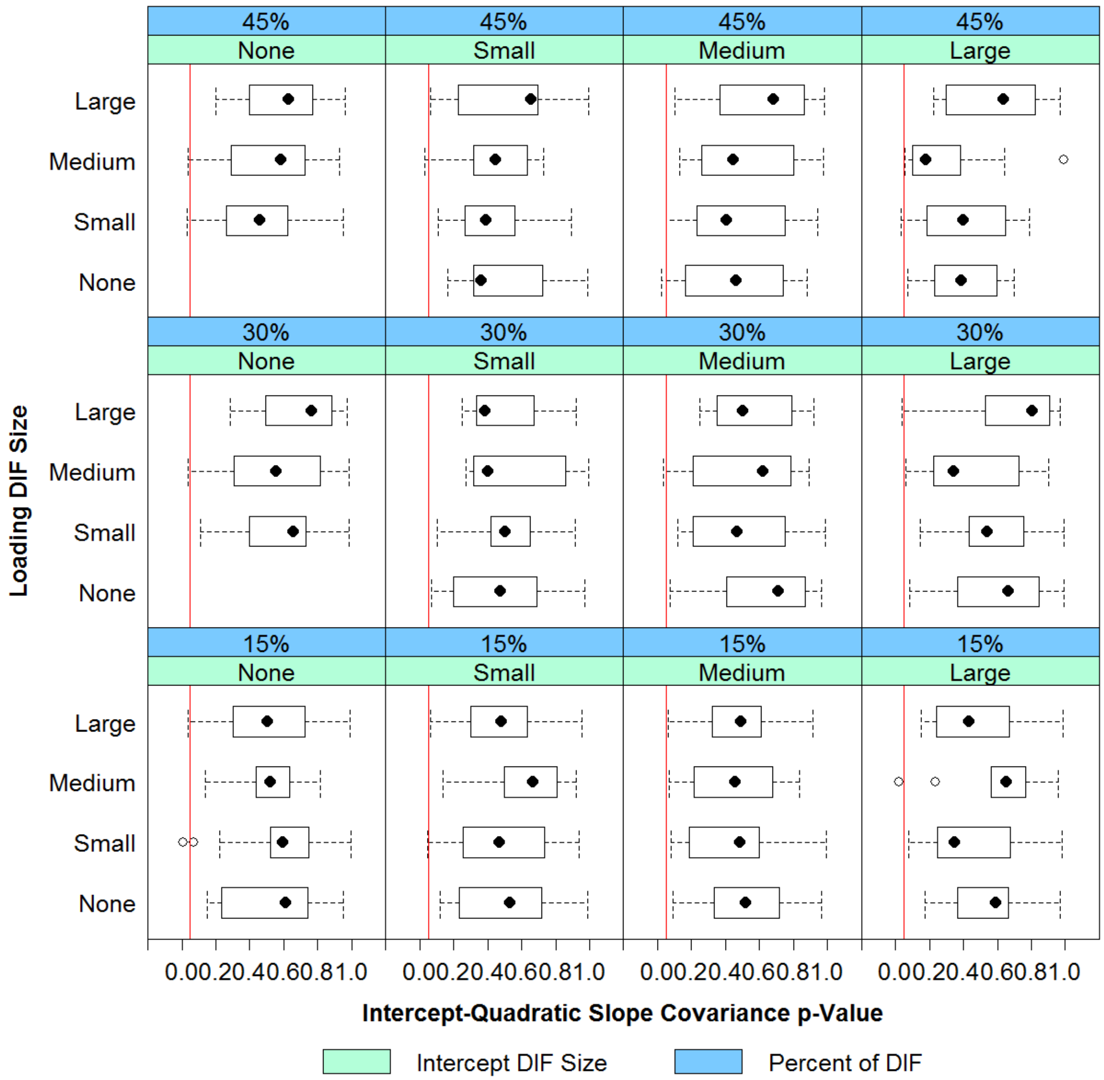

Figure 75. Intercept-Quadratic Slope Covariance *p*-Value for the First-Order Non-Linear Latent Growth Model.

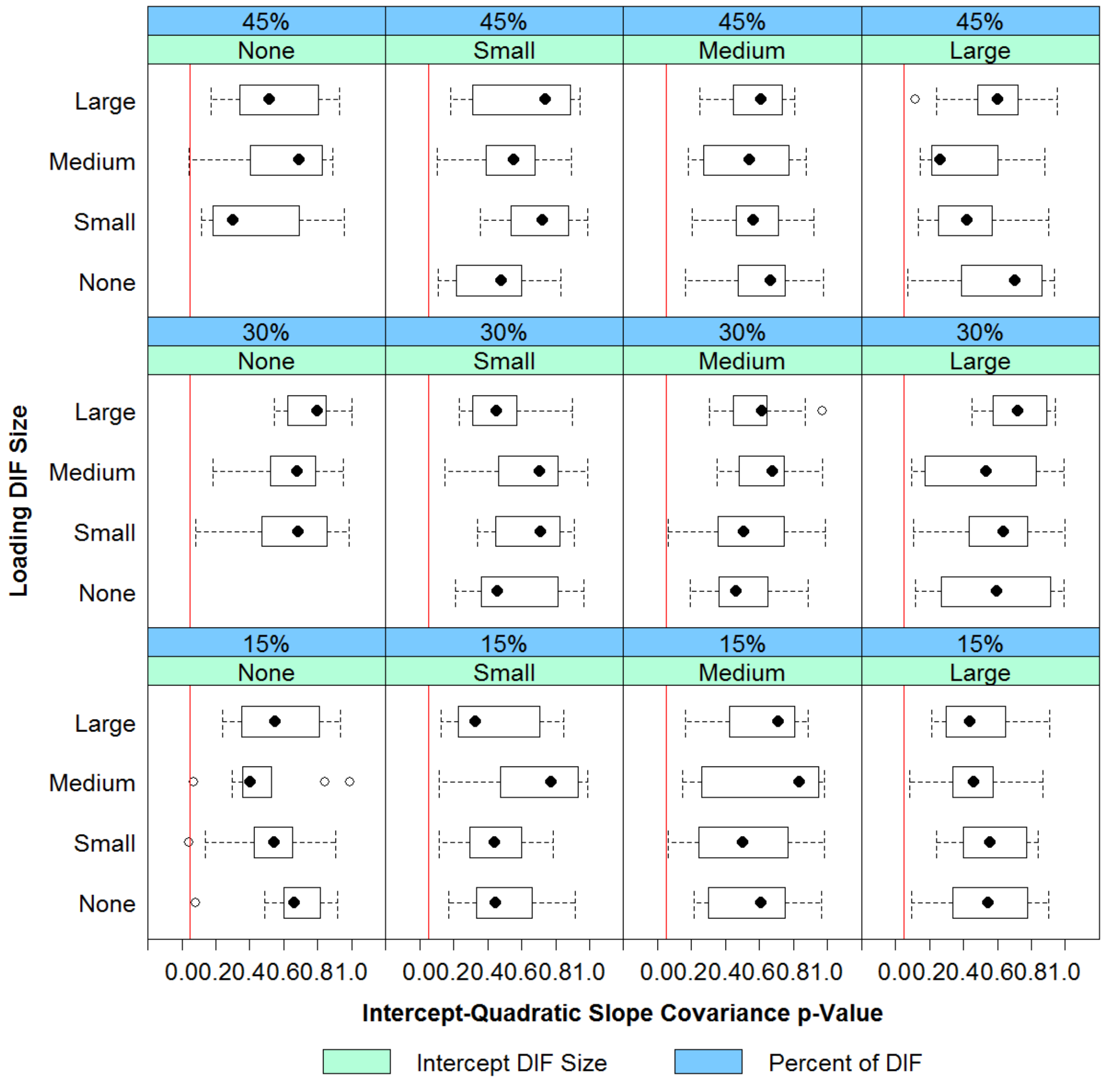

Figure 76. Intercept-Quadratic Slope Covariance *p*-Value for the Constrained Second-Order Non-Linear Latent Growth Model.

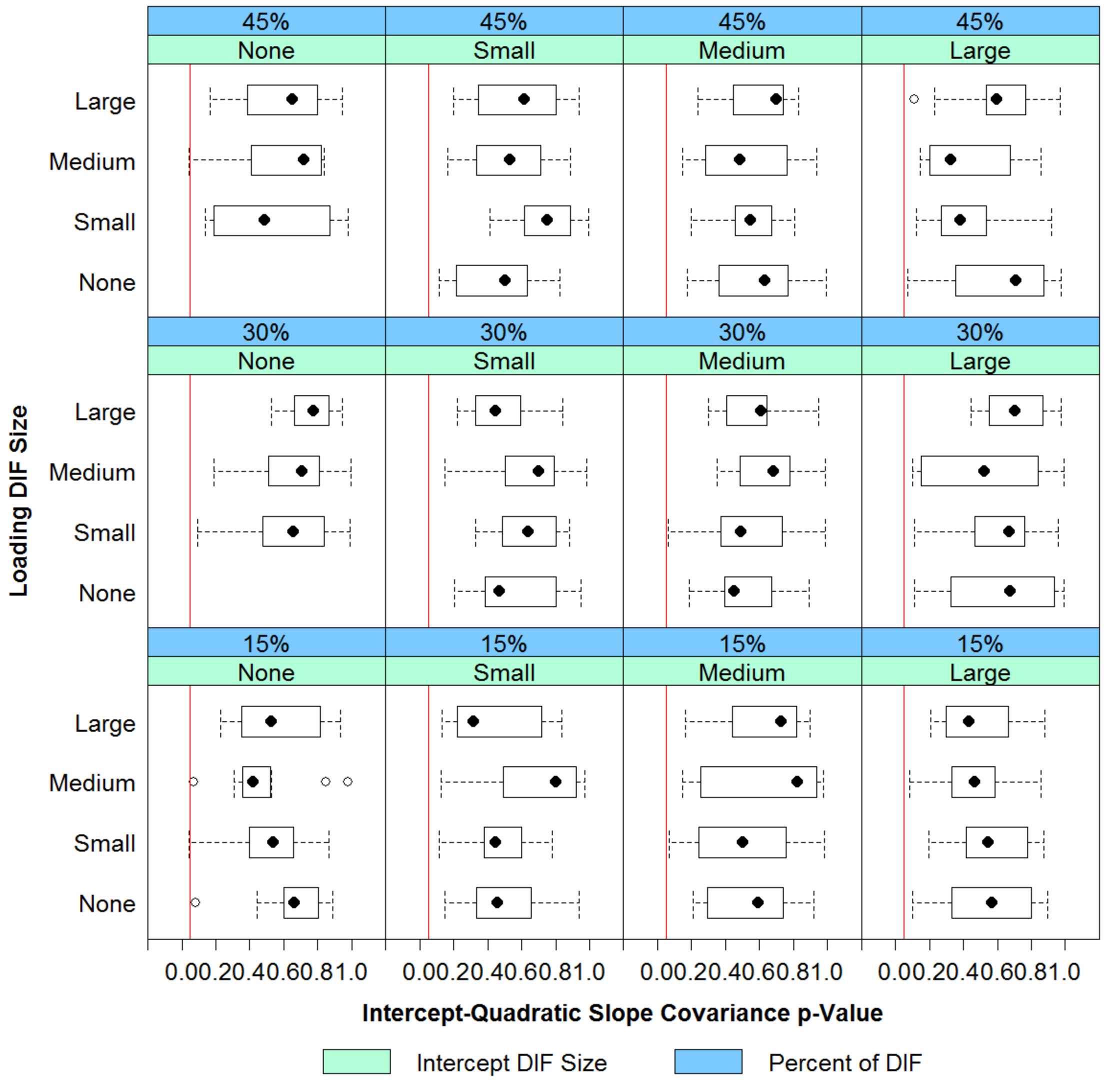

Figure 77. Intercept-Quadratic Slope Covariance *p*-Value for the Free Second-Order Non-Linear Latent Growth Model.

The estimated values for the covariance between the intercept and quadratic slope are presented in Figures 78-80. The mean and standard deviation values for the intercept-quadratic slope covariance estimates are included in Appendix N. As suggested by the non-significant *p*-values, the estimates for the intercept-quadratic slope covariance were near zero for all non-linear models across all conditions.

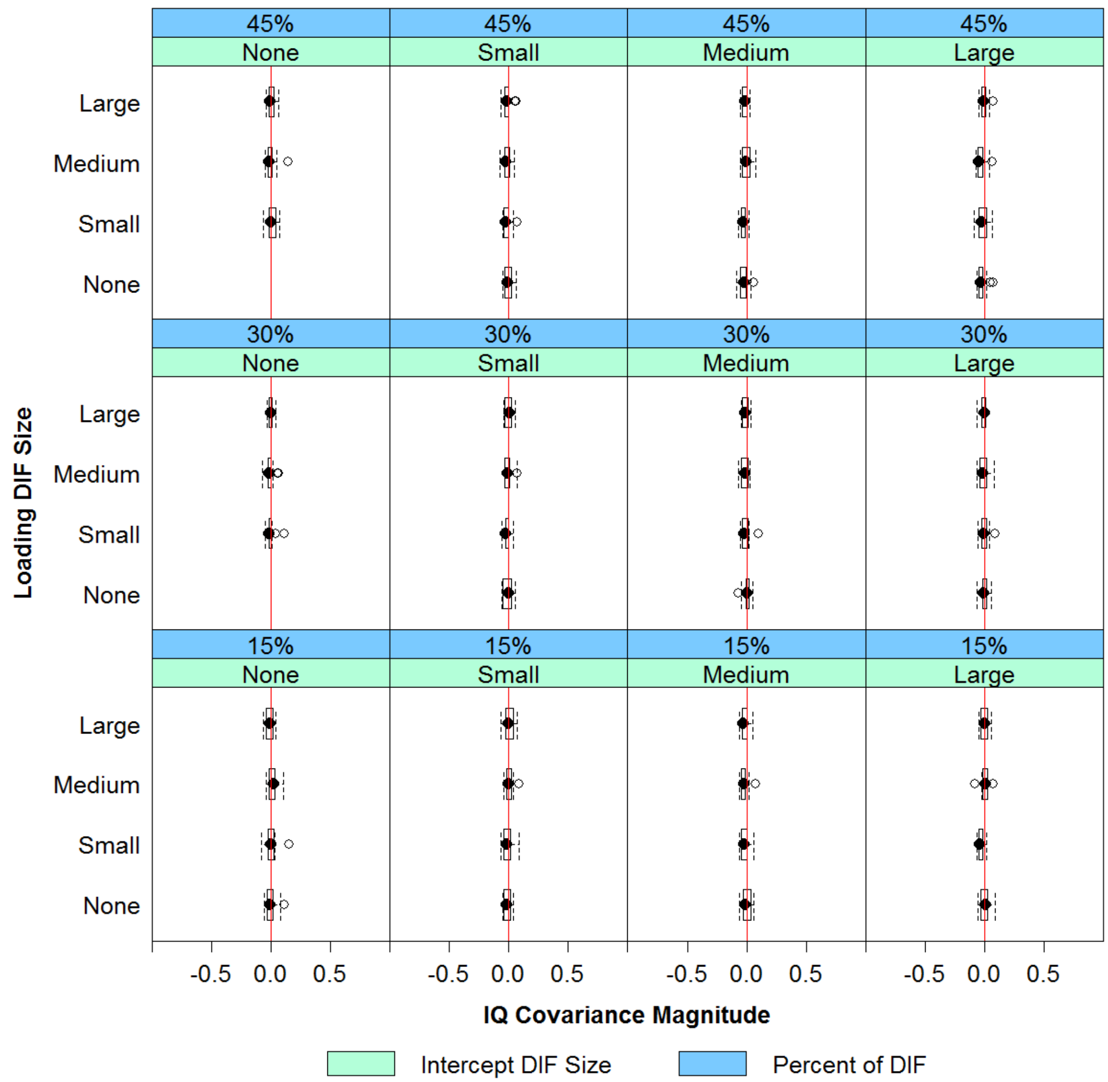

Figure 78. Intercept-Quadratic Slope Covariance for the First-Order Non-Linear Latent Growth Model.

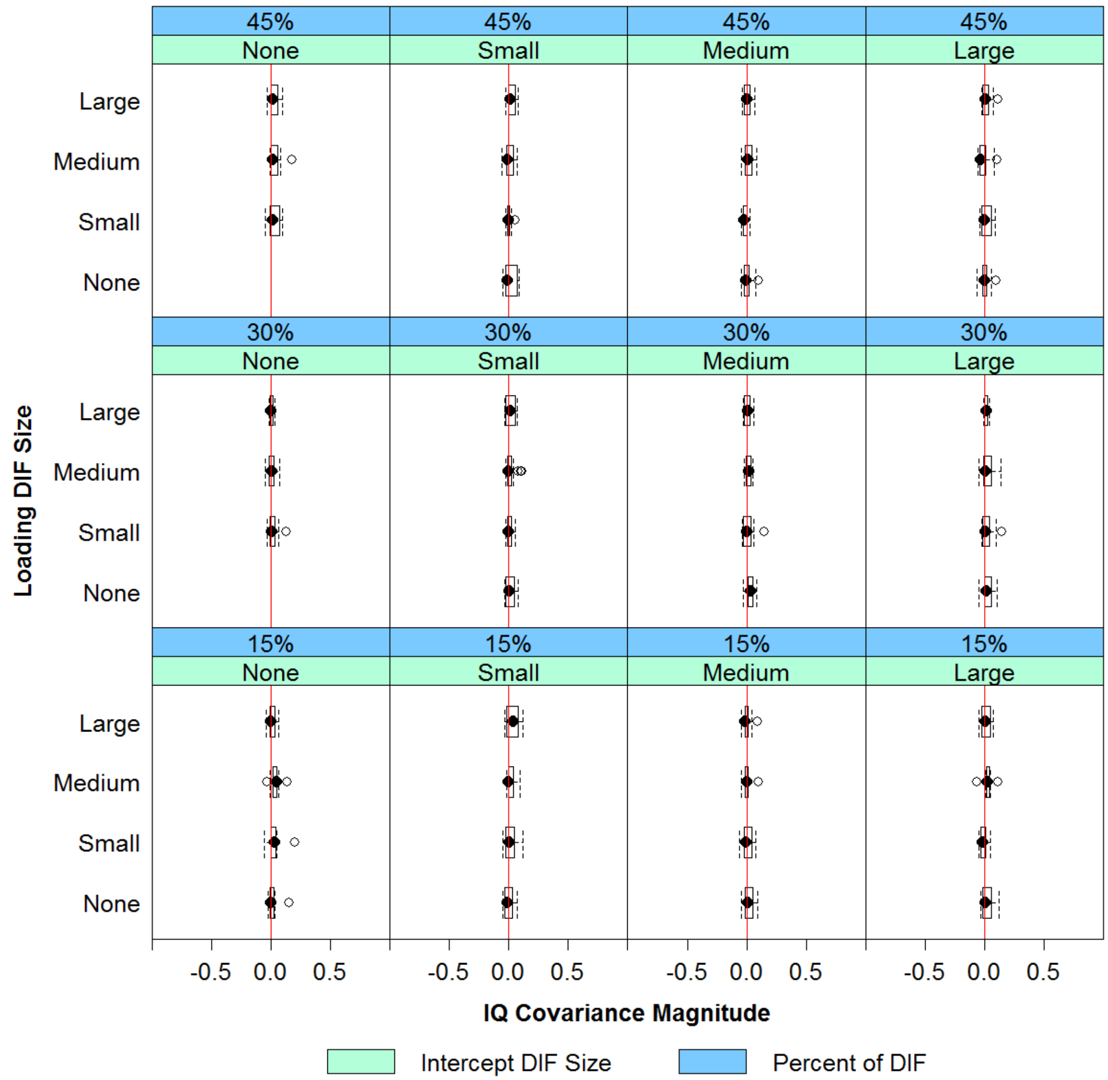

Figure 79. Intercept-Quadratic Slope Covariance for the Constrained Second-Order Non-Linear Latent Growth Model.

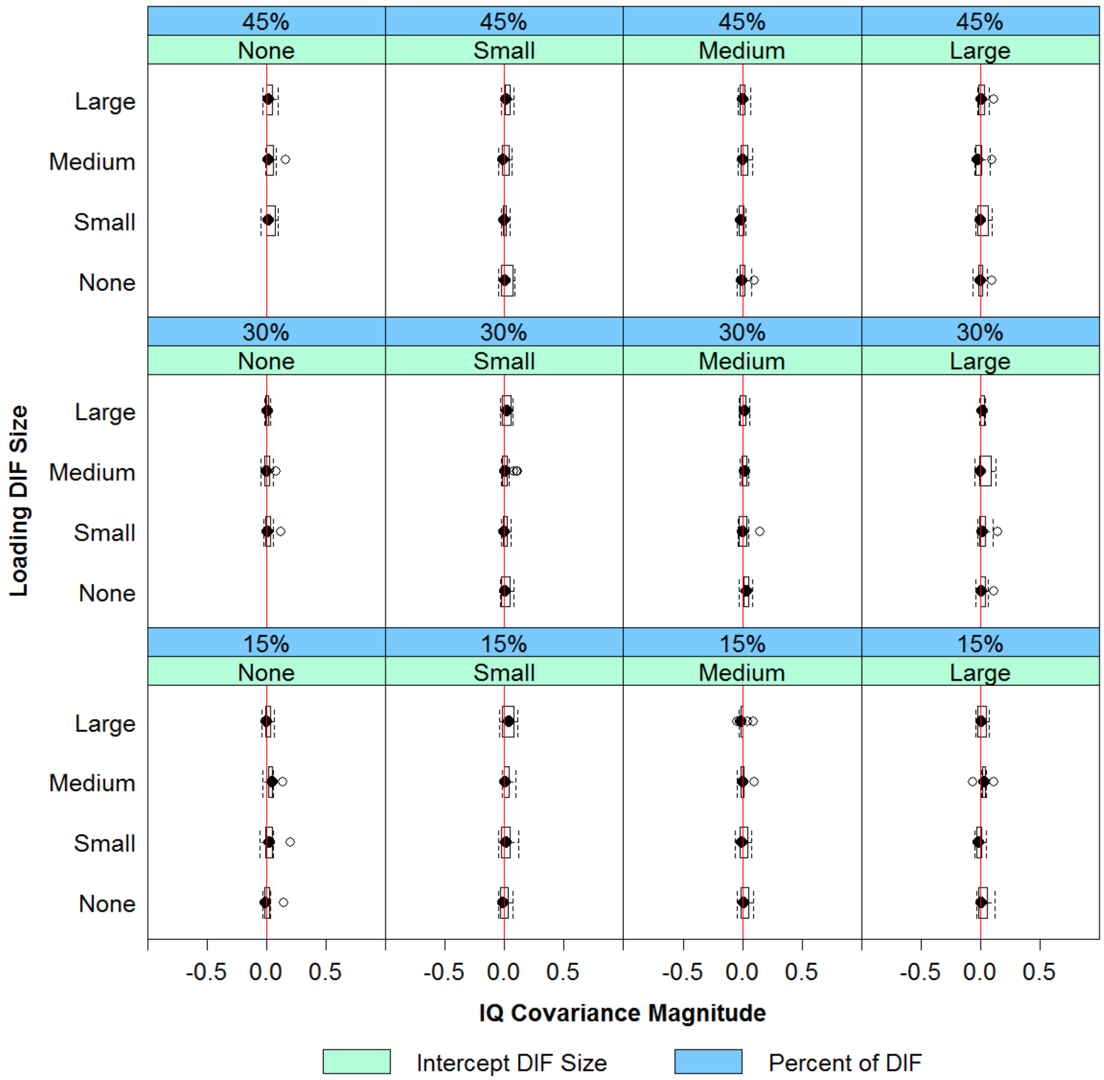

Figure 80. Intercept-Quadratic Slope Covariance for the Free Second-Order Non-Linear Latent Growth Model.

Slope-Quadratic Slope Covariance. The *p*-values associated with the estimate of the slope-quadratic slope covariance are presented in Figures 81-83. The mean and standard deviation values for the *p*-values for the slope-quadratic slope covariance are included in Appendix O. A red line is drawn at 0.05 to represent the critical value with which the *p*-value is often compared to determine statistical significance. As with the other covariance estimates, the *p*-values suggested that the non-linear models recovered the slope-quadratic slope parameter well. The *p*-values were almost all non-significant suggesting that the estimated value for the slope-quadratic slope covariance estimate was not significantly different than zero. The Type I error rate across all conditions for the NLGM, 2NLGMC, and 2NLGMF was 10%, 6%, and 6%, respectively.

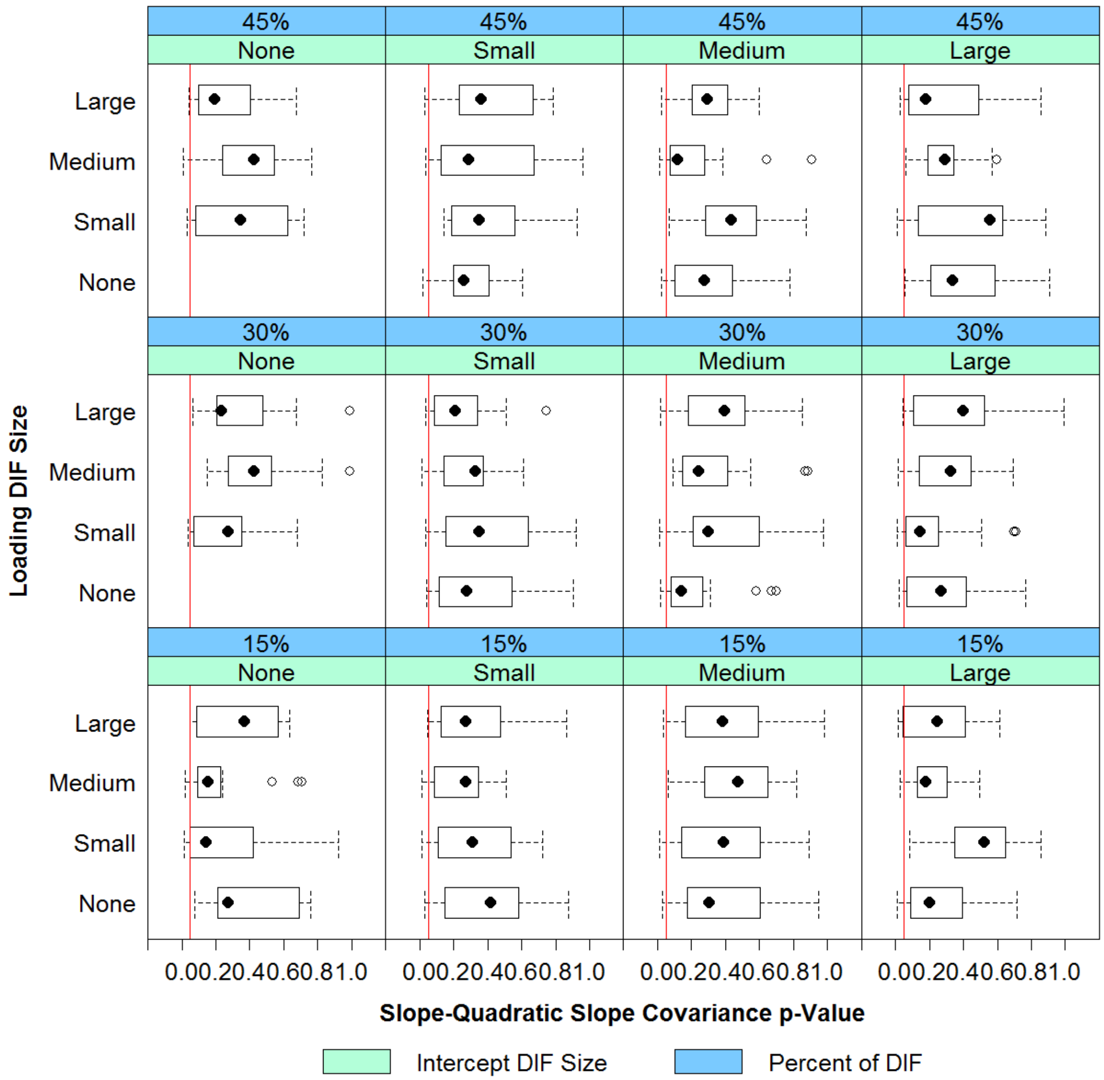

Figure 81. Slope-Quadratic Slope Covariance *p*-Value for the First-Order Non-Linear Latent Growth Model.

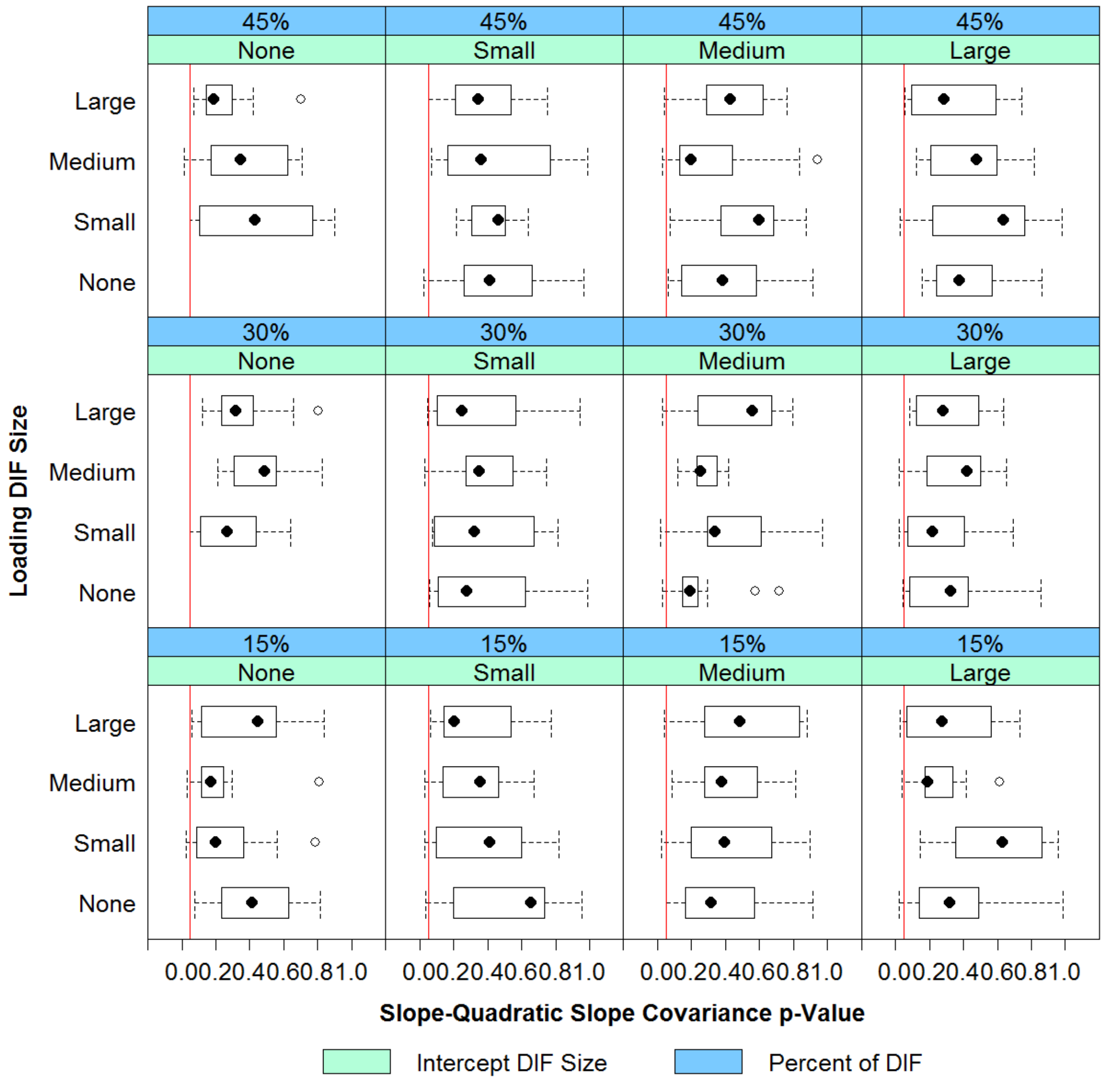

Figure 82. Slope-Quadratic Slope Covariance *p*-Value for the Constrained Second-Order Non-Linear Latent Growth Model.

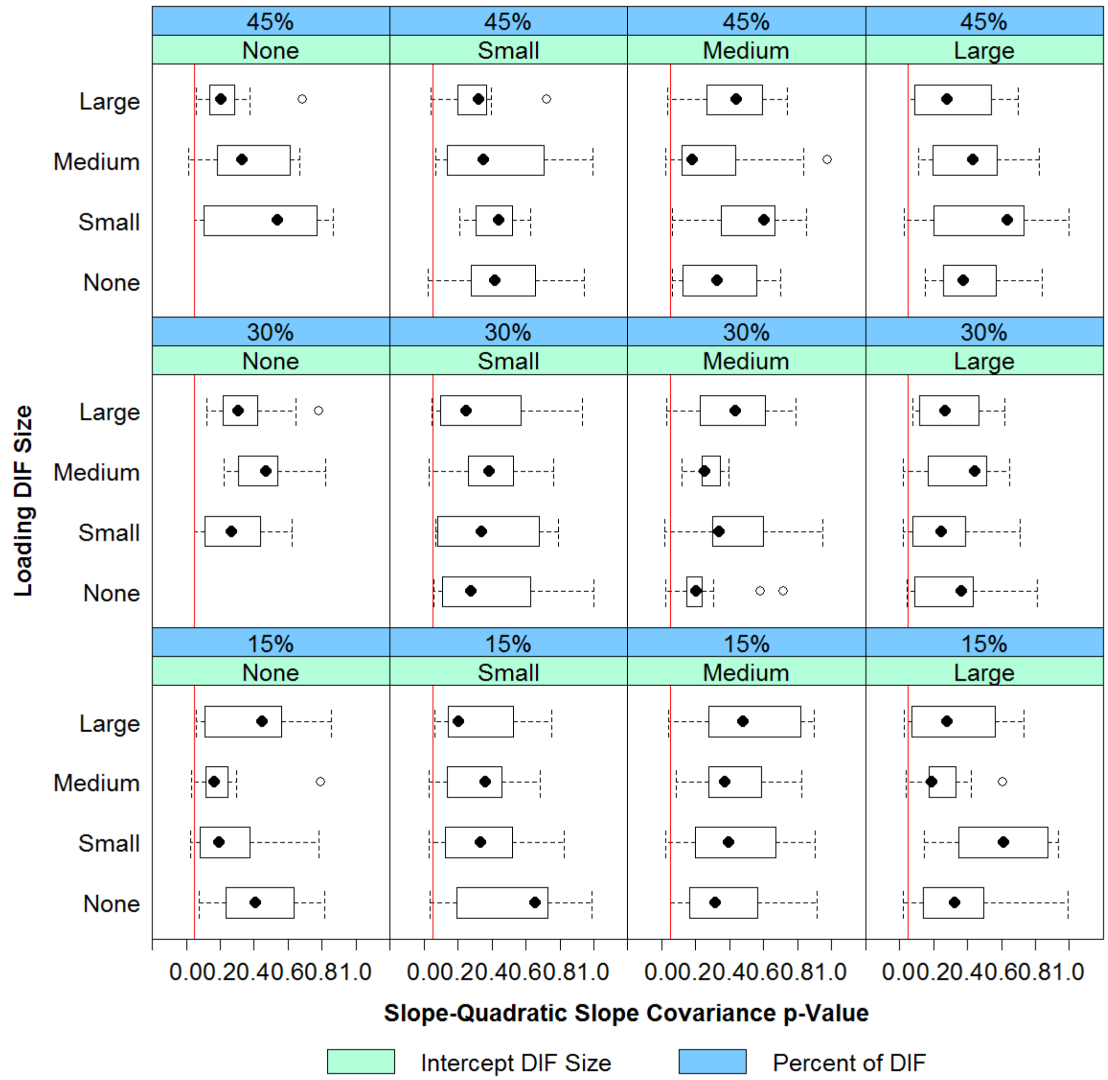

Figure 83. Slope-Quadratic Slope Covariance *p*-Value for the Free Second-Order Non-Linear Latent Growth Model.

The estimated parameters for the covariance between the linear slope and the quadratic slope are presented in Figures 84-86. The mean and standard deviation values for the slope-quadratic slope covariance estimates are included in Appendix P. As with the other covariance parameters in the non-linear models, the estimated values were near zero. The small magnitude was unsurprising given the non-significant *p*-values associated with the slope-quadratic slope covariance estimates.

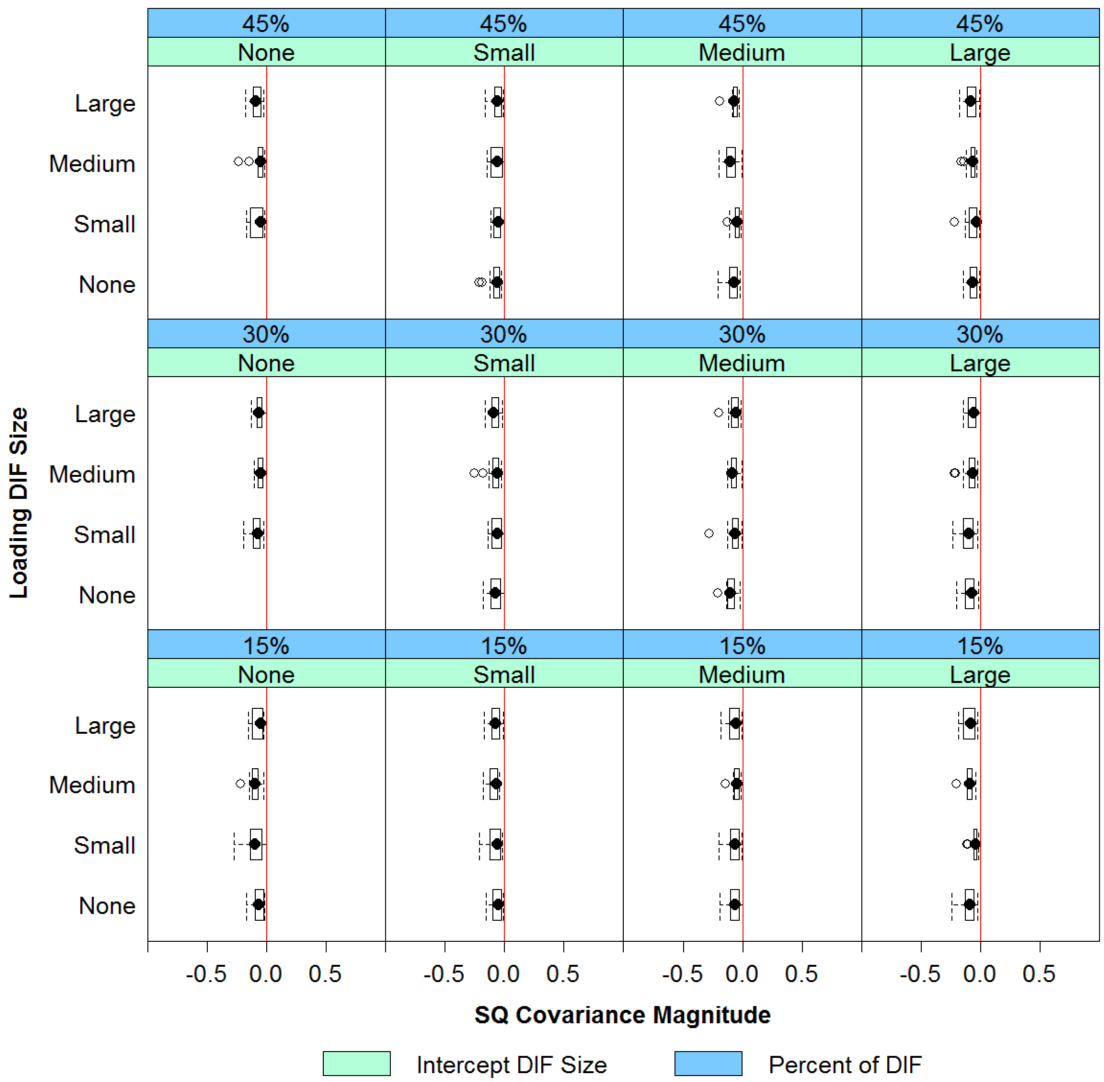

Figure 84. Slope-Quadratic Slope Covariance for the First-Order Non-Linear Latent Growth Model.

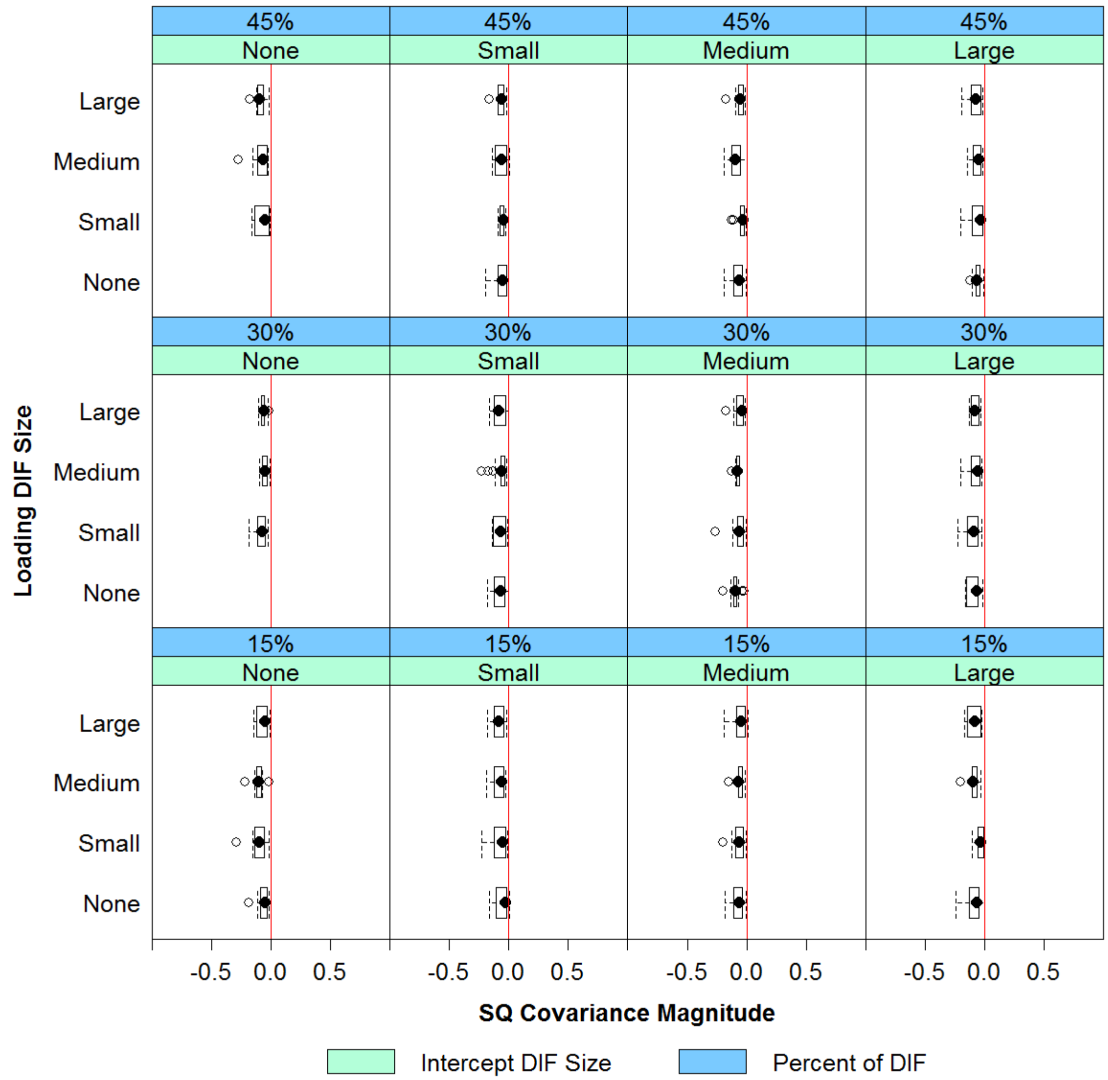

Figure 85. Slope-Quadratic Slope Covariance for the Constrained Second-Order Non-Linear Latent Growth Model.

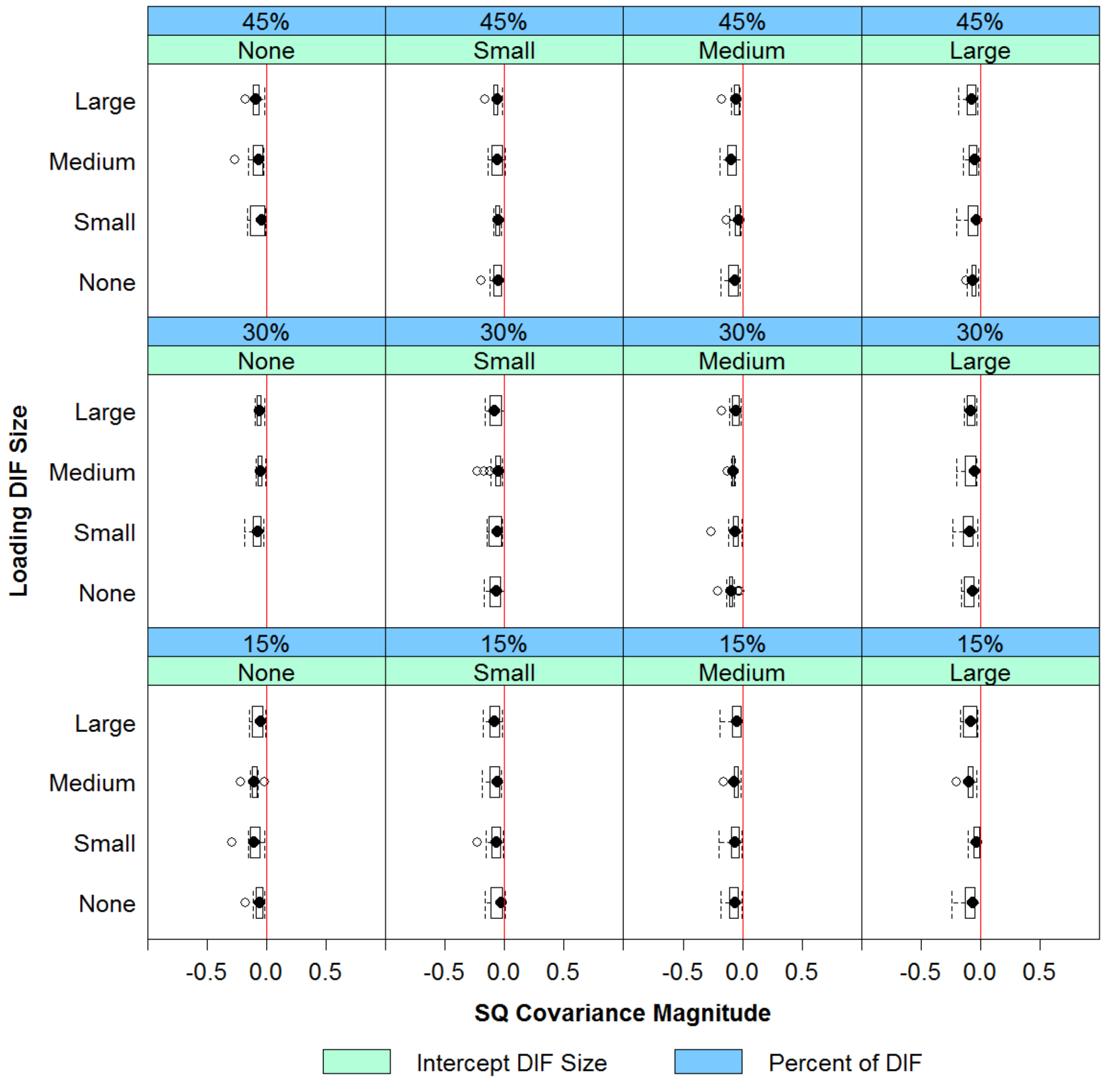

Figure 86. Slope-Quadratic Slope Covariance for the Free Second-Order Non-Linear Latent Growth Model.

Quadratic Slope. The *p*-values associated with the estimate of the quadratic slope estimate are presented in Figures 87-89. A red line is drawn at 0.05 to represent the critical value with which the *p*-value is often compared to determine statistical significance. The mean and standard deviation values for the *p*-value for the quadratic slope are included in Appendix Q. For all non-linear models (NLGM, 2NLGMC, and 2NLGMF), the *p*-values suggested that the quadratic slope was often statistically significant. For the first-order non-linear model, the median *p*-values in the conditions with a low proportion of DIF and/or small intercept DIF were frequently non-significant. The variability in *p*-values for all models was much smaller in the conditions with 45% of items containing medium or large intercept DIF. For these conditions, the quadratic slope was almost always statistically significant. The Type I error rate was 71% in the condition with 45% of items with large intercept DIF. For the second-order models, the quadratic slope parameter was often estimated as statistically significant. For the conditions with 45% of items with large intercept DIF, the Type I error rate was 89% for the 2NLGMC and 100% for the 2NLGMF.

Figure 87. Quadratic Slope Estimate *p*-Value for the First-Order Non-Linear Latent Growth Model.

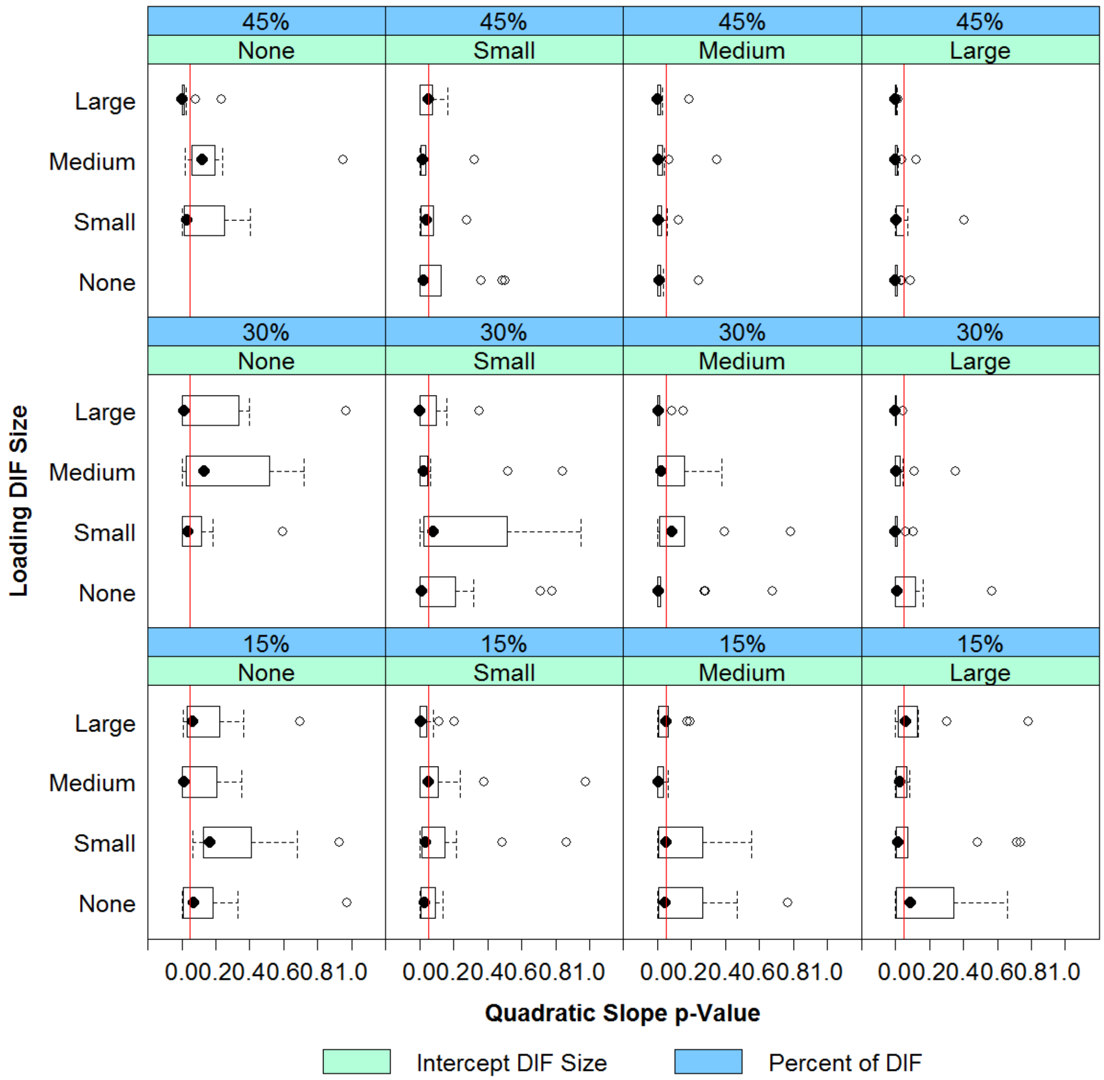

Figure 88. Quadratic Slope Estimate *p*-Value for the Constrained Second-Order Non-Linear Latent Growth Model.

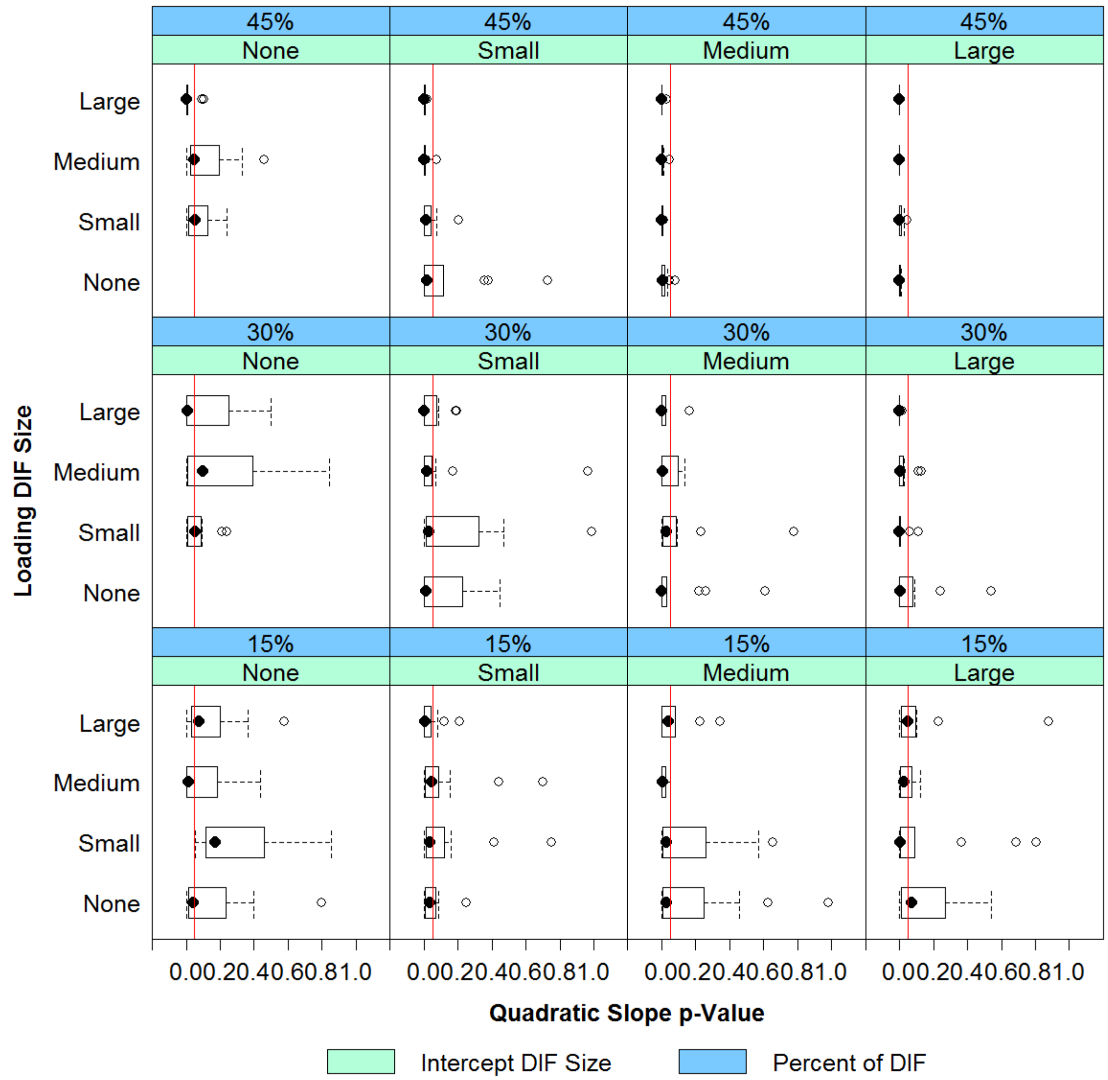

Figure 89. Quadratic Slope Estimate *p*-Value for the Free Second-Order Non-Linear Latent Growth Model.

The estimated quadratic slope parameters for the non-linear models are presented in Figures 90-92. The mean and standard deviation values for the quadratic slope estimates are included in Appendix R. For all non-linear models, the estimated value of the quadratic slope parameter was essentially zero across all conditions. Whereas the *p*-

values suggested that the quadratic slope was statistically significant in the conditions with a large percent of large intercept DIF, the estimated values are not practically significant.

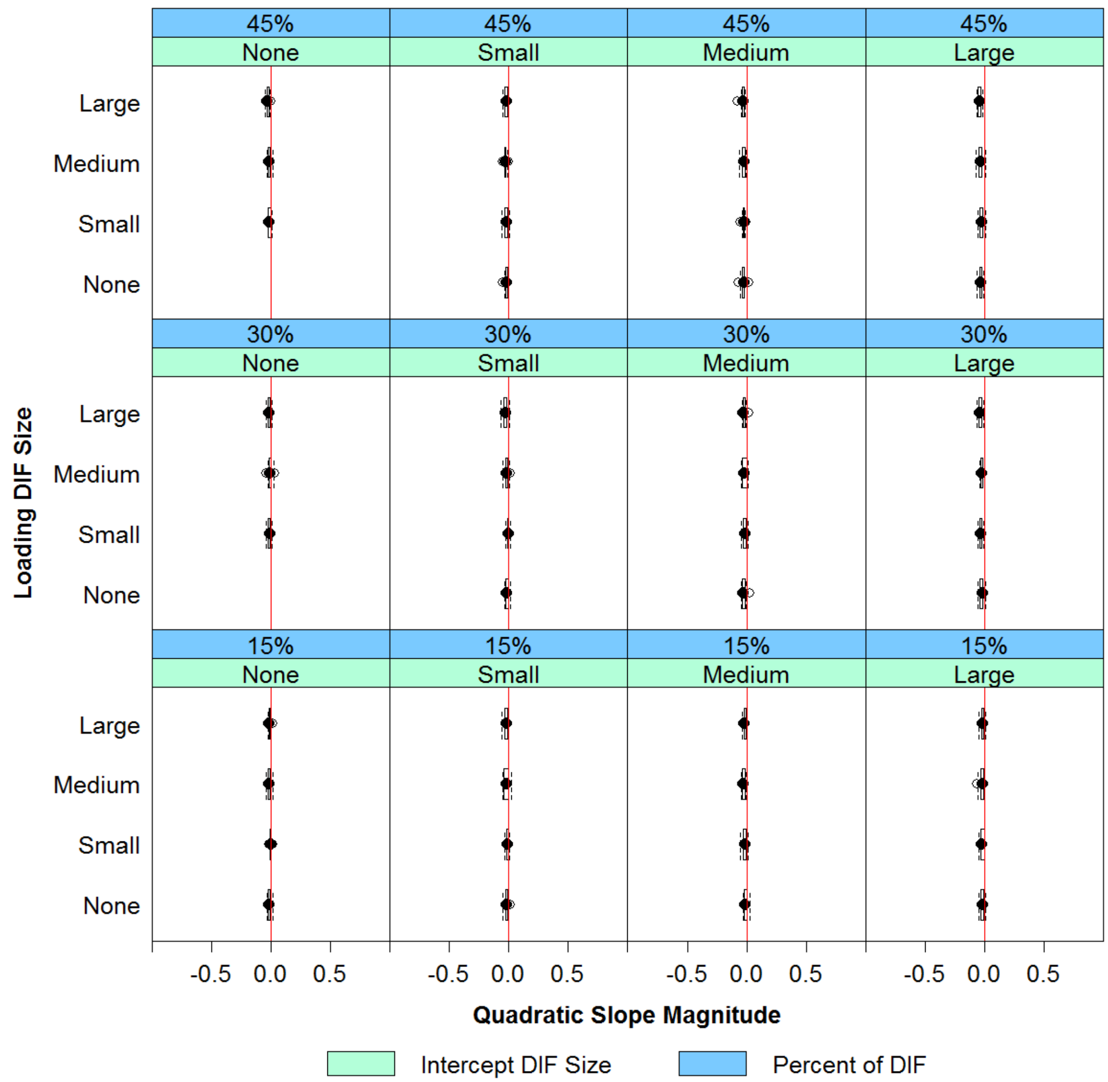

Figure 90. Quadratic Slope Estimate for the Constrained Second-Order Non-Linear Latent Growth Model.

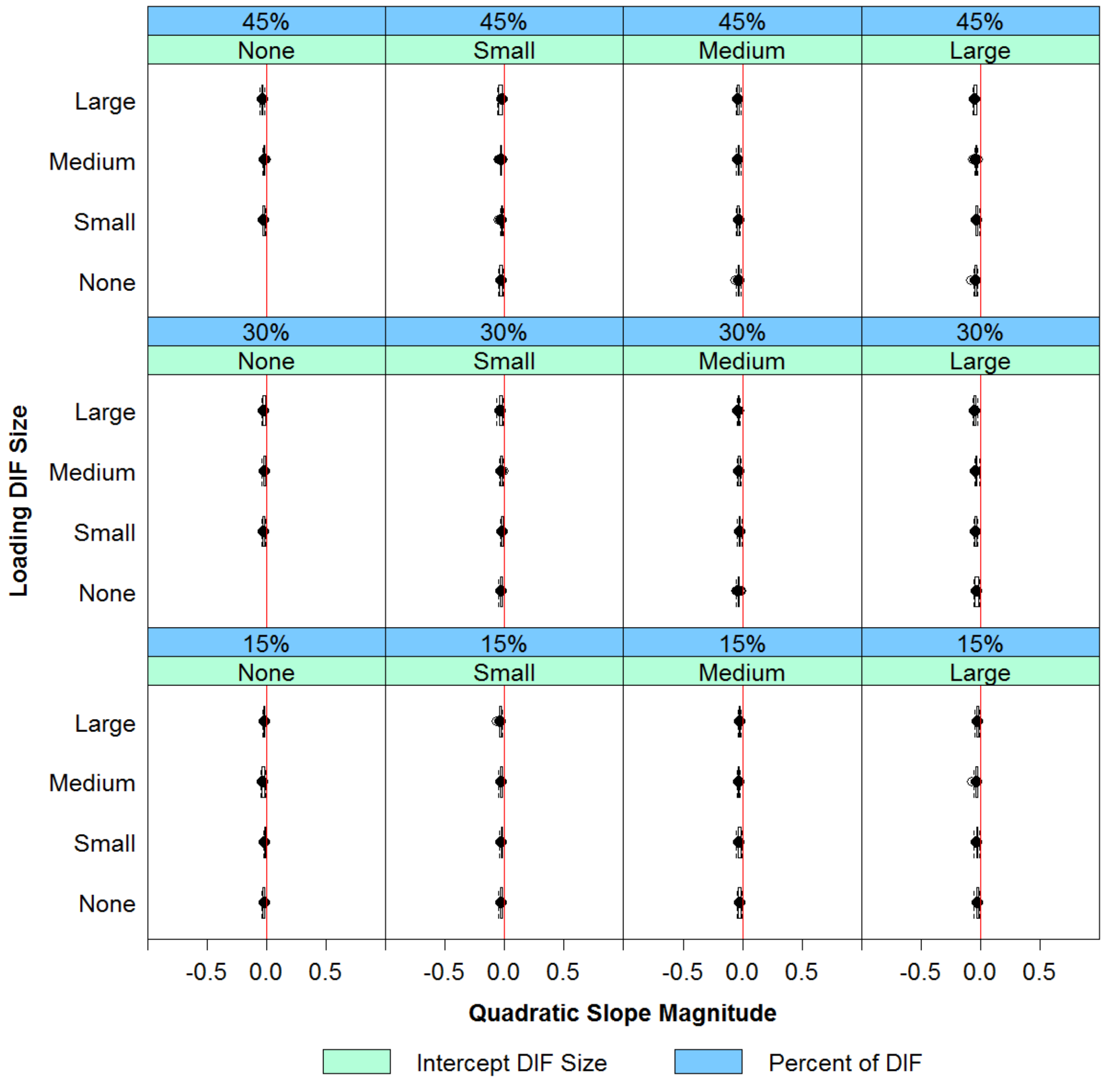

Figure 91. Quadratic Slope Estimate for the Constrained Second-Order Non-Linear Latent Growth Model.

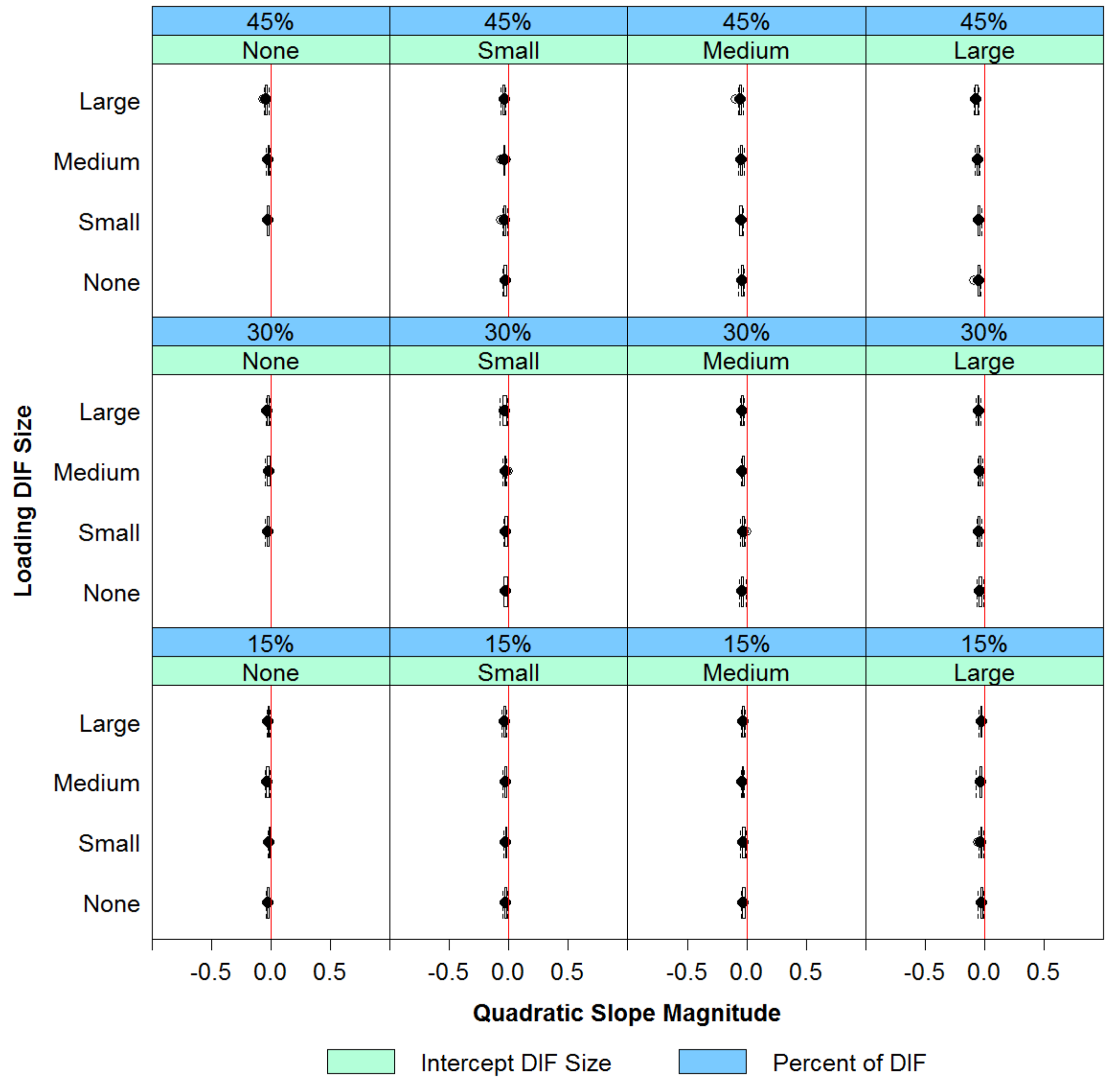

Figure 92. Quadratic Slope Estimate for the Free Second-Order Non-Linear Latent Growth Model.

Quadratic Slope Variance. The *p*-values associated with the estimate of the quadratic slope variance estimate are presented in Figures 93-95. A red line is drawn at 0.05 to represent the critical value with which the *p*-value is often compared to determine statistical significance. The mean and standard deviation values for the *p*-values for the quadratic slope variance are included in Appendix S. Across all conditions the *p*-values for the non-linear models (NLGM, 2NLGMC, and 2NLGMF) suggested that the quadratic slope variance was typically estimated to be non-significant. The variability in *p*-values was relatively large, but almost all *p*-values for all conditions were above the 0.05 critical value. Across all conditions, the Type I error rate for the NLGM, 2NLGMC, and 2NLGMF was 10%, 5%, and 5%, respectively.

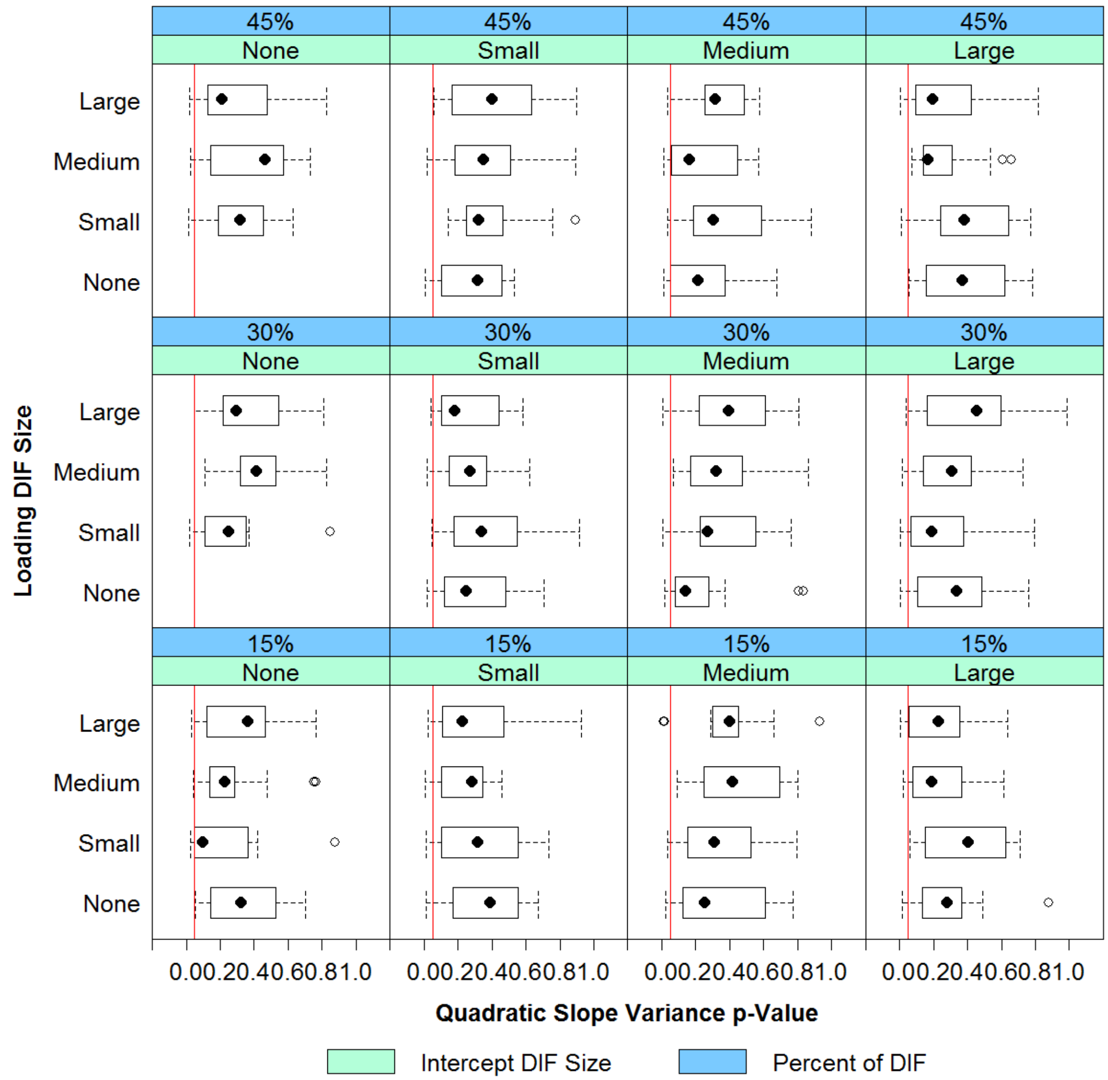

Figure 93. Quadratic Slope Variance Estimate *p*-Value for the First-Order Non-Linear Latent Growth Model.

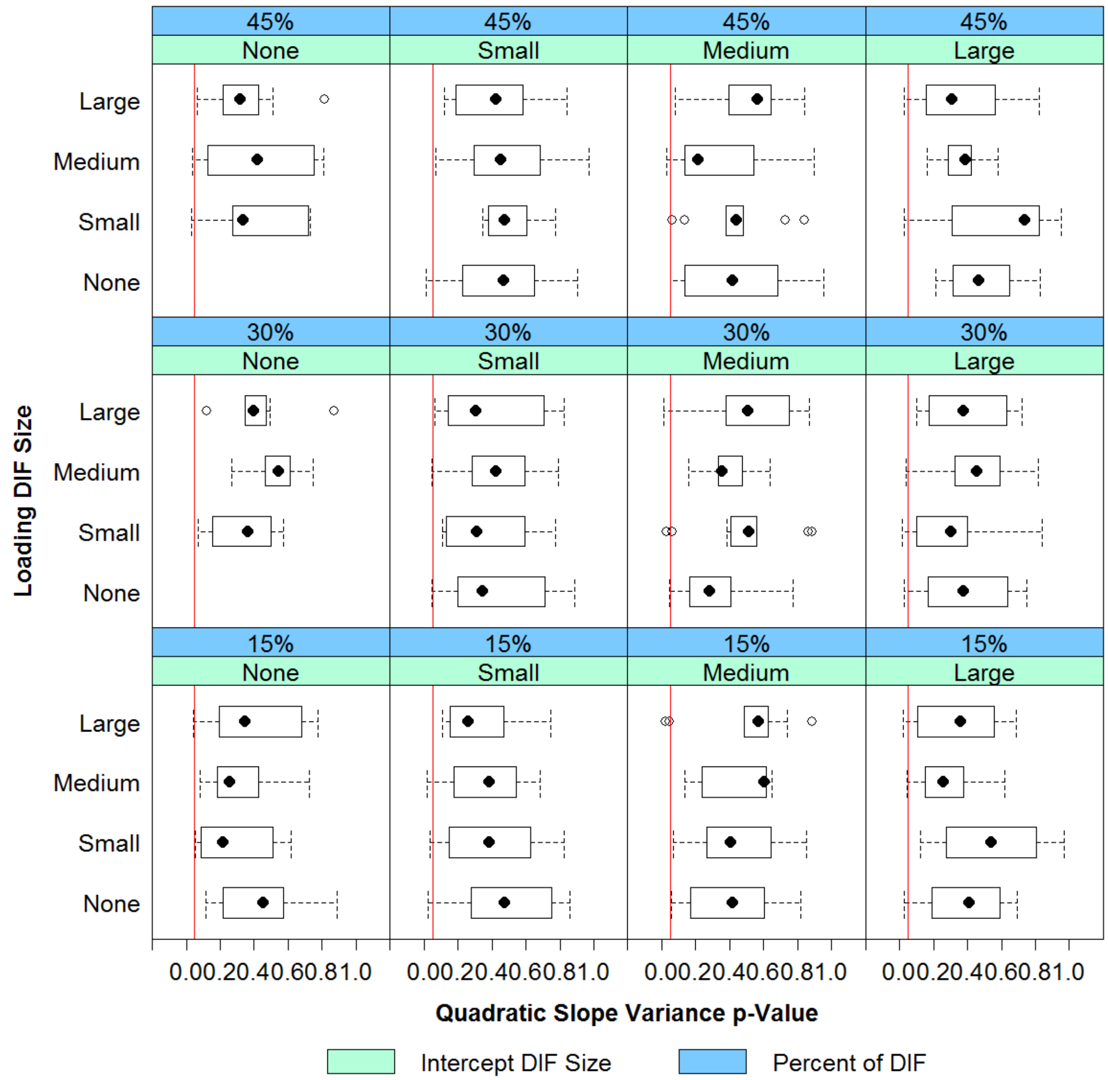

Figure 94. Quadratic Slope Variance Estimate *p*-Value for the Constrained Second-Order Non-Linear Latent Growth Model.

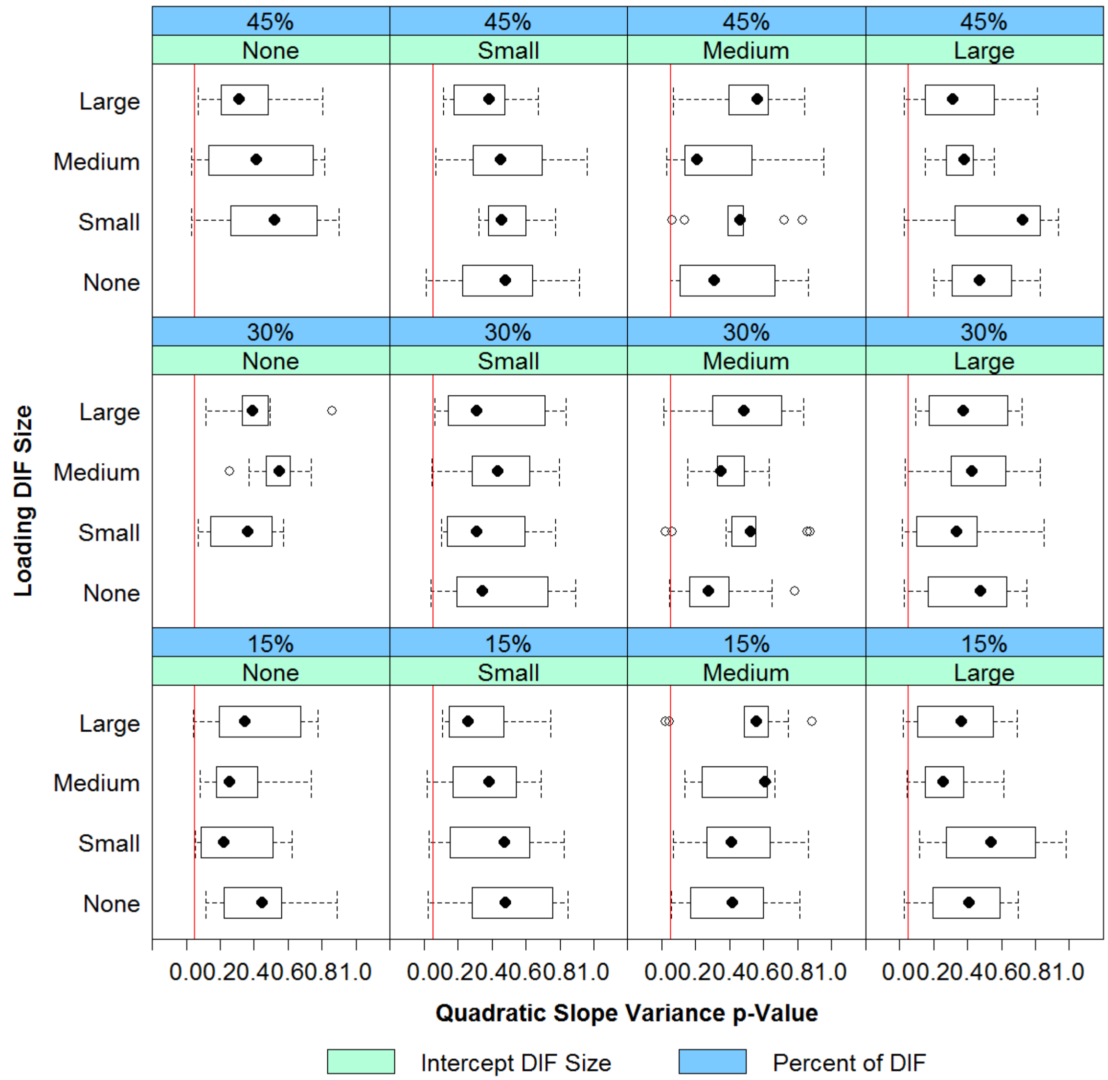

Figure 95. Quadratic Slope Variance Estimate *p*-Value for the Free Second-Order Non-Linear Latent Growth Model.

The estimates for the quadratic slope variance parameter in the non-linear models are presented in Figures 96-98. The mean and standard deviation values for the quadratic slope variance estimates are included in Appendix T. The estimates for all non-linear models were essentially zero in all conditions. This result was not surprising given that the *p*-values suggested that the estimates were not significant.

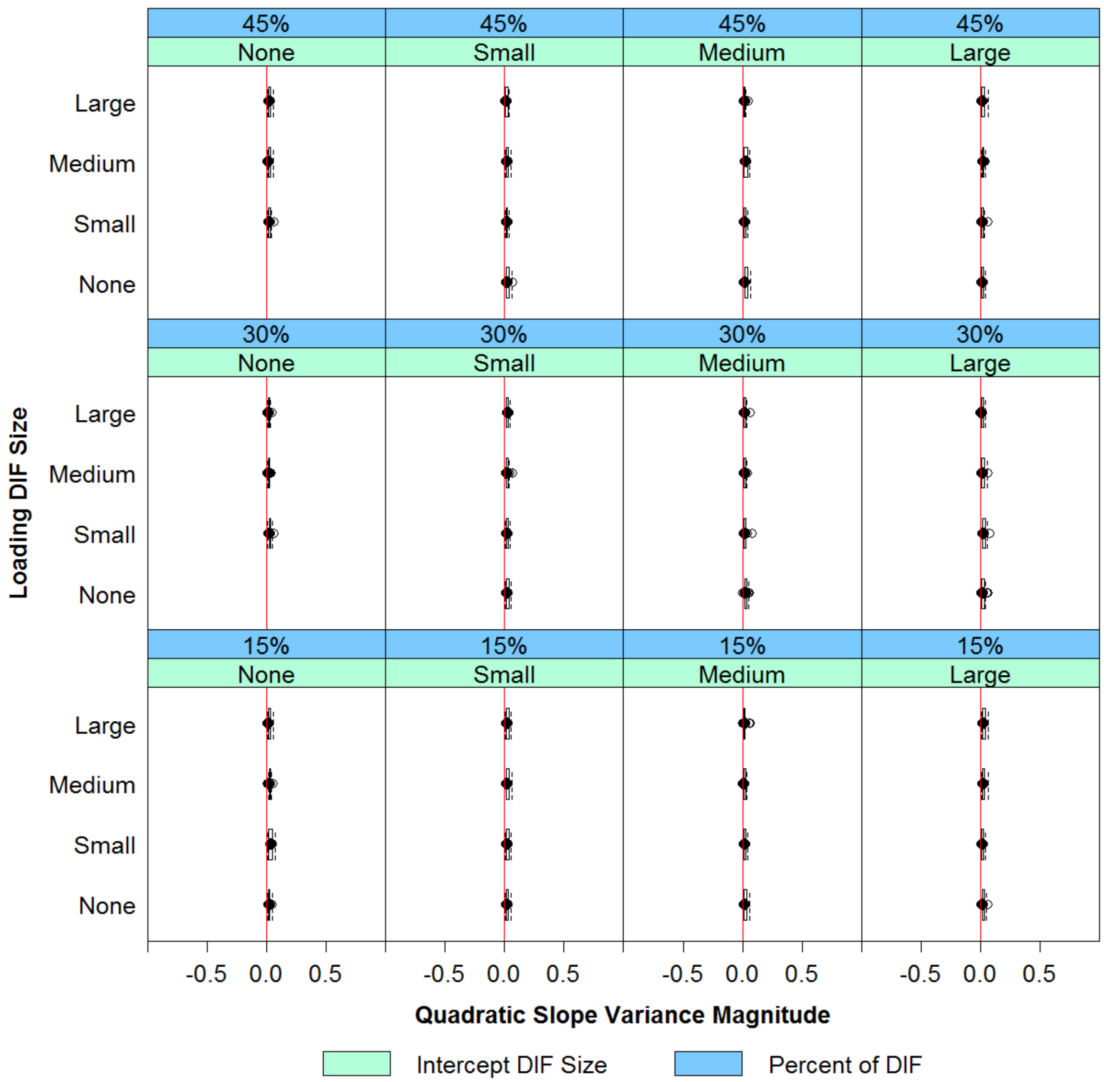

Figure 96. Quadratic Slope Variance Estimate for the First-Order Non-Linear Latent Growth Model.

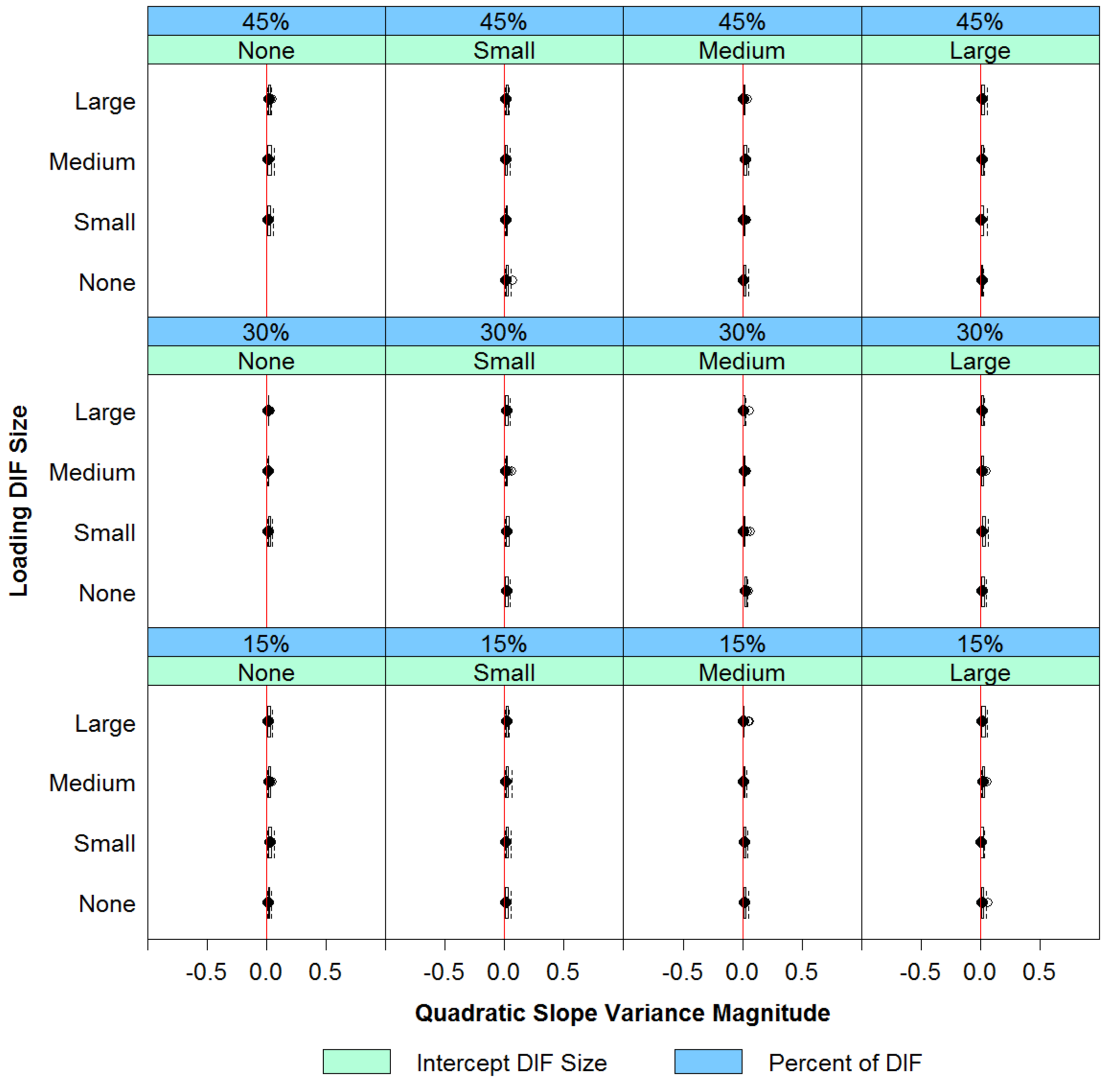

Figure 97. Quadratic Slope Variance Estimate for the Constrained Second-Order Non-Linear Latent Growth Model.

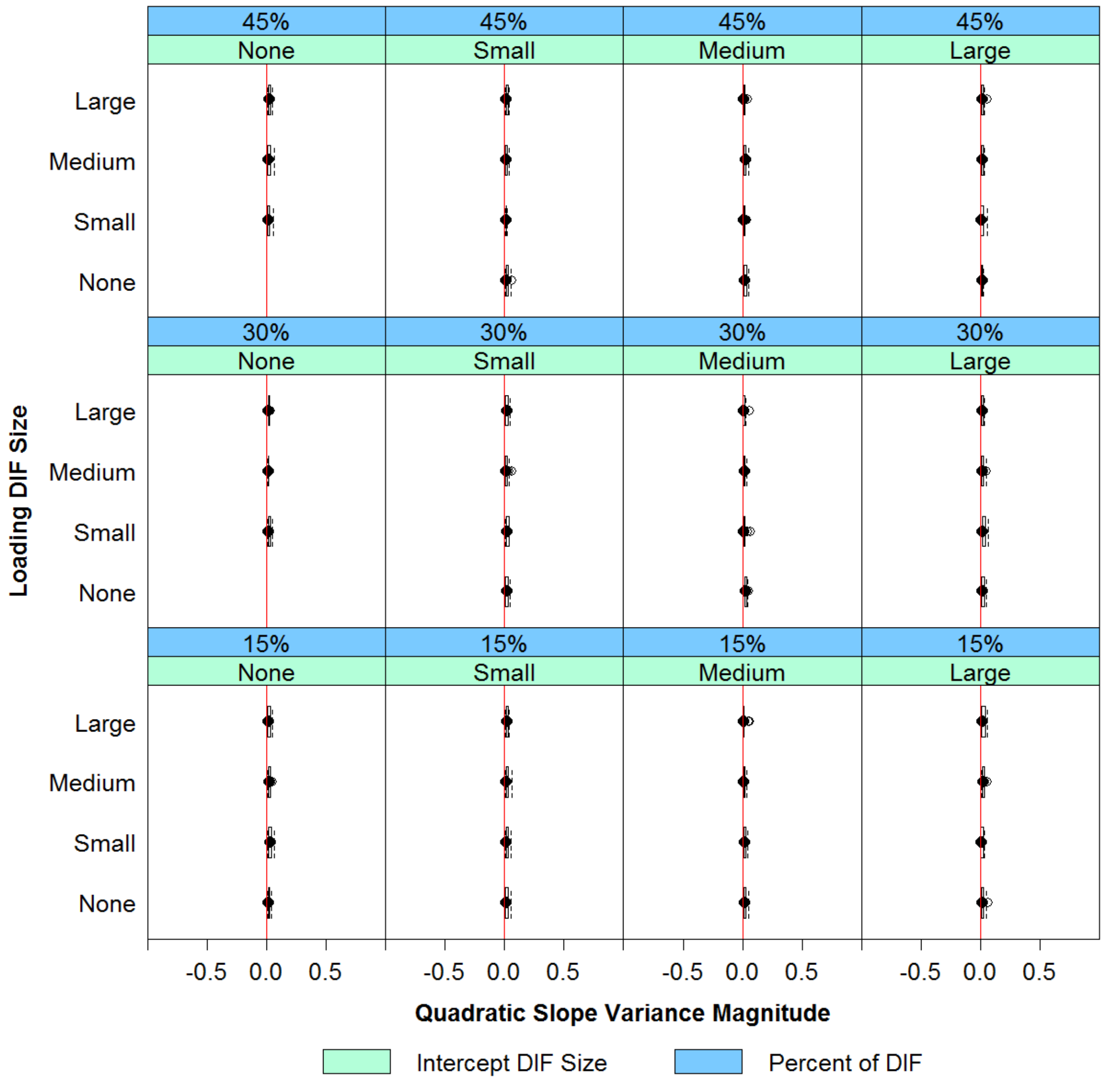

Figure 98. Quadratic Slope Variance Estimate for the Free Second-Order Non-Linear Latent Growth Model.

Preliminary Examination of Test Length

Based on the results of the 20-item conditions, two conditions were prioritized for examination with a 40-item exam. The results of the 20-item exam suggested that the size of loading DIF often had little to no impact on the results. The proportion of items with

DIF and the size of intercept DIF often had an impact that increased as the proportion of items and size of DIF increased. Based on these results, two conditions were examined with a 40-item exam: 15% of items with large intercept DIF and no loading DIF, and 30% of items with large intercept DIF and no loading DIF. Due to estimation time constraints, 5 replications of each condition were simulated and estimated for the 2LGMF and 2LGMC modeling approaches.

Results for the limited number of replications for the 40-item conditions looked very similar to the 20-item exam. The 2LGMC and 2LGMF results for the selected 20- and 40-item conditions are presented in Appendix U and V, respectively. The fit indices, parameter recovery, and the incorrectly specified non-linear growth parameters were all comparable to the corresponding 20-item exam results. These results suggest that type of DIF and proportion of DIF items have a large impact, but that the length of the test itself does not seem to be an important factor.

CHAPTER V

DISCUSSION

The purpose of the current study was to examine the impact of longitudinal measurement invariance violations on growth models. Six longitudinal growth models were fit to data with varying degrees of measurement invariance violations. A first-order linear growth model was fit to examine the impact of ignoring invariance violations using a summated score. A second-order linear growth model that constrained all item parameters to be equivalent across measurement occasions allowed the examination of invariance violations when item parameters were modeled with a 2PL. This model differs from the first-order model such that invariance violations can be examined when the items are allowed to be weighted by their factor loadings (*a*-parameter). A second-order linear growth model in which invariant item parameters were fixed to be equivalent across time and non-invariant item parameters were allowed to be freely estimated was also fit to the data. This model examined the effect of modeling the invariance violation (i.e., allowing for partial measurement invariance). In addition to the three linear growth models described, analogous non-linear growth models were estimated. The non-linear models examined whether invariance violations influence the estimated shape of growth.

Model convergence, model fit, growth model parameter recovery, and the shape of growth were evaluated. All of the linear models converged to a solution, whereas many of the non-linear models did not converge. This is likely due to the fact that the non-linear models are grossly misspecified (Diallo, Morin, & Parker, 2014). Notably, the non-convergence of the non-linear models should be an indicator to researchers that the model may be misspecified.

The model fit was evaluated using three fit indices: chi-squared, RMSEA, and CFI. For the first-order linear growth model, the chi-squared values were often significant, suggesting that the model does not fit the data well. The chi-squared values were most consistently significant in the conditions with a moderate or large proportion of DIF and/or moderate or large intercept DIF. The RMSEA and CFI values for the first-order linear growth model all suggest that the model fits the data well. The fit statistics for the first-order non-linear models all suggest that the non-linear growth model adequately fits the data across all conditions. Thus, the traditional fit indices provided no indication of model misspecification in the evaluation of overall growth model fit. The fit statistics for all of the second-order models, both linear and non-linear, all followed the same pattern. The chi-squared values were unrelated to the conditions varied in the study and suggested that none of the second-order models fit the data well. The RMSEA and CFI were also unrelated to the conditions in the study and all suggested that the second-order models sufficiently fit the data. The traditional fit indices were not able to identify the invariance violations or that the second-order free latent growth model was the true model.

Three nested model comparisons were conducted to examine whether a chi-squared difference test could detect differences between nested models. The first model comparison examined whether the addition of a non-linear slope (and the accompanying variance components) significantly improved model fit in the first-order models. The results suggested that the non-linear model often suggested that the non-linear parameters significantly improved model fit over the linear model. As the proportion of DIF and the size of the intercept DIF increased, the chi-squared difference test more consistently suggested that the non-linear model significantly improved fit. In addition to the traditional fit indices, the chi-squared difference test also suggested that the non-linear model is the best fitting model to the data. This is consistent with the results found in Wirth (2008) that suggest DIF may lead to incorrect conclusions about the shape of growth. The second model comparison also examined the differences between a linear and non-linear model, but used the second-order models. This comparison provided insight about whether the inclusion of a measurement model (with the non-invariant item parameters freely estimated) could alleviate the shape of growth issues created by invariance violations and identify the true (i.e., 2LGMF) model. The results of the chi-squared difference test between the second-order free linear model and the second-order free non-linear model suggested that the non-linear model often significantly improved model fit over the linear model. This finding was more prevalent in conditions with a large proportion of items with large intercept and loading DIF. The inclusion of the measurement model may have marginally alleviated the shape of growth issues caused by invariance violations as the number of significant chi-squared difference tests for the second-order models was slightly lower than the first-order models. The results, however, still suggest that when DIF is present a researcher would often choose a non-linear model over the true linear model based on the results of the chi-squared difference tests. The final model comparison examined the differences between the second-order free and second-order constrained models. This comparison provided insight regarding whether modeling partial measurement invariance significantly improved model fit. The results suggested that modeling the partial measurement invariance almost always significantly improved model fit. Thus, the chi-squared difference test was able to identify the true model (i.e., 2LGMF). Overall, the nested model comparisons suggest that researchers may be able to correctly model partial measurement invariance, but that DIF may cause researchers to come to erroneous conclusions about the shape of growth, regardless of whether partial measurement invariance is modeled.

The intercept variance, slope, and slope variance parameters were examined to determine how well each model recovered the growth model parameters. The pattern of parameter recovery was similar across all three linear models (LGM, 2LGMC, and 2LGMF). The intercept variance parameter was modestly recovered by all three linear growth models for all conditions. For the second-order models, the intercept parameter was often slightly overestimated. The slope parameter was recovered well by the linear models in some of the conditions. The slope parameter was overestimated in the conditions with 30% of items with large intercept DIF, 45% of items with moderate intercept DIF, and 45% of items with large intercept DIF. This finding was most prominent in the second-order free latent growth model, which is the “true” model. This finding suggests that researchers estimating growth models in situations where the measure contains longitudinal measurement invariance violations could be misled about the magnitude of growth. In the case where items get easier over time, growth is overestimated. The reverse is presumably true, although it was not addressed in the current research. The slope variance was not well recovered by any of the linear models across all conditions. As with the linear models, the pattern of parameter recovery was similar across the three non-linear models (NLGM, 2NLGMC, and 2NLGMF). Notably, many of the non-linear models did not converge so the results should be interpreted with caution. The range of intercept variance parameters was notably larger than in the linear models. The slope parameter was overestimated in the non-linear models. The overestimation was most severe in conditions with 30% of items with moderate intercept DIF, 30% of items with large intercept DIF, 45% of items with moderate intercept DIF, and 45% of items with large intercept DIF. As with the linear models, the results suggest that, when longitudinal measurement invariance violations are present, researchers may make inaccurate conclusions about the magnitude of change over time. Again, in this study the growth was overestimated, but varying DIF directions were not examined. The slope variance parameter was not recovered well by any of the non-linear models in any of the conditions. As with the intercept variance, the slope variance estimates were far more variable in the non-linear models compared to the linear models.

The model parameters estimated in the growth model that were not included in the generating model (i.e., factor covariances, non-linear growth) were examined to investigate how well the models recover the, ideally, non-significant parameters. The covariance between the intercept and slope is the only parameter examined across all six modeling approaches. For the first-order linear modeling approach, the intercept-slope covariance had a wide range of *p*-values, but was consistently estimated near zero. The *p*-

values were less variable as the proportion of DIF and size of intercept DIF increased. The pattern of recovery for the intercept-slope covariance in the second-order linear models was similar. The *p*-values for the parameter were often significant, but the estimated values were near zero. The linear model results tend to suggest a statistically, but not practically, significant covariance between the intercept and slope factors. For all three non-linear models, the *p*-values typically suggested a non-significant intercept slope covariance across all conditions. The intercept-slope covariance estimates for the non-linear models were near zero and had a wider range than their linear model counterparts. The additional covariance parameters for the non-linear models (intercept-quadratic slope, slope-quadratic slope) followed the same pattern as the intercept-slope covariance for the non-linear models. Thus, the *p*-values correctly identified the parameters as non-significant and the estimate was near zero for all three non-linear models across all conditions. The quadratic slope *p*-values for the first-order non-linear model were often non-significant, but were related to the conditions varied in the study. The *p*-values were more consistent and lower (i.e., significant) as the amount of DIF and the size of intercept DIF increased. The estimate of the quadratic slope was consistently near zero. The first-order non-linear growth model may suggest a statistically, but not practically significant quadratic slope parameter in contexts with a large amount of large intercept DIF. For both second-order non-linear models, the *p*-values suggested that the quadratic slope was statistically significant, whereas the near-zero estimates suggested that the quadratic slope was not practically significant across most conditions. The quadratic slope variance was estimated as statistically and practically non-significant for all three non-linear models. Overall, the nested model comparisons and quadratic slope significance tests may lead researchers to choose non-linear models over linear models, particularly when a large amount of intercept DIF is present. The magnitude of the estimate, however, should provide evidence that, while statistically significant, the non-linear model does not add any practical significance above and beyond the linear model.

In summary, substantial longitudinal item invariance violations may be problematic when estimating growth models. If the model converges to a solution, traditional fit indices may not identify a misspecified model. Interestingly, the first-order non-linear model often had a non-significant chi-squared value despite a large sample size. In addition, chi-squared difference tests often suggest a non-linear model fits significantly better than a linear model. Thus, in growth models, particularly in first-order models, DIF may manifest as a non-linear slope parameter. The slope and slope variance parameters, which are arguably the most important parameters of interest in a growth model, were often not recovered well, regardless of the type of model used. The proportion of items with DIF and the size of the intercept DIF had the largest impact on the parameter recovery. The size of DIF in item loadings had essentially no impact on how well each model recovered the true growth model parameters. Practitioners should also take care to consider the statistical and practical significance of the growth model parameters. For several of the parameters modeled that were not included in the generating model, the *p*-values suggested statistical significance, whereas the actual estimate was likely not practically significant. Most notably, the second-order free latent growth model, the true model which allowed for partial measurement invariance, did not alleviate the issues caused by the item invariance violations. In fact, the slope parameter recovery was worst for the models that freely estimated the non-invariant item parameters. This result suggests that modeling the partial measurement invariance does not sufficiently alleviate the concerns with invariance violations. Practitioners should take care to examine longitudinal measurement invariance before interpreting the results of a growth model. Ultimately, the results of this study suggest that if item parameters are not invariant across measurement occasions, growth modeling is not appropriate.

Limitations and Future Directions

There are several limitations and potential future directions for this research. One limitation of the study is the fact that several of the non-linear models did not converge. This resulted in fewer replications per condition in the non-linear modeling approaches. Because there are fewer replications per condition, the results of the non-linear models should all be interpreted with caution as they may be less stable. In addition, the common features of the non-converged results are unknown and thus, the results of the non-linear models may be biased in unknown ways. Because the only results able to be interpreted for the non-linear models were the replications that were able to converge to a solution even when the model was grossly misspecified, the non-linear results may be biased (i.e., may include the most extreme results). Future studies may want to run more replications in order to ensure a sufficient number of replications per condition.

Another limitation is that the 40-item condition, which would have allowed researchers to examine whether the impact of DIF was mitigated by having more items overall, was dropped due to long estimation times. The preliminary results suggested that there was minimal difference between the 20- and 40-item exams. Future research may want to investigate whether the impact of longitudinal measurement invariance violations are related to the total number of items on an exam.

The current study limited the examination of parameter recovery to the growth model parameters. These are often the parameters of interest for researchers. The results of the study suggested that the second-order models fit the data well, regardless of whether they were the correct model, but the growth model parameters were often not well recovered. Future studies may want to examine the recovery of item parameters to determine how item parameter recovery is related to growth model parameter recovery.

An additional limitation is that the current study is only one version of the potential issues with invariance violations in growth models. This study examined a variety of factors that may impact growth models, but was certainly not a comprehensive examination of all possible factors. Future studies should examine a variety of additional factors. One additional factor that could be examined in future studies is sample size and whether invariance violation issues are exacerbated with smaller sample sizes. Another factor that could be further explored is the type of invariance violation. This study examined invariance violations when the violations for all items occurred in the same direction and were the same size (within a condition). Future studies could examine the impact of DIF on growth models when the size and direction are varied within a condition. A third factor that could be examined in future studies is the type of model specifications. In this study we examined whether DIF impacted the shape of growth (i.e., quadratic slope). Future studies may want to examine the impact of DIF on predictors of growth, growth comparisons across groups, etc. Another specification approach could be to examine the impact of modeling partial invariance incorrectly (i.e., freeing invariant items and constraining non-invariant items).

This study provided a preliminary investigation of growth modeling concerns in the presence of longitudinal measurement invariance violations. The results suggest that the proportion of non-invariant items and the size of intercept invariance violations have the most significant impact on results. Researchers should use extreme caution when estimating growth models when DIF is present as it may lead to spurious conclusions about change over time.

REFERENCES

Ackerman, T. (1992). Assessing construct validity using multidimensional item response theory. Presented at the American Educational Research Association, San Francisco, CA.

Bible, L., & Tadros, H. (2014). Do values change over time? An exploratory study of business majors. *Journal of Legal, Ethical and Regulatory Issues*, *17*(1), 97–120.

Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, *44*(11), S176–S181.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456 – 466.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005 – 1018.

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.

Diallo, T. M. O., Morin, A. J. S., & Parker, P. D. (2014). Statistical power of latent growth curve models to detect quadratic growth. *Behavior Research Methods, 46*(2), 357–371. https://doi.org/10.3758/s13428-013-0395-1

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *4*(1), 22.

Fleishman, J. A., Spector, W. D., & Altman, B. M. (2002). Impact of differential item functioning on age and gender differences in functional disability. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(5), S275–S284.

Goldstein, H. (1983). Measuring changes in educational attainment over time: Problems and possibilities. *Journal of Educational Measurement*, 369–377.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Springer Science & Business Media.

Hoogland, J. J., & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods & Research*, *26*(3), 329–367. https://doi.org/10.1177/0049124198026003003

Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 258–286). Cambridge: Cambridge University Press.

Jaakkola, T., Wang, J., Yli-Piipari, S., & Liukkonen, J. (2015). A multilevel latent growth modelling of the longitudinal changes in motivation regulations in physical education. *Journal of Sports Science and Medicine*, *14*, 163–171.

Jones, R. N., & Gallo, J. J. (2002). Education and sex differences in the mini-mental state examination effects of differential item functioning. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(6), P548–P558.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, 79–93.

King, R. B., & McInerney, D. M. (2014). Mapping changes in students' English and math self-concepts: a latent growth model study. *Educational Psychology*, *34*(5), 581–597.

Lamote, C., Pinxten, M., Van Den Noortgate, W., & Van Damme, J. (2014). Is the cure worse than the disease? A longitudinal study on the effect of grade retention in secondary education on achievement and academic self-concept. *Educational Studies*, *40*(5), 496–514.

Leite, W. L. (2007). A Comparison of Latent Growth Models for Constructs Measured by Multiple Items. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 581–610. https://doi.org/10.1080/10705510701575438

Millsap, R. E., & Meredith, W. (2004). Factorial invariance: Historical trends and new developments. In *factor analysis at 100" conference*.

Muthén, B. (2012). Binary CFA vs IRT. *Mplus Discussion*.

Nye, C. (2011). *The development and validation of effect size measures for IRT and CFA studies of measurement equivalence* (Dissertation). University of Illinois at Urbana-Champaign, Urbana, Illinois.

Olivera-Aguilar, M. (2013). *Impact of Violations of Longitudinal Measurement Invariance in Latent Growth Models and Autoregressive Quasi-simplex Models* (Dissertation). Arizona State University, Arizona.

Ouweneel, E., Schaufeli, W. B., & Le Blanc, P. M. (2013). Believe, and You Will Achieve: Changes over Time in Self-Efficacy, Engagement, and Performance: Self-Efficacy, Engagement, And Performance. *Applied Psychology: Health and Well-Being*, *5*(2), 225–247.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*(3), 517–529.

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

Ryoo, J. H., Molfese, V. J., Heaton, R., Zhou, X., Brown, E. T., Prokasky, A., & Davis, E. (2014). Early Mathematics Skills From Prekindergarten to First Grade Score Changes and Ability Group Differences in Kentucky, Nebraska, and Shanghai Samples. *Journal of Advanced Academics*, *25*(3), 162–188.

Schmitt, N., Golubovich, J., & Leong, F. T. L. (2011). Impact of Measurement Invariance on Construct Correlations, Mean Differences, and Relations With External Correlates: An Illustrative Example Using Big Five and RIASEC Measures. *Assessment*, *18*(4), 412–427.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222.

Setzer, J. C. (2008). *Parameter recovery of the explanatory multidimensional Rasch model* (Unpublished Doctoral Dissertation). James Madison University, Harrisonburg, VA.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, *25*(1), 78–107.

Steinmetz, H. (2013). Analyzing Observed Composite Differences Across Groups. *Methodology*, *9*(1), 1–12.

Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*(3), 393–408.

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70.

West, B. T., Welch, K. B., & Galecki, A. T. (2006). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton: Chapman and Hall/CRC.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial Invariance Within Longitudinal Structural Equation Models: Measuring the Same Construct Across Time. *Child Development Perspectives*, *4*(1), 10–18.

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.

Wirth, R. J. (2008). *The effects of measurement non-invariance on parameter estimation in latent growth models* (Dissertation). University of North Carolina at Chapel Hill, North Carolina.

Zwick, R. (2012). A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report Series*, *2012*(1), i–30.

APPENDIX A

CHI-SQUARED RESULTS

First-Order Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.110 (0.177)	0.031 (0.077)	0.003 (0.006)	0.003 (0.009)
	Medium	0.058 (0.093)	0.040 (0.108)	0.018 (0.069)	0.003 (0.015)
	Small	0.131 (0.262)	0.050 (0.083)	0.025 (0.050)	0.023 (0.060)
	None	-- --	0.030 (0.046)	0.017 (0.041)	0.012 (0.027)
30%	Large	0.107 (0.156)	0.055 (0.101)	0.021 (0.054)	0.002 (0.003)
	Medium	0.126 (0.201)	0.082 (0.187)	0.033 (0.089)	0.020 (0.049)
	Small	0.029 (0.063)	0.089 (0.145)	0.057 (0.092)	0.010 (0.038)
	None	-- --	0.061 (0.136)	0.043 (0.117)	0.029 (0.087)
15%	Large	0.064 (0.088)	0.109 (0.175)	0.085 (0.183)	0.023 (0.056)
	Medium	0.092 (0.222)	0.045 (0.093)	0.097 (0.164)	0.050 (0.133)
	Small	0.090 (0.171)	0.086 (0.169)	0.078 (0.099)	0.102 (0.215)
	None	0.147 (0.173)	0.098 (0.153)	0.071 (0.150)	0.078 (0.133)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	None	-- --	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
30%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.000 (0.001)	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)
	None	-- --	0.001 (0.002)	0.001 (0.004)	0.000 (0.000)
15%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.000 (0.001)	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)
	None	0.001 (0.003)	0.001 (0.004)	0.006 (0.027)	0.000 (0.000)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.001 (0.002)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.018 (0.060)	0.003 (0.008)	0.002 (0.006)	0.000 (0.001)
	None	-- --	0.002 (0.006)	0.005 (0.011)	0.001 (0.002)
30%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.001 (0.004)	0.000 (0.001)	0.004 (0.018)	0.000 (0.000)
	None	-- --	0.004 (0.010)	0.003 (0.014)	0.000 (0.000)
15%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.001 (0.004)	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)
	None	0.002 (0.006)	0.002 (0.003)	0.017 (0.080)	0.000 (0.000)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.372 (0.310)	0.395 (0.367)	0.416 (0.328)	0.302 (0.267)
	Medium	0.580 (0.314)	0.367 (0.315)	0.471 (0.335)	0.393 (0.278)
	Small	0.443 (0.336)	0.433 (0.215)	0.332 (0.373)	0.375 (0.285)
	None	-- --	0.295 (0.267)	0.409 (0.261)	0.350 (0.298)
30%	Large	0.502 (0.279)	0.378 (0.253)	0.342 (0.286)	0.425 (0.294)
	Medium	0.484 (0.275)	0.386 (0.368)	0.246 (0.256)	0.239 (0.249)
	Small	0.391 (0.272)	0.443 (0.349)	0.330 (0.251)	0.542 (0.249)
	None	-- --	0.310 (0.272)	0.283 (0.264)	0.360 (0.306)
15%	Large	0.519 (0.325)	0.476 (0.366)	0.429 (0.282)	0.504 (0.336)
	Medium	0.458 (0.314)	0.316 (0.284)	0.456 (0.251)	0.223 (0.194)
	Small	0.280 (0.226)	0.444 (0.286)	0.491 (0.287)	0.415 (0.323)
	None	0.598 (0.282)	0.356 (0.346)	0.477 (0.377)	0.496 (0.343)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	None	#N/A #N/A	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
30%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.002 (0.006)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	None	#N/A #N/A	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)
15%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.001 (0.004)	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)
	None	0.000 (0.000)	0.001 (0.002)	0.001 (0.001)	0.000 (0.000)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.001)	0.001 (0.002)	0.000 (0.000)	0.000 (0.000)
	Small	0.026 (0.035)	0.022 (0.063)	0.002 (0.004)	0.000 (0.000)
	None	-- --	0.011 (0.023)	0.009 (0.021)	0.000 (0.000)
30%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.004 (0.005)	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)
	None	-- --	0.002 (0.003)	0.002 (0.008)	0.000 (0.000)
15%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.002 (0.006)	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)
	None	0.000 (0.000)	0.001 (0.002)	0.001 (0.002)	0.000 (0.000)

APPENDIX B

ROOT MEAN SQUARE ERROR OF APPROXIMATION RESULTS

First-Order Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.024 (0.013)	0.030 (0.010)	0.036 (0.010)	0.042 (0.011)
	Medium	0.028 (0.012)	0.031 (0.012)	0.036 (0.011)	0.041 (0.008)
	Small	0.022 (0.011)	0.029 (0.013)	0.032 (0.012)	0.032 (0.010)
	None	-- --	0.029 (0.011)	0.034 (0.011)	0.036 (0.013)
30%	Large	0.022 (0.012)	0.029 (0.013)	0.030 (0.009)	0.038 (0.010)
	Medium	0.020 (0.011)	0.026 (0.012)	0.032 (0.011)	0.030 (0.008)
	Small	0.026 (0.006)	0.022 (0.010)	0.026 (0.011)	0.036 (0.010)
	None	-- --	0.028 (0.012)	0.030 (0.011)	0.031 (0.011)
15%	Large	0.024 (0.009)	0.023 (0.012)	0.023 (0.009)	0.031 (0.010)
	Medium	0.024 (0.010)	0.028 (0.011)	0.023 (0.011)	0.030 (0.012)
	Small	0.024 (0.011)	0.024 (0.011)	0.023 (0.012)	0.027 (0.014)
	None	0.018 (0.010)	0.024 (0.012)	0.024 (0.009)	0.026 (0.013)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.012 (0.002)	0.012 (0.002)	0.012 (0.002)	0.014 (0.002)
	Medium	0.009 (0.001)	0.009 (0.001)	0.010 (0.001)	0.011 (0.001)
	Small	0.007 (0.001)	0.007 (0.001)	0.008 (0.001)	0.009 (0.001)
	None	-- --	0.007 (0.001)	0.007 (0.001)	0.008 (0.001)
30%	Large	0.011 (0.002)	0.011 (0.002)	0.011 (0.002)	0.012 (0.001)
	Medium	0.008 (0.001)	0.008 (0.001)	0.009 (0.001)	0.010 (0.001)
	Small	0.007 (0.001)	0.007 (0.001)	0.008 (0.001)	0.009 (0.001)
	None	-- --	0.007 (0.001)	0.007 (0.001)	0.009 (0.001)
15%	Large	0.009 (0.002)	0.009 (0.002)	0.010 (0.002)	0.010 (0.001)
	Medium	0.007 (0.001)	0.008 (0.001)	0.008 (0.001)	0.009 (0.001)
	Small	0.007 (0.001)	0.007 (0.001)	0.007 (0.001)	0.008 (0.001)
	None	0.006 (0.001)	0.006 (0.001)	0.006 (0.001)	0.008 (0.001)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.010 (0.002)	0.010 (0.002)	0.010 (0.002)	0.011 (0.002)
	Medium	0.007 (0.001)	0.008 (0.001)	0.008 (0.001)	0.008 (0.001)
	Small	0.006 (0.001)	0.006 (0.001)	0.006 (0.001)	0.007 (0.001)
	None	-- --	0.006 (0.001)	0.006 (0.001)	0.006 (0.001)
30%	Large	0.011 (0.002)	0.010 (0.002)	0.011 (0.002)	0.012 (0.002)
	Medium	0.008 (0.001)	0.008 (0.001)	0.009 (0.001)	0.010 (0.001)
	Small	0.006 (0.001)	0.007 (0.001)	0.007 (0.001)	0.008 (0.001)
	None	-- --	0.006 (0.001)	0.007 (0.001)	0.008 (0.001)
15%	Large	0.009 (0.002)	0.009 (0.002)	0.010 (0.002)	0.010 (0.001)
	Medium	0.007 (0.001)	0.007 (0.001)	0.008 (0.001)	0.009 (0.001)
	Small	0.006 (0.001)	0.007 (0.001)	0.007 (0.001)	0.008 (0.001)
	None	0.006 (0.001)	0.006 (0.001)	0.006 (0.001)	0.008 (0.001)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.012 (0.014)	0.014 (0.015)	0.010 (0.016)	0.018 (0.025)
	Medium	0.004 (0.007)	0.012 (0.014)	0.009 (0.012)	0.010 (0.013)
	Small	0.010 (0.013)	0.004 (0.008)	0.018 (0.016)	0.012 (0.019)
	None	-- --	0.015 (0.015)	0.008 (0.012)	0.015 (0.017)
30%	Large	0.005 (0.009)	0.008 (0.011)	0.012 (0.012)	0.008 (0.012)
	Medium	0.007 (0.014)	0.014 (0.016)	0.020 (0.018)	0.018 (0.014)
	Small	0.009 (0.015)	0.010 (0.012)	0.015 (0.020)	0.004 (0.008)
	None	-- --	0.015 (0.017)	0.015 (0.014)	0.015 (0.017)
15%	Large	0.006 (0.009)	0.011 (0.015)	0.006 (0.009)	0.010 (0.020)
	Medium	0.008 (0.011)	0.015 (0.017)	0.005 (0.008)	0.020 (0.023)
	Small	0.013 (0.013)	0.009 (0.016)	0.006 (0.010)	0.010 (0.015)
	None	0.003 (0.007)	0.017 (0.018)	0.012 (0.019)	0.009 (0.014)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.012 (0.002)	0.013 (0.003)	0.013 (0.001)	0.014 (0.002)
	Medium	0.009 (0.001)	0.009 (0.001)	0.010 (0.001)	0.011 (0.001)
	Small	0.006 (0.001)	0.007 (0.001)	0.008 (0.001)	0.009 (0.001)
	None	#N/A #N/A	0.007 (0.001)	0.008 (0.001)	0.009 (0.001)
30%	Large	0.012 (0.002)	0.011 (0.002)	0.012 (0.002)	0.013 (0.001)
	Medium	0.008 (0.001)	0.008 (0.001)	0.009 (0.001)	0.010 (0.001)
	Small	0.006 (0.001)	0.007 (0.001)	0.008 (0.001)	0.009 (0.001)
	None	#N/A #N/A	0.006 (0.001)	0.007 (0.001)	0.009 (0.001)
15%	Large	0.010 (0.002)	0.010 (0.002)	0.011 (0.001)	0.011 (0.002)
	Medium	0.007 (0.001)	0.008 (0.001)	0.008 (0.002)	0.009 (0.001)
	Small	0.006 (0.001)	0.007 (0.001)	0.007 (0.001)	0.008 (0.001)
	None	0.007 (0.001)	0.006 (0.001)	0.006 (0.001)	0.008 (0.001)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.010 (0.002)	0.011 (0.002)	0.010 (0.002)	0.011 (0.002)
	Medium	0.007 (0.002)	0.007 (0.001)	0.008 (0.001)	0.008 (0.001)
	Small	0.005 (0.001)	0.005 (0.001)	0.006 (0.001)	0.007 (0.001)
	None	-- --	0.005 (0.001)	0.006 (0.001)	0.007 (0.001)
30%	Large	0.011 (0.002)	0.011 (0.003)	0.011 (0.002)	0.012 (0.001)
	Medium	0.008 (0.001)	0.008 (0.001)	0.009 (0.001)	0.009 (0.001)
	Small	0.006 (0.001)	0.006 (0.001)	0.007 (0.001)	0.008 (0.001)
	None	-- --	0.006 (0.001)	0.007 (0.002)	0.008 (0.001)
15%	Large	0.010 (0.003)	0.010 (0.002)	0.011 (0.002)	0.011 (0.002)
	Medium	0.007 (0.001)	0.007 (0.001)	0.008 (0.002)	0.009 (0.001)
	Small	0.006 (0.001)	0.007 (0.001)	0.007 (0.001)	0.008 (0.001)
	None	0.007 (0.001)	0.006 (0.001)	0.006 (0.001)	0.007 (0.001)

APPENDIX C

COMPARATIVE FIT INDEX RESULTS

First-Order Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.995 (0.004)	0.993 (0.004)	0.991 (0.005)	0.987 (0.007)
	Medium	0.994 (0.004)	0.993 (0.005)	0.991 (0.005)	0.988 (0.004)
	Small	0.996 (0.003)	0.993 (0.006)	0.992 (0.006)	0.993 (0.004)
	None	-- --	0.993 (0.005)	0.992 (0.005)	0.990 (0.008)
30%	Large	0.996 (0.003)	0.993 (0.005)	0.993 (0.003)	0.990 (0.005)
	Medium	0.997 (0.003)	0.994 (0.004)	0.993 (0.004)	0.993 (0.003)
	Small	0.995 (0.002)	0.996 (0.003)	0.995 (0.004)	0.991 (0.005)
	None	-- --	0.994 (0.004)	0.993 (0.004)	0.993 (0.004)
15%	Large	0.996 (0.003)	0.996 (0.004)	0.996 (0.002)	0.993 (0.004)
	Medium	0.996 (0.003)	0.994 (0.004)	0.996 (0.003)	0.993 (0.004)
	Small	0.996 (0.003)	0.996 (0.003)	0.996 (0.005)	0.994 (0.005)
	None	0.997 (0.002)	0.995 (0.004)	0.996 (0.003)	0.995 (0.004)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.995 (0.002)	0.995 (0.002)	0.994 (0.002)	0.992 (0.002)
	Medium	0.997 (0.001)	0.997 (0.001)	0.996 (0.001)	0.996 (0.001)
	Small	0.998 (0.000)	0.998 (0.000)	0.998 (0.001)	0.997 (0.001)
	None	-- --	0.998 (0.000)	0.998 (0.001)	0.998 (0.001)
30%	Large	0.996 (0.002)	0.996 (0.002)	0.995 (0.002)	0.995 (0.001)
	Medium	0.998 (0.001)	0.998 (0.001)	0.997 (0.001)	0.996 (0.001)
	Small	0.998 (0.000)	0.998 (0.000)	0.998 (0.001)	0.997 (0.001)
	None	-- --	0.999 (0.000)	0.998 (0.001)	0.998 (0.001)
15%	Large	0.997 (0.001)	0.997 (0.001)	0.997 (0.001)	0.996 (0.001)
	Medium	0.998 (0.001)	0.998 (0.000)	0.998 (0.001)	0.998 (0.001)
	Small	0.999 (0.001)	0.998 (0.000)	0.998 (0.001)	0.998 (0.001)
	None	0.999 (0.000)	0.999 (0.001)	0.999 (0.000)	0.998 (0.000)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.996 (0.001)	0.996 (0.002)	0.996 (0.002)	0.995 (0.002)
	Medium	0.998 (0.001)	0.998 (0.001)	0.998 (0.001)	0.998 (0.001)
	Small	0.999 (0.000)	0.999 (0.000)	0.999 (0.001)	0.998 (0.001)
	None	-- --	0.999 (0.000)	0.999 (0.001)	0.999 (0.001)
30%	Large	0.996 (0.002)	0.996 (0.002)	0.996 (0.002)	0.995 (0.001)
	Medium	0.998 (0.001)	0.998 (0.001)	0.997 (0.001)	0.997 (0.001)
	Small	0.999 (0.001)	0.999 (0.000)	0.998 (0.001)	0.998 (0.001)
	None	-- --	0.999 (0.000)	0.999 (0.001)	0.998 (0.001)
15%	Large	0.997 (0.001)	0.997 (0.001)	0.997 (0.001)	0.996 (0.001)
	Medium	0.998 (0.001)	0.998 (0.000)	0.998 (0.001)	0.998 (0.001)
	Small	0.999 (0.001)	0.998 (0.000)	0.998 (0.001)	0.998 (0.001)
	None	0.999 (0.000)	0.999 (0.001)	0.999 (0.000)	0.998 (0.000)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	1.000 (0.001)	0.999 (0.001)	1.000 (0.001)	0.999 (0.002)
	Medium	1.000 (0.000)	1.000 (0.001)	1.000 (0.001)	1.000 (0.000)
	Small	1.000 (0.001)	1.000 (0.000)	0.999 (0.001)	0.999 (0.001)
	None	-- --	0.999 (0.001)	1.000 (0.001)	0.999 (0.001)
30%	Large	1.000 (0.000)	1.000 (0.000)	1.000 (0.001)	1.000 (0.001)
	Medium	1.000 (0.001)	0.999 (0.001)	0.999 (0.001)	0.999 (0.001)
	Small	1.000 (0.001)	1.000 (0.000)	0.999 (0.002)	1.000 (0.000)
	None	-- --	0.999 (0.001)	0.999 (0.001)	0.999 (0.001)
15%	Large	1.000 (0.000)	1.000 (0.001)	1.000 (0.000)	0.999 (0.001)
	Medium	1.000 (0.000)	0.999 (0.001)	1.000 (0.000)	0.999 (0.003)
	Small	1.000 (0.001)	1.000 (0.001)	1.000 (0.000)	1.000 (0.001)
	None	1.000 (0.000)	0.999 (0.001)	0.999 (0.001)	1.000 (0.001)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.995 (0.002)	0.993 (0.004)	0.994 (0.002)	0.993 (0.002)
	Medium	0.997 (0.001)	0.997 (0.001)	0.996 (0.001)	0.996 (0.001)
	Small	0.999 (0.000)	0.998 (0.000)	0.998 (0.000)	0.997 (0.001)
	None	#N/A #N/A	0.999 (0.000)	0.998 (0.000)	0.997 (0.000)
30%	Large	0.995 (0.002)	0.995 (0.002)	0.995 (0.002)	0.994 (0.001)
	Medium	0.998 (0.001)	0.998 (0.001)	0.997 (0.001)	0.997 (0.001)
	Small	0.999 (0.000)	0.998 (0.000)	0.998 (0.000)	0.997 (0.001)
	None	#N/A #N/A	0.999 (0.000)	0.998 (0.001)	0.998 (0.000)
15%	Large	0.996 (0.002)	0.997 (0.001)	0.996 (0.001)	0.996 (0.001)
	Medium	0.998 (0.000)	0.998 (0.001)	0.998 (0.001)	0.998 (0.001)
	Small	0.999 (0.000)	0.999 (0.000)	0.998 (0.001)	0.998 (0.000)
	None	0.999 (0.000)	0.999 (0.000)	0.999 (0.001)	0.998 (0.000)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.997 (0.001)	0.995 (0.003)	0.996 (0.001)	0.996 (0.002)
	Medium	0.998 (0.001)	0.998 (0.001)	0.998 (0.000)	0.998 (0.001)
	Small	0.999 (0.000)	0.999 (0.000)	0.999 (0.000)	0.998 (0.000)
	None	-- --	0.999 (0.000)	0.999 (0.000)	0.998 (0.000)
30%	Large	0.996 (0.002)	0.996 (0.002)	0.996 (0.002)	0.995 (0.001)
	Medium	0.998 (0.001)	0.998 (0.001)	0.997 (0.001)	0.997 (0.001)
	Small	0.999 (0.000)	0.999 (0.000)	0.998 (0.000)	0.998 (0.001)
	None	-- --	0.999 (0.000)	0.999 (0.001)	0.998 (0.000)
15%	Large	0.997 (0.002)	0.997 (0.001)	0.996 (0.001)	0.996 (0.001)
	Medium	0.998 (0.000)	0.998 (0.001)	0.998 (0.001)	0.998 (0.001)
	Small	0.999 (0.000)	0.999 (0.000)	0.998 (0.001)	0.998 (0.000)
	None	0.999 (0.000)	0.999 (0.000)	0.999 (0.001)	0.998 (0.000)

APPENDIX D

NESTED MODEL COMPARISON RESULTS

First-Order Linear vs. First-Order Non-Linear					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.008 (0.014)	0.037 (0.043)	0.005 (0.009)	0.007 (0.025)
	Medium	0.105 (0.136)	0.016 (0.042)	0.019 (0.066)	0.012 (0.047)
	Small	0.208 (0.333)	0.042 (0.050)	0.039 (0.065)	0.030 (0.063)
	None	-- --	0.053 (0.078)	0.021 (0.062)	0.023 (0.037)
30%	Large	0.090 (0.176)	0.039 (0.096)	0.028 (0.063)	0.002 (0.004)
	Medium	0.154 (0.213)	0.071 (0.129)	0.072 (0.178)	0.020 (0.043)
	Small	0.032 (0.059)	0.115 (0.204)	0.084 (0.144)	0.009 (0.032)
	None	-- --	0.048 (0.109)	0.022 (0.046)	0.036 (0.093)
15%	Large	0.042 (0.062)	0.118 (0.151)	0.100 (0.280)	0.038 (0.072)
	Medium	0.066 (0.168)	0.026 (0.044)	0.108 (0.201)	0.093 (0.219)
	Small	0.074 (0.153)	0.057 (0.097)	0.071 (0.126)	0.099 (0.168)
	None	0.137 (0.176)	0.074 (0.122)	0.109 (0.168)	0.056 (0.092)

Second-Order Linear vs. Second-Order Non-Linear					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.036 (0.064)	0.145 (0.172)	0.050 (0.081)	0.062 (0.131)
	Medium	0.181 (0.156)	0.124 (0.191)	0.084 (0.200)	0.029 (0.039)
	Small	0.304 (0.324)	0.104 (0.131)	0.101 (0.110)	0.177 (0.210)
	None	-- --	0.145 (0.182)	0.101 (0.138)	0.136 (0.130)
30%	Large	0.157 (0.231)	0.089 (0.199)	0.173 (0.210)	0.083 (0.110)
	Medium	0.181 (0.216)	0.102 (0.163)	0.053 (0.065)	0.144 (0.227)
	Small	0.072 (0.120)	0.155 (0.213)	0.131 (0.154)	0.040 (0.041)
	None	-- --	0.145 (0.191)	0.162 (0.145)	0.084 (0.096)
15%	Large	0.096 (0.114)	0.174 (0.185)	0.135 (0.152)	0.031 (0.033)
	Medium	0.037 (0.045)	0.115 (0.156)	0.133 (0.253)	0.139 (0.159)
	Small	0.073 (0.101)	0.055 (0.066)	0.147 (0.145)	0.235 (0.224)
	None	0.208 (0.159)	0.166 (0.171)	0.202 (0.186)	0.161 (0.206)

Second-Order Free vs. Second-Order Constrained					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Small	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	None	-- --	0.000 (0.002)	0.000 (0.000)	0.000 (0.000)
30%	Large	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Medium	0.000 (0.000)	0.000 (0.000)	0.001 (0.003)	0.000 (0.000)
	Small	0.017 (0.044)	0.003 (0.006)	0.002 (0.007)	0.001 (0.003)
	None	-- --	0.001 (0.002)	0.008 (0.036)	0.000 (0.001)
15%	Large	0.045 (0.071)	0.023 (0.038)	0.040 (0.079)	0.010 (0.031)
	Medium	0.051 (0.088)	0.098 (0.213)	0.038 (0.064)	0.036 (0.108)
	Small	0.023 (0.044)	0.065 (0.153)	0.071 (0.155)	0.007 (0.015)
	None	0.082 (0.155)	0.057 (0.187)	0.013 (0.022)	0.016 (0.024)

APPENDIX E

INTERCEPT VARIANCE PARAMETER BIAS ESTIMATES

First-Order Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.033 (0.099)	-0.028 (0.090)	-0.030 (0.084)	-0.028 (0.084)
	Medium	-0.013 (0.098)	-0.019 (0.090)	-0.025 (0.103)	-0.014 (0.091)
	Small	-0.054 (0.091)	-0.026 (0.093)	-0.017 (0.096)	0.015 (0.094)
	None	-- --	-0.043 (0.094)	-0.018 (0.113)	-0.005 (0.103)
30%	Large	-0.045 (0.108)	-0.043 (0.106)	-0.004 (0.116)	0.012 (0.101)
	Medium	-0.046 (0.122)	-0.034 (0.089)	-0.067 (0.096)	-0.037 (0.120)
	Small	-0.034 (0.095)	-0.040 (0.068)	-0.002 (0.089)	0.017 (0.104)
	None	-- --	-0.032 (0.106)	-0.048 (0.095)	-0.018 (0.083)
15%	Large	-0.053 (0.122)	0.001 (0.133)	-0.007 (0.126)	-0.036 (0.117)
	Medium	-0.047 (0.095)	-0.043 (0.080)	-0.024 (0.106)	-0.063 (0.108)
	Small	-0.037 (0.102)	-0.024 (0.118)	-0.036 (0.125)	-0.014 (0.091)
	None	0.009 (0.096)	-0.043 (0.100)	-0.027 (0.121)	-0.012 (0.078)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.121 (0.102)	0.105 (0.095)	0.105 (0.090)	0.098 (0.092)
	Medium	0.157 (0.096)	0.127 (0.120)	0.118 (0.118)	0.113 (0.104)
	Small	0.091 (0.111)	0.112 (0.108)	0.106 (0.096)	0.124 (0.090)
	None	-- --	0.098 (0.109)	0.106 (0.122)	0.094 (0.114)
30%	Large	0.117 (0.132)	0.107 (0.121)	0.143 (0.138)	0.162 (0.108)
	Medium	0.109 (0.131)	0.122 (0.104)	0.075 (0.108)	0.083 (0.123)
	Small	0.125 (0.114)	0.106 (0.084)	0.123 (0.093)	0.165 (0.119)
	None	-- --	0.117 (0.125)	0.079 (0.102)	0.099 (0.086)
15%	Large	0.104 (0.143)	0.153 (0.146)	0.138 (0.137)	0.119 (0.133)
	Medium	0.112 (0.107)	0.123 (0.095)	0.120 (0.119)	0.070 (0.121)
	Small	0.126 (0.128)	0.132 (0.153)	0.110 (0.143)	0.122 (0.108)
	None	0.168 (0.112)	0.104 (0.114)	0.113 (0.133)	0.128 (0.100)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.105 (0.103)	0.090 (0.096)	0.087 (0.088)	0.074 (0.085)
	Medium	0.133 (0.096)	0.108 (0.116)	0.092 (0.113)	0.095 (0.099)
	Small	0.067 (0.104)	0.096 (0.105)	0.086 (0.095)	0.110 (0.089)
	None	-- --	0.085 (0.110)	0.097 (0.117)	0.084 (0.111)
30%	Large	0.110 (0.130)	0.095 (0.120)	0.133 (0.135)	0.150 (0.107)
	Medium	0.103 (0.132)	0.114 (0.107)	0.063 (0.110)	0.074 (0.125)
	Small	0.117 (0.111)	0.096 (0.082)	0.116 (0.090)	0.153 (0.120)
	None	-- --	0.108 (0.119)	0.073 (0.104)	0.096 (0.084)
15%	Large	0.102 (0.143)	0.149 (0.146)	0.135 (0.138)	0.114 (0.133)
	Medium	0.109 (0.106)	0.119 (0.094)	0.116 (0.118)	0.068 (0.122)
	Small	0.122 (0.126)	0.130 (0.150)	0.105 (0.143)	0.122 (0.107)
	None	0.166 (0.113)	0.104 (0.115)	0.113 (0.133)	0.127 (0.099)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.038 (0.229)	-0.062 (0.208)	-0.046 (0.161)	-0.023 (0.218)
	Medium	0.002 (0.286)	-0.070 (0.213)	0.018 (0.243)	-0.149 (0.258)
	Small	0.002 (0.277)	-0.110 (0.199)	-0.141 (0.203)	-0.076 (0.195)
	None	-- --	-0.041 (0.249)	-0.115 (0.232)	-0.096 (0.234)
30%	Large	-0.027 (0.173)	-0.032 (0.220)	-0.044 (0.153)	-0.036 (0.161)
	Medium	-0.041 (0.246)	-0.046 (0.241)	-0.095 (0.240)	-0.001 (0.295)
	Small	-0.057 (0.240)	-0.098 (0.171)	-0.030 (0.269)	0.065 (0.280)
	None	-- --	-0.050 (0.260)	-0.024 (0.213)	0.018 (0.177)
15%	Large	-0.108 (0.261)	0.019 (0.248)	-0.013 (0.293)	0.000 (0.197)
	Medium	0.080 (0.201)	-0.022 (0.209)	-0.145 (0.183)	-0.014 (0.219)
	Small	-0.030 (0.335)	-0.048 (0.243)	-0.102 (0.211)	-0.159 (0.177)
	None	0.069 (0.330)	-0.087 (0.169)	-0.029 (0.289)	0.069 (0.229)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.271 (0.277)	0.203 (0.240)	0.141 (0.213)	0.163 (0.298)
	Medium	0.311 (0.362)	0.146 (0.257)	0.203 (0.278)	0.048 (0.336)
	Small	0.203 (0.346)	0.051 (0.184)	-0.039 (0.179)	0.168 (0.251)
	None	-- --	0.141 (0.333)	0.099 (0.273)	0.102 (0.292)
30%	Large	0.120 (0.195)	0.133 (0.250)	0.166 (0.179)	0.158 (0.168)
	Medium	0.193 (0.303)	0.163 (0.293)	0.157 (0.218)	0.225 (0.372)
	Small	0.167 (0.245)	0.126 (0.166)	0.182 (0.344)	0.289 (0.401)
	None	-- --	0.191 (0.291)	0.224 (0.233)	0.205 (0.194)
15%	Large	0.098 (0.247)	0.257 (0.315)	0.174 (0.275)	0.176 (0.214)
	Medium	0.305 (0.252)	0.181 (0.235)	0.104 (0.259)	0.223 (0.257)
	Small	0.240 (0.398)	0.157 (0.303)	0.103 (0.281)	0.023 (0.218)
	None	0.218 (0.351)	0.066 (0.220)	0.192 (0.362)	0.260 (0.281)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.267 (0.279)	0.217 (0.237)	0.133 (0.207)	0.169 (0.284)
	Medium	0.304 (0.341)	0.155 (0.263)	0.222 (0.291)	0.060 (0.331)
	Small	0.193 (0.307)	0.054 (0.174)	-0.027 (0.181)	0.187 (0.267)
	None	-- --	0.158 (0.322)	0.112 (0.280)	0.107 (0.298)
30%	Large	0.125 (0.198)	0.140 (0.249)	0.177 (0.197)	0.174 (0.174)
	Medium	0.194 (0.298)	0.166 (0.281)	0.160 (0.215)	0.240 (0.367)
	Small	0.170 (0.235)	0.134 (0.174)	0.187 (0.342)	0.290 (0.385)
	None	-- --	0.185 (0.288)	0.226 (0.235)	0.198 (0.197)
15%	Large	0.103 (0.251)	0.257 (0.311)	0.177 (0.271)	0.174 (0.217)
	Medium	0.307 (0.254)	0.179 (0.226)	0.106 (0.261)	0.225 (0.257)
	Small	0.243 (0.395)	0.180 (0.302)	0.104 (0.279)	0.022 (0.218)
	None	0.215 (0.341)	0.073 (0.225)	0.196 (0.371)	0.264 (0.286)

APPENDIX F

INTERCEPT VARIANCE PARAMETER ROOT MEAN SQUARE ERROR ESTIMATES

First-Order Latent Growth Model

Percent of DIF	Loading DIF	Intercept DIF: None	Small	Medium	Large
45%	Large	0.083	0.080	0.062	0.066
		(0.061)	(0.049)	(0.064)	(0.058)
	Medium	0.084	0.077	0.084	0.071
		(0.050)	(0.049)	(0.062)	(0.057)
	Small	0.089	0.076	0.078	0.070
		(0.055)	(0.058)	(0.056)	(0.064)
	None	--	0.076	0.091	0.085
		--	(0.068)	(0.067)	(0.056)
30%	Large	0.091	0.089	0.092	0.076
		(0.071)	(0.070)	(0.069)	(0.065)
	Medium	0.096	0.075	0.090	0.101
		(0.086)	(0.057)	(0.074)	(0.071)
	Small	0.080	0.063	0.070	0.078
		(0.060)	(0.046)	(0.053)	(0.069)
	None	--	0.092	0.087	0.067
		--	(0.060)	(0.058)	(0.051)
15%	Large	0.115	0.106	0.097	0.102
		(0.064)	(0.077)	(0.078)	(0.064)
	Medium	0.084	0.077	0.088	0.098
		(0.063)	(0.047)	(0.061)	(0.076)
	Small	0.088	0.076	0.095	0.069
		(0.061)	(0.093)	(0.086)	(0.058)
	None	0.077	0.084	0.098	0.062
		(0.056)	(0.068)	(0.072)	(0.047)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.125 (0.098)	0.109 (0.089)	0.121 (0.066)	0.109 (0.078)
	Medium	0.157 (0.096)	0.130 (0.117)	0.127 (0.108)	0.130 (0.081)
	Small	0.116 (0.082)	0.134 (0.079)	0.123 (0.072)	0.135 (0.072)
	None	-- --	0.124 (0.077)	0.128 (0.099)	0.124 (0.078)
30%	Large	0.150 (0.091)	0.121 (0.105)	0.157 (0.121)	0.172 (0.091)
	Medium	0.145 (0.086)	0.122 (0.104)	0.107 (0.075)	0.123 (0.081)
	Small	0.144 (0.089)	0.109 (0.080)	0.131 (0.081)	0.169 (0.113)
	None	-- --	0.134 (0.105)	0.108 (0.069)	0.103 (0.081)
15%	Large	0.128 (0.121)	0.170 (0.125)	0.161 (0.108)	0.140 (0.110)
	Medium	0.120 (0.098)	0.127 (0.088)	0.137 (0.098)	0.115 (0.077)
	Small	0.137 (0.115)	0.135 (0.150)	0.158 (0.085)	0.145 (0.073)
	None	0.168 (0.112)	0.131 (0.080)	0.141 (0.102)	0.128 (0.100)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.114 (0.093)	0.100 (0.086)	0.108 (0.058)	0.091 (0.066)
	Medium	0.133 (0.095)	0.114 (0.110)	0.105 (0.101)	0.113 (0.076)
	Small	0.096 (0.077)	0.122 (0.071)	0.109 (0.066)	0.121 (0.073)
	None	-- --	0.114 (0.078)	0.122 (0.089)	0.118 (0.072)
30%	Large	0.144 (0.088)	0.114 (0.102)	0.149 (0.117)	0.162 (0.087)
	Medium	0.143 (0.085)	0.114 (0.106)	0.102 (0.074)	0.118 (0.083)
	Small	0.137 (0.084)	0.100 (0.077)	0.125 (0.078)	0.157 (0.114)
	None	-- --	0.126 (0.100)	0.105 (0.070)	0.100 (0.080)
15%	Large	0.129 (0.119)	0.167 (0.124)	0.159 (0.108)	0.135 (0.110)
	Medium	0.118 (0.095)	0.124 (0.087)	0.134 (0.096)	0.115 (0.077)
	Small	0.134 (0.112)	0.134 (0.147)	0.153 (0.086)	0.144 (0.073)
	None	0.166 (0.113)	0.132 (0.080)	0.141 (0.102)	0.127 (0.099)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.193 (0.116)	0.176 (0.117)	0.139 (0.086)	0.167 (0.134)
	Medium	0.193 (0.203)	0.196 (0.099)	0.172 (0.167)	0.271 (0.108)
	Small	0.241 (0.106)	0.206 (0.082)	0.215 (0.113)	0.189 (0.073)
	None	-- --	0.224 (0.099)	0.222 (0.128)	0.210 (0.132)
30%	Large	0.143 (0.094)	0.175 (0.129)	0.131 (0.084)	0.119 (0.109)
	Medium	0.208 (0.125)	0.203 (0.129)	0.194 (0.162)	0.232 (0.173)
	Small	0.178 (0.164)	0.172 (0.088)	0.211 (0.159)	0.216 (0.183)
	None	-- --	0.217 (0.142)	0.174 (0.117)	0.152 (0.085)
15%	Large	0.228 (0.154)	0.215 (0.113)	0.245 (0.141)	0.175 (0.075)
	Medium	0.153 (0.150)	0.165 (0.122)	0.204 (0.105)	0.135 (0.167)
	Small	0.225 (0.242)	0.200 (0.137)	0.183 (0.142)	0.205 (0.117)
	None	0.264 (0.197)	0.154 (0.106)	0.216 (0.185)	0.183 (0.148)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.309 (0.229)	0.243 (0.193)	0.212 (0.134)	0.199 (0.273)
	Medium	0.311 (0.362)	0.198 (0.216)	0.223 (0.261)	0.243 (0.225)
	Small	0.299 (0.251)	0.139 (0.123)	0.152 (0.088)	0.204 (0.219)
	None	-- --	0.282 (0.216)	0.187 (0.219)	0.247 (0.175)
30%	Large	0.170 (0.148)	0.223 (0.166)	0.201 (0.135)	0.202 (0.100)
	Medium	0.250 (0.253)	0.224 (0.247)	0.218 (0.144)	0.321 (0.287)
	Small	0.198 (0.218)	0.163 (0.125)	0.269 (0.276)	0.323 (0.373)
	None	-- --	0.252 (0.235)	0.249 (0.204)	0.227 (0.167)
15%	Large	0.186 (0.181)	0.326 (0.237)	0.237 (0.215)	0.232 (0.144)
	Medium	0.308 (0.248)	0.219 (0.197)	0.189 (0.194)	0.246 (0.233)
	Small	0.285 (0.365)	0.228 (0.249)	0.258 (0.132)	0.163 (0.139)
	None	0.254 (0.323)	0.169 (0.148)	0.293 (0.277)	0.293 (0.243)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.308 (0.228)	0.254 (0.189)	0.202 (0.132)	0.192 (0.267)
	Medium	0.304 (0.341)	0.201 (0.226)	0.237 (0.279)	0.238 (0.228)
	Small	0.266 (0.237)	0.132 (0.119)	0.150 (0.091)	0.216 (0.240)
	None	-- --	0.283 (0.212)	0.206 (0.216)	0.252 (0.181)
30%	Large	0.176 (0.149)	0.222 (0.172)	0.216 (0.147)	0.217 (0.105)
	Medium	0.248 (0.250)	0.221 (0.237)	0.221 (0.137)	0.335 (0.275)
	Small	0.193 (0.214)	0.170 (0.133)	0.272 (0.273)	0.322 (0.356)
	None	-- --	0.245 (0.235)	0.252 (0.205)	0.215 (0.177)
15%	Large	0.188 (0.186)	0.327 (0.230)	0.237 (0.213)	0.230 (0.150)
	Medium	0.309 (0.251)	0.215 (0.189)	0.193 (0.194)	0.246 (0.234)
	Small	0.287 (0.361)	0.246 (0.247)	0.259 (0.129)	0.164 (0.137)
	None	0.251 (0.313)	0.171 (0.157)	0.304 (0.280)	0.299 (0.245)

APPENDIX G

SLOPE PARAMETER BIAS ESTIMATES

First-Order Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.068 (0.067)	0.021 (0.093)	0.069 (0.089)	0.169 (0.108)
	Medium	-0.050 (0.100)	0.017 (0.088)	0.083 (0.100)	0.162 (0.104)
	Small	-0.052 (0.062)	0.038 (0.082)	0.122 (0.091)	0.193 (0.109)
	None	-- --	0.026 (0.081)	0.119 (0.121)	0.179 (0.090)
30%	Large	-0.085 (0.077)	-0.024 (0.088)	0.060 (0.087)	0.056 (0.104)
	Medium	-0.071 (0.085)	0.003 (0.078)	0.041 (0.106)	0.113 (0.063)
	Small	-0.062 (0.082)	0.016 (0.093)	0.071 (0.092)	0.081 (0.097)
	None	-- --	-0.004 (0.080)	0.041 (0.089)	0.124 (0.080)
15%	Large	-0.054 (0.088)	-0.029 (0.077)	0.009 (0.073)	-0.015 (0.095)
	Medium	-0.072 (0.076)	-0.052 (0.074)	0.016 (0.089)	0.026 (0.083)
	Small	-0.073 (0.069)	-0.005 (0.084)	-0.005 (0.093)	0.021 (0.073)
	None	-0.033 (0.046)	-0.042 (0.063)	0.015 (0.087)	0.038 (0.068)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.066	0.023	0.074	0.151
		(0.042)	(0.052)	(0.049)	(0.047)
	Medium	-0.046	0.027	0.091	0.161
		(0.062)	(0.050)	(0.055)	(0.055)
	Small	-0.048	0.044	0.117	0.185
		(0.042)	(0.046)	(0.044)	(0.051)
	None	--	0.038	0.125	0.192
		--	(0.047)	(0.063)	(0.042)
30%	Large	-0.069	-0.013	0.052	0.069
		(0.048)	(0.048)	(0.052)	(0.055)
	Medium	-0.067	0.009	0.046	0.095
		(0.058)	(0.049)	(0.063)	(0.037)
	Small	-0.051	0.011	0.063	0.100
		(0.050)	(0.052)	(0.052)	(0.050)
	None	--	0.013	0.054	0.119
		--	(0.049)	(0.046)	(0.036)
15%	Large	-0.046	-0.027	0.002	0.001
		(0.056)	(0.046)	(0.046)	(0.058)
	Medium	-0.058	-0.031	0.009	0.022
		(0.047)	(0.038)	(0.049)	(0.040)
	Small	-0.056	-0.007	-0.005	0.026
		(0.042)	(0.056)	(0.052)	(0.041)
	None	-0.029	-0.025	0.014	0.038
		(0.028)	(0.042)	(0.053)	(0.036)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.101 (0.056)	0.087 (0.064)	0.183 (0.077)	0.300 (0.077)
	Medium	-0.060 (0.077)	0.096 (0.060)	0.206 (0.063)	0.330 (0.070)
	Small	-0.064 (0.045)	0.101 (0.069)	0.230 (0.053)	0.362 (0.074)
	None	-- --	0.115 (0.055)	0.271 (0.076)	0.399 (0.043)
30%	Large	-0.080 (0.048)	-0.001 (0.061)	0.087 (0.056)	0.119 (0.064)
	Medium	-0.075 (0.070)	0.028 (0.066)	0.076 (0.069)	0.150 (0.059)
	Small	-0.056 (0.056)	0.038 (0.055)	0.103 (0.057)	0.168 (0.061)
	None	-- --	0.037 (0.056)	0.101 (0.050)	0.186 (0.046)
15%	Large	-0.043 (0.056)	-0.022 (0.048)	0.011 (0.051)	0.009 (0.058)
	Medium	-0.060 (0.050)	-0.027 (0.038)	0.017 (0.053)	0.036 (0.042)
	Small	-0.059 (0.042)	-0.005 (0.057)	0.001 (0.053)	0.041 (0.043)
	None	-0.031 (0.033)	-0.022 (0.039)	0.024 (0.058)	0.053 (0.041)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.108 (0.098)	0.179 (0.080)	0.285 (0.113)	0.413 (0.089)
	Medium	0.044 (0.067)	0.162 (0.107)	0.271 (0.095)	0.333 (0.113)
	Small	0.046 (0.053)	0.160 (0.099)	0.262 (0.081)	0.333 (0.076)
	None	-- --	0.160 (0.115)	0.260 (0.083)	0.372 (0.068)
30%	Large	0.022 (0.075)	0.138 (0.084)	0.192 (0.087)	0.295 (0.113)
	Medium	-0.005 (0.112)	0.122 (0.089)	0.191 (0.086)	0.267 (0.109)
	Small	0.055 (0.114)	0.084 (0.080)	0.167 (0.133)	0.299 (0.098)
	None	-- --	0.123 (0.112)	0.210 (0.096)	0.262 (0.125)
15%	Large	0.018 (0.086)	0.113 (0.075)	0.114 (0.101)	0.125 (0.091)
	Medium	0.067 (0.105)	0.082 (0.096)	0.135 (0.081)	0.196 (0.119)
	Small	-0.019 (0.093)	0.063 (0.071)	0.089 (0.069)	0.118 (0.108)
	None	0.059 (0.078)	0.076 (0.084)	0.088 (0.124)	0.165 (0.112)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.164 (0.096)	0.222 (0.098)	0.310 (0.115)	0.427 (0.099)
	Medium	0.088 (0.054)	0.200 (0.107)	0.319 (0.117)	0.381 (0.126)
	Small	0.118 (0.070)	0.202 (0.097)	0.295 (0.102)	0.359 (0.086)
	None	-- --	0.210 (0.144)	0.325 (0.080)	0.442 (0.064)
30%	Large	0.066 (0.079)	0.185 (0.097)	0.244 (0.094)	0.326 (0.113)
	Medium	0.084 (0.105)	0.180 (0.094)	0.249 (0.104)	0.322 (0.113)
	Small	0.119 (0.117)	0.127 (0.072)	0.204 (0.128)	0.353 (0.122)
	None	-- --	0.195 (0.118)	0.285 (0.085)	0.339 (0.129)
15%	Large	0.069 (0.087)	0.206 (0.072)	0.163 (0.077)	0.165 (0.095)
	Medium	0.136 (0.119)	0.145 (0.091)	0.189 (0.093)	0.265 (0.125)
	Small	0.067 (0.100)	0.124 (0.071)	0.150 (0.084)	0.164 (0.106)
	None	0.113 (0.073)	0.139 (0.097)	0.185 (0.120)	0.208 (0.110)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.168 (0.092)	0.343 (0.120)	0.525 (0.126)	0.731 (0.107)
	Medium	0.109 (0.044)	0.337 (0.117)	0.531 (0.151)	0.655 (0.130)
	Small	0.122 (0.050)	0.328 (0.124)	0.477 (0.098)	0.616 (0.092)
	None	-- --	0.329 (0.131)	0.506 (0.093)	0.687 (0.084)
30%	Large	0.068 (0.078)	0.224 (0.097)	0.297 (0.104)	0.412 (0.113)
	Medium	0.094 (0.103)	0.220 (0.099)	0.319 (0.115)	0.421 (0.126)
	Small	0.117 (0.113)	0.167 (0.079)	0.267 (0.122)	0.459 (0.124)
	None	-- --	0.236 (0.125)	0.360 (0.093)	0.410 (0.116)
15%	Large	0.077 (0.095)	0.216 (0.072)	0.178 (0.078)	0.187 (0.097)
	Medium	0.137 (0.123)	0.154 (0.089)	0.205 (0.103)	0.286 (0.128)
	Small	0.068 (0.096)	0.143 (0.073)	0.165 (0.093)	0.184 (0.105)
	None	0.111 (0.079)	0.148 (0.104)	0.200 (0.129)	0.231 (0.114)

APPENDIX H

SLOPE PARAMETER ROOT MEAN SQUARE ERROR ESTIMATES

First-Order Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.037 (0.029)	0.041 (0.023)	0.047 (0.031)	0.085 (0.053)
	Medium	0.043 (0.035)	0.035 (0.027)	0.052 (0.039)	0.086 (0.043)
	Small	0.031 (0.026)	0.035 (0.028)	0.064 (0.041)	0.096 (0.055)
	None	-- --	0.034 (0.024)	0.074 (0.040)	0.092 (0.040)
30%	Large	0.047 (0.033)	0.032 (0.032)	0.042 (0.032)	0.048 (0.034)
	Medium	0.046 (0.030)	0.029 (0.025)	0.043 (0.037)	0.057 (0.031)
	Small	0.042 (0.029)	0.038 (0.026)	0.048 (0.032)	0.053 (0.034)
	None	-- --	0.031 (0.024)	0.037 (0.031)	0.063 (0.040)
15%	Large	0.043 (0.028)	0.035 (0.021)	0.028 (0.022)	0.040 (0.026)
	Medium	0.045 (0.025)	0.037 (0.025)	0.036 (0.026)	0.037 (0.022)
	Small	0.040 (0.030)	0.030 (0.028)	0.037 (0.027)	0.030 (0.022)
	None	0.023 (0.017)	0.032 (0.020)	0.036 (0.024)	0.033 (0.019)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.033 (0.021)	0.022 (0.017)	0.039 (0.020)	0.076 (0.023)
	Medium	0.031 (0.022)	0.022 (0.018)	0.046 (0.027)	0.080 (0.028)
	Small	0.025 (0.019)	0.024 (0.020)	0.059 (0.022)	0.092 (0.025)
	None	-- --	0.026 (0.015)	0.064 (0.029)	0.096 (0.021)
30%	Large	0.037 (0.021)	0.018 (0.017)	0.031 (0.019)	0.037 (0.024)
	Medium	0.037 (0.023)	0.019 (0.016)	0.029 (0.025)	0.047 (0.018)
	Small	0.030 (0.018)	0.022 (0.014)	0.033 (0.024)	0.050 (0.025)
	None	-- --	0.020 (0.015)	0.029 (0.021)	0.059 (0.018)
15%	Large	0.031 (0.019)	0.023 (0.014)	0.018 (0.014)	0.023 (0.016)
	Medium	0.034 (0.016)	0.020 (0.013)	0.020 (0.015)	0.020 (0.010)
	Small	0.028 (0.020)	0.020 (0.019)	0.021 (0.015)	0.021 (0.011)
	None	0.015 (0.013)	0.020 (0.014)	0.023 (0.015)	0.021 (0.015)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.050 (0.028)	0.044 (0.031)	0.091 (0.039)	0.150 (0.039)
	Medium	0.040 (0.028)	0.053 (0.021)	0.103 (0.031)	0.165 (0.035)
	Small	0.032 (0.022)	0.051 (0.033)	0.115 (0.026)	0.181 (0.037)
	None	-- --	0.058 (0.028)	0.136 (0.038)	0.200 (0.022)
30%	Large	0.042 (0.020)	0.021 (0.022)	0.045 (0.025)	0.059 (0.032)
	Medium	0.043 (0.027)	0.030 (0.020)	0.041 (0.031)	0.075 (0.029)
	Small	0.032 (0.023)	0.027 (0.019)	0.051 (0.028)	0.084 (0.031)
	None	-- --	0.025 (0.022)	0.051 (0.025)	0.093 (0.023)
15%	Large	0.028 (0.020)	0.022 (0.015)	0.019 (0.017)	0.024 (0.016)
	Medium	0.034 (0.018)	0.020 (0.012)	0.022 (0.016)	0.024 (0.014)
	Small	0.030 (0.020)	0.021 (0.019)	0.021 (0.015)	0.027 (0.013)
	None	0.018 (0.014)	0.018 (0.012)	0.025 (0.018)	0.029 (0.017)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.060 (0.040)	0.089 (0.040)	0.143 (0.056)	0.206 (0.045)
	Medium	0.031 (0.025)	0.082 (0.051)	0.136 (0.048)	0.166 (0.057)
	Small	0.025 (0.025)	0.080 (0.050)	0.131 (0.040)	0.167 (0.038)
	None	-- --	0.081 (0.055)	0.130 (0.041)	0.186 (0.034)
30%	Large	0.029 (0.025)	0.072 (0.038)	0.096 (0.043)	0.147 (0.057)
	Medium	0.046 (0.030)	0.063 (0.042)	0.096 (0.041)	0.133 (0.054)
	Small	0.045 (0.043)	0.047 (0.034)	0.085 (0.064)	0.150 (0.049)
	None	-- --	0.067 (0.049)	0.105 (0.048)	0.131 (0.062)
15%	Large	0.034 (0.026)	0.057 (0.037)	0.060 (0.047)	0.065 (0.041)
	Medium	0.047 (0.040)	0.052 (0.035)	0.068 (0.041)	0.098 (0.059)
	Small	0.037 (0.029)	0.041 (0.024)	0.045 (0.034)	0.065 (0.047)
	None	0.034 (0.035)	0.042 (0.038)	0.052 (0.055)	0.082 (0.056)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.082 (0.048)	0.111 (0.049)	0.155 (0.057)	0.213 (0.049)
	Medium	0.044 (0.027)	0.100 (0.053)	0.159 (0.059)	0.191 (0.063)
	Small	0.059 (0.035)	0.101 (0.049)	0.147 (0.051)	0.179 (0.043)
	None	-- --	0.105 (0.072)	0.162 (0.040)	0.221 (0.032)
30%	Large	0.040 (0.032)	0.092 (0.048)	0.122 (0.047)	0.163 (0.057)
	Medium	0.048 (0.046)	0.090 (0.047)	0.124 (0.052)	0.161 (0.056)
	Small	0.064 (0.053)	0.065 (0.031)	0.102 (0.064)	0.176 (0.061)
	None	-- --	0.098 (0.059)	0.143 (0.043)	0.169 (0.065)
15%	Large	0.044 (0.033)	0.103 (0.036)	0.082 (0.038)	0.083 (0.047)
	Medium	0.070 (0.057)	0.075 (0.040)	0.094 (0.046)	0.132 (0.062)
	Small	0.046 (0.038)	0.062 (0.035)	0.075 (0.042)	0.082 (0.053)
	None	0.056 (0.036)	0.069 (0.049)	0.092 (0.060)	0.104 (0.055)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.084 (0.046)	0.172 (0.060)	0.262 (0.063)	0.366 (0.054)
	Medium	0.054 (0.022)	0.168 (0.058)	0.266 (0.075)	0.327 (0.065)
	Small	0.061 (0.025)	0.164 (0.062)	0.239 (0.049)	0.308 (0.046)
	None	-- --	0.165 (0.066)	0.253 (0.047)	0.343 (0.042)
30%	Large	0.042 (0.029)	0.112 (0.049)	0.148 (0.052)	0.206 (0.056)
	Medium	0.051 (0.047)	0.110 (0.050)	0.160 (0.057)	0.210 (0.063)
	Small	0.066 (0.048)	0.083 (0.039)	0.133 (0.061)	0.229 (0.062)
	None	-- --	0.118 (0.062)	0.180 (0.047)	0.205 (0.058)
15%	Large	0.047 (0.037)	0.108 (0.036)	0.089 (0.039)	0.093 (0.048)
	Medium	0.072 (0.058)	0.078 (0.042)	0.103 (0.051)	0.143 (0.064)
	Small	0.046 (0.036)	0.072 (0.036)	0.082 (0.046)	0.092 (0.053)
	None	0.055 (0.039)	0.074 (0.052)	0.100 (0.065)	0.116 (0.057)

APPENDIX I

SLOPE VARIANCE PARAMETER BIAS ESTIMATES

First-Order Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.178 (0.098)	-0.207 (0.142)	-0.204 (0.155)	-0.155 (0.187)
	Medium	-0.102 (0.169)	-0.188 (0.124)	-0.179 (0.155)	-0.162 (0.175)
	Small	-0.154 (0.135)	-0.126 (0.161)	-0.136 (0.171)	-0.116 (0.199)
	None	-- --	-0.112 (0.163)	-0.098 (0.249)	-0.138 (0.163)
30%	Large	-0.189 (0.138)	-0.184 (0.156)	-0.123 (0.143)	-0.167 (0.165)
	Medium	-0.124 (0.183)	-0.125 (0.139)	-0.138 (0.192)	-0.098 (0.133)
	Small	-0.129 (0.162)	-0.085 (0.180)	-0.092 (0.180)	-0.136 (0.172)
	None	-- --	-0.100 (0.163)	-0.136 (0.193)	-0.090 (0.206)
15%	Large	-0.149 (0.178)	-0.074 (0.184)	-0.098 (0.164)	-0.188 (0.162)
	Medium	-0.132 (0.137)	-0.145 (0.145)	-0.058 (0.180)	-0.140 (0.164)
	Small	-0.134 (0.147)	-0.070 (0.228)	-0.077 (0.207)	-0.137 (0.121)
	None	-0.063 (0.139)	-0.155 (0.145)	-0.076 (0.175)	-0.068 (0.158)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.265 (0.103)	-0.260 (0.100)	-0.250 (0.101)	-0.232 (0.096)
	Medium	-0.186 (0.091)	-0.232 (0.121)	-0.226 (0.088)	-0.216 (0.115)
	Small	-0.236 (0.104)	-0.196 (0.107)	-0.224 (0.095)	-0.198 (0.076)
	None	-- --	-0.173 (0.097)	-0.176 (0.116)	-0.188 (0.106)
30%	Large	-0.226 (0.116)	-0.227 (0.095)	-0.200 (0.127)	-0.180 (0.105)
	Medium	-0.208 (0.116)	-0.190 (0.106)	-0.211 (0.103)	-0.197 (0.104)
	Small	-0.199 (0.100)	-0.200 (0.110)	-0.200 (0.096)	-0.151 (0.103)
	None	-- --	-0.152 (0.115)	-0.190 (0.093)	-0.179 (0.118)
15%	Large	-0.211 (0.128)	-0.165 (0.140)	-0.185 (0.127)	-0.220 (0.111)
	Medium	-0.190 (0.087)	-0.178 (0.104)	-0.169 (0.097)	-0.224 (0.091)
	Small	-0.182 (0.106)	-0.165 (0.149)	-0.178 (0.112)	-0.205 (0.099)
	None	-0.151 (0.103)	-0.201 (0.106)	-0.180 (0.104)	-0.149 (0.120)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.325 (0.101)	-0.319 (0.099)	-0.313 (0.096)	-0.309 (0.086)
	Medium	-0.244 (0.089)	-0.284 (0.111)	-0.285 (0.083)	-0.268 (0.104)
	Small	-0.279 (0.097)	-0.233 (0.099)	-0.264 (0.092)	-0.235 (0.073)
	None	-- --	-0.186 (0.098)	-0.185 (0.114)	-0.201 (0.104)
30%	Large	-0.251 (0.113)	-0.256 (0.092)	-0.226 (0.122)	-0.210 (0.102)
	Medium	-0.226 (0.116)	-0.210 (0.106)	-0.232 (0.105)	-0.219 (0.105)
	Small	-0.215 (0.096)	-0.217 (0.109)	-0.214 (0.094)	-0.171 (0.104)
	None	-- --	-0.161 (0.109)	-0.197 (0.092)	-0.184 (0.120)
15%	Large	-0.215 (0.128)	-0.172 (0.138)	-0.192 (0.126)	-0.229 (0.112)
	Medium	-0.197 (0.087)	-0.184 (0.103)	-0.175 (0.096)	-0.229 (0.092)
	Small	-0.189 (0.105)	-0.169 (0.147)	-0.184 (0.112)	-0.208 (0.097)
	None	-0.153 (0.106)	-0.202 (0.106)	-0.180 (0.103)	-0.151 (0.118)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	1.372 (0.919)	0.776 (0.943)	1.062 (0.827)	1.282 (1.178)
	Medium	1.080 (1.393)	0.858 (1.004)	1.606 (1.200)	0.887 (0.879)
	Small	1.191 (1.328)	0.794 (0.837)	0.596 (0.656)	0.895 (1.183)
	None	-- --	1.207 (1.109)	1.087 (1.110)	0.854 (0.823)
30%	Large	0.957 (0.785)	1.234 (0.891)	1.024 (0.905)	0.953 (0.873)
	Medium	0.761 (0.746)	1.249 (1.231)	1.081 (0.799)	1.370 (1.286)
	Small	1.219 (1.103)	1.039 (0.965)	1.119 (1.361)	1.693 (1.170)
	None	-- --	1.196 (1.210)	1.483 (0.963)	1.443 (1.096)
15%	Large	1.004 (1.037)	1.193 (0.959)	1.109 (1.252)	1.400 (1.094)
	Medium	1.743 (1.083)	1.312 (0.848)	0.700 (0.771)	1.443 (0.980)
	Small	1.545 (1.526)	1.223 (1.142)	0.948 (1.083)	0.374 (0.696)
	None	1.118 (1.108)	0.883 (1.070)	1.104 (1.189)	1.689 (1.251)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	1.565 (0.976)	1.077 (1.103)	0.885 (0.933)	1.221 (1.296)
	Medium	1.510 (1.861)	0.890 (1.126)	1.297 (1.242)	0.703 (1.109)
	Small	1.176 (1.531)	0.728 (0.611)	0.400 (0.836)	1.001 (1.410)
	None	-- --	1.033 (1.331)	0.995 (1.094)	0.865 (0.811)
30%	Large	0.867 (0.630)	1.216 (1.156)	0.879 (0.926)	1.162 (0.751)
	Medium	0.772 (0.675)	1.195 (1.254)	1.337 (0.529)	1.355 (1.385)
	Small	1.231 (1.061)	1.049 (1.090)	1.115 (1.412)	1.776 (1.339)
	None	-- --	1.295 (1.297)	1.590 (0.900)	1.489 (1.109)
15%	Large	1.095 (1.078)	1.365 (1.152)	0.899 (1.379)	1.323 (1.142)
	Medium	1.905 (1.107)	1.287 (0.971)	0.915 (1.078)	1.502 (1.075)
	Small	1.724 (1.581)	1.272 (1.440)	0.963 (1.262)	0.298 (0.787)
	None	1.008 (1.179)	0.675 (1.220)	1.138 (1.201)	1.570 (1.490)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	1.468 (0.947)	1.143 (1.087)	0.865 (0.922)	1.193 (1.216)
	Medium	1.438 (1.767)	0.879 (1.110)	1.326 (1.275)	0.722 (1.093)
	Small	1.049 (1.381)	0.724 (0.589)	0.423 (0.889)	1.031 (1.427)
	None	-- --	1.069 (1.279)	1.092 (1.083)	0.866 (0.819)
30%	Large	0.871 (0.636)	1.213 (1.147)	0.943 (0.957)	1.200 (0.753)
	Medium	0.765 (0.658)	1.165 (1.209)	1.333 (0.491)	1.328 (1.414)
	Small	1.235 (1.032)	1.058 (1.095)	1.118 (1.403)	1.731 (1.296)
	None	-- --	1.262 (1.279)	1.586 (0.908)	1.362 (1.022)
15%	Large	1.108 (1.090)	1.359 (1.132)	0.903 (1.347)	1.305 (1.134)
	Medium	1.906 (1.098)	1.270 (0.943)	0.915 (1.087)	1.502 (1.074)
	Small	1.731 (1.573)	1.332 (1.397)	0.960 (1.260)	0.291 (0.786)
	None	0.995 (1.142)	0.700 (1.237)	1.152 (1.224)	1.582 (1.507)

APPENDIX J

SLOPE VARIANCE PARAMETER ROOT MEAN SQUARE ERROR ESTIMATES

First-Order Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.036 (0.018)	0.042 (0.027)	0.042 (0.029)	0.040 (0.027)
	Medium	0.033 (0.022)	0.039 (0.023)	0.042 (0.022)	0.039 (0.026)
	Small	0.036 (0.020)	0.036 (0.019)	0.033 (0.028)	0.036 (0.029)
	None	-- --	0.032 (0.023)	0.044 (0.029)	0.035 (0.025)
30%	Large	0.038 (0.027)	0.043 (0.021)	0.030 (0.023)	0.038 (0.027)
	Medium	0.035 (0.027)	0.031 (0.021)	0.039 (0.025)	0.027 (0.019)
	Small	0.035 (0.021)	0.034 (0.020)	0.032 (0.024)	0.037 (0.024)
	None	-- --	0.030 (0.024)	0.039 (0.027)	0.038 (0.023)
15%	Large	0.042 (0.018)	0.032 (0.023)	0.031 (0.023)	0.042 (0.026)
	Medium	0.033 (0.018)	0.034 (0.022)	0.030 (0.023)	0.034 (0.027)
	Small	0.035 (0.019)	0.035 (0.032)	0.036 (0.024)	0.030 (0.021)
	None	0.026 (0.016)	0.038 (0.019)	0.032 (0.021)	0.029 (0.019)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.053 (0.021)	0.052 (0.020)	0.050 (0.020)	0.046 (0.019)
	Medium	0.037 (0.018)	0.047 (0.023)	0.045 (0.018)	0.045 (0.019)
	Small	0.047 (0.021)	0.039 (0.021)	0.045 (0.018)	0.040 (0.015)
	None	-- --	0.035 (0.018)	0.038 (0.018)	0.038 (0.021)
30%	Large	0.046 (0.022)	0.045 (0.019)	0.041 (0.023)	0.036 (0.021)
	Medium	0.042 (0.023)	0.039 (0.020)	0.042 (0.021)	0.040 (0.020)
	Small	0.040 (0.020)	0.040 (0.022)	0.040 (0.019)	0.032 (0.018)
	None	-- --	0.032 (0.021)	0.038 (0.019)	0.039 (0.018)
15%	Large	0.045 (0.020)	0.038 (0.021)	0.040 (0.021)	0.045 (0.020)
	Medium	0.038 (0.017)	0.036 (0.020)	0.035 (0.018)	0.045 (0.018)
	Small	0.038 (0.019)	0.039 (0.022)	0.036 (0.021)	0.041 (0.020)
	None	0.032 (0.017)	0.040 (0.021)	0.036 (0.021)	0.032 (0.021)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.065	0.064	0.063	0.062
		(0.020)	(0.020)	(0.019)	(0.017)
	Medium	0.049	0.057	0.057	0.054
		(0.018)	(0.022)	(0.017)	(0.021)
	Small	0.056	0.047	0.053	0.047
		(0.019)	(0.020)	(0.018)	(0.015)
	None	--	0.038	0.040	0.040
		--	(0.019)	(0.018)	(0.021)
30%	Large	0.050	0.051	0.046	0.042
		(0.022)	(0.018)	(0.022)	(0.020)
	Medium	0.045	0.042	0.046	0.044
		(0.023)	(0.020)	(0.021)	(0.021)
	Small	0.043	0.043	0.043	0.035
		(0.019)	(0.022)	(0.019)	(0.019)
	None	--	0.033	0.039	0.040
		--	(0.021)	(0.018)	(0.018)
15%	Large	0.046	0.039	0.041	0.047
		(0.020)	(0.021)	(0.021)	(0.020)
	Medium	0.039	0.037	0.035	0.046
		(0.017)	(0.021)	(0.018)	(0.018)
	Small	0.039	0.039	0.038	0.042
		(0.019)	(0.021)	(0.021)	(0.019)
	None	0.033	0.040	0.036	0.032
		(0.018)	(0.021)	(0.021)	(0.021)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.274 (0.184)	0.175 (0.169)	0.215 (0.161)	0.266 (0.224)
	Medium	0.221 (0.274)	0.198 (0.172)	0.339 (0.213)	0.180 (0.173)
	Small	0.254 (0.249)	0.172 (0.153)	0.136 (0.113)	0.192 (0.225)
	None	-- --	0.241 (0.222)	0.217 (0.222)	0.179 (0.155)
30%	Large	0.214 (0.121)	0.253 (0.168)	0.206 (0.179)	0.207 (0.153)
	Medium	0.173 (0.122)	0.253 (0.243)	0.232 (0.133)	0.274 (0.257)
	Small	0.244 (0.221)	0.208 (0.193)	0.241 (0.256)	0.339 (0.234)
	None	-- --	0.256 (0.223)	0.297 (0.193)	0.289 (0.219)
15%	Large	0.209 (0.199)	0.240 (0.191)	0.251 (0.218)	0.285 (0.212)
	Medium	0.349 (0.217)	0.262 (0.170)	0.147 (0.146)	0.299 (0.179)
	Small	0.329 (0.281)	0.245 (0.228)	0.209 (0.197)	0.095 (0.126)
	None	0.229 (0.216)	0.212 (0.176)	0.235 (0.222)	0.341 (0.245)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.313 (0.195)	0.231 (0.202)	0.185 (0.178)	0.250 (0.253)
	Medium	0.303 (0.371)	0.231 (0.165)	0.288 (0.212)	0.175 (0.194)
	Small	0.270 (0.271)	0.146 (0.121)	0.122 (0.136)	0.226 (0.258)
	None	-- --	0.234 (0.241)	0.202 (0.215)	0.185 (0.147)
30%	Large	0.193 (0.089)	0.255 (0.216)	0.190 (0.169)	0.232 (0.150)
	Medium	0.175 (0.103)	0.244 (0.246)	0.267 (0.106)	0.272 (0.276)
	Small	0.246 (0.212)	0.219 (0.208)	0.245 (0.261)	0.355 (0.268)
	None	-- --	0.275 (0.241)	0.318 (0.180)	0.302 (0.216)
15%	Large	0.231 (0.201)	0.274 (0.229)	0.245 (0.212)	0.272 (0.219)
	Medium	0.383 (0.218)	0.257 (0.194)	0.196 (0.201)	0.313 (0.194)
	Small	0.353 (0.306)	0.258 (0.284)	0.236 (0.208)	0.106 (0.129)
	None	0.207 (0.231)	0.201 (0.188)	0.240 (0.226)	0.332 (0.276)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.294 (0.189)	0.244 (0.196)	0.181 (0.176)	0.241 (0.240)
	Medium	0.288 (0.353)	0.225 (0.167)	0.291 (0.222)	0.178 (0.190)
	Small	0.228 (0.259)	0.145 (0.118)	0.128 (0.146)	0.233 (0.260)
	None	-- --	0.237 (0.233)	0.218 (0.217)	0.183 (0.152)
30%	Large	0.192 (0.094)	0.254 (0.215)	0.203 (0.174)	0.240 (0.151)
	Medium	0.173 (0.102)	0.239 (0.236)	0.267 (0.098)	0.267 (0.282)
	Small	0.247 (0.206)	0.215 (0.215)	0.245 (0.260)	0.346 (0.259)
	None	-- --	0.268 (0.239)	0.317 (0.182)	0.272 (0.204)
15%	Large	0.236 (0.201)	0.272 (0.226)	0.244 (0.206)	0.268 (0.217)
	Medium	0.381 (0.220)	0.254 (0.189)	0.197 (0.202)	0.313 (0.194)
	Small	0.354 (0.305)	0.272 (0.273)	0.234 (0.210)	0.105 (0.128)
	None	0.205 (0.222)	0.203 (0.194)	0.243 (0.231)	0.334 (0.280)

APPENDIX K

INTERCEPT-SLOPE COVARIANCE PARAMETER *P*-VALUE

First-Order Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.185 (0.265)	0.137 (0.236)	0.087 (0.139)	0.076 (0.203)
	Medium	0.137 (0.224)	0.173 (0.265)	0.069 (0.119)	0.072 (0.136)
	Small	0.272 (0.297)	0.140 (0.259)	0.134 (0.179)	0.045 (0.055)
	None	-- --	0.186 (0.223)	0.111 (0.152)	0.110 (0.228)
30%	Large	0.149 (0.227)	0.160 (0.246)	0.088 (0.137)	0.061 (0.104)
	Medium	0.235 (0.301)	0.173 (0.242)	0.248 (0.349)	0.143 (0.199)
	Small	0.233 (0.289)	0.240 (0.333)	0.127 (0.179)	0.079 (0.160)
	None	-- --	0.094 (0.108)	0.178 (0.266)	0.165 (0.222)
15%	Large	0.170 (0.252)	0.162 (0.239)	0.154 (0.210)	0.172 (0.226)
	Medium	0.222 (0.260)	0.195 (0.291)	0.174 (0.216)	0.198 (0.282)
	Small	0.211 (0.284)	0.202 (0.250)	0.177 (0.221)	0.168 (0.239)
	None	0.161 (0.198)	0.179 (0.247)	0.198 (0.243)	0.185 (0.243)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.020 (0.030)	0.018 (0.026)	0.028 (0.062)	0.013 (0.029)
	Medium	0.009 (0.020)	0.025 (0.042)	0.023 (0.047)	0.034 (0.098)
	Small	0.072 (0.122)	0.041 (0.082)	0.032 (0.044)	0.018 (0.031)
	None	-- --	0.060 (0.122)	0.065 (0.110)	0.075 (0.160)
30%	Large	0.037 (0.075)	0.027 (0.051)	0.040 (0.102)	0.017 (0.035)
	Medium	0.052 (0.127)	0.020 (0.030)	0.069 (0.140)	0.045 (0.117)
	Small	0.061 (0.137)	0.034 (0.053)	0.041 (0.103)	0.021 (0.040)
	None	-- --	0.040 (0.103)	0.068 (0.130)	0.047 (0.065)
15%	Large	0.053 (0.101)	0.041 (0.079)	0.031 (0.055)	0.050 (0.093)
	Medium	0.032 (0.042)	0.042 (0.072)	0.022 (0.035)	0.083 (0.141)
	Small	0.033 (0.059)	0.041 (0.076)	0.068 (0.166)	0.073 (0.198)
	None	0.022 (0.040)	0.082 (0.169)	0.093 (0.216)	0.038 (0.055)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.006 (0.009)	0.005 (0.008)	0.009 (0.019)	0.006 (0.019)
	Medium	0.005 (0.011)	0.013 (0.025)	0.011 (0.023)	0.020 (0.067)
	Small	0.055 (0.094)	0.034 (0.071)	0.024 (0.037)	0.011 (0.019)
	None	-- --	0.067 (0.128)	0.065 (0.113)	0.073 (0.143)
30%	Large	0.024 (0.048)	0.021 (0.042)	0.032 (0.086)	0.011 (0.024)
	Medium	0.045 (0.113)	0.016 (0.023)	0.064 (0.130)	0.038 (0.102)
	Small	0.057 (0.129)	0.031 (0.049)	0.037 (0.095)	0.020 (0.041)
	None	-- --	0.043 (0.111)	0.072 (0.138)	0.046 (0.066)
15%	Large	0.052 (0.099)	0.040 (0.078)	0.030 (0.054)	0.048 (0.089)
	Medium	0.031 (0.040)	0.041 (0.072)	0.021 (0.033)	0.079 (0.133)
	Small	0.031 (0.055)	0.041 (0.073)	0.069 (0.174)	0.071 (0.192)
	None	0.022 (0.041)	0.085 (0.177)	0.095 (0.224)	0.035 (0.051)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.553 (0.289)	0.591 (0.324)	0.574 (0.157)	0.600 (0.242)
	Medium	0.622 (0.314)	0.603 (0.303)	0.524 (0.256)	0.376 (0.176)
	Small	0.503 (0.333)	0.580 (0.251)	0.588 (0.286)	0.485 (0.328)
	None	-- --	0.562 (0.282)	0.518 (0.296)	0.594 (0.244)
30%	Large	0.642 (0.206)	0.502 (0.235)	0.577 (0.156)	0.619 (0.213)
	Medium	0.553 (0.271)	0.601 (0.269)	0.495 (0.293)	0.519 (0.324)
	Small	0.683 (0.305)	0.586 (0.209)	0.534 (0.241)	0.584 (0.291)
	None	-- --	0.451 (0.205)	0.541 (0.247)	0.583 (0.240)
15%	Large	0.472 (0.271)	0.496 (0.285)	0.535 (0.292)	0.510 (0.227)
	Medium	0.520 (0.245)	0.573 (0.274)	0.586 (0.278)	0.557 (0.273)
	Small	0.560 (0.272)	0.524 (0.229)	0.539 (0.329)	0.515 (0.220)
	None	0.536 (0.268)	0.556 (0.192)	0.585 (0.279)	0.480 (0.233)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.405 (0.260)	0.454 (0.248)	0.575 (0.305)	0.527 (0.293)
	Medium	0.466 (0.302)	0.539 (0.271)	0.494 (0.290)	0.527 (0.271)
	Small	0.455 (0.351)	0.677 (0.260)	0.674 (0.239)	0.502 (0.339)
	None	-- --	0.478 (0.310)	0.614 (0.289)	0.561 (0.279)
30%	Large	0.583 (0.202)	0.497 (0.311)	0.545 (0.286)	0.539 (0.243)
	Medium	0.546 (0.265)	0.599 (0.319)	0.515 (0.279)	0.417 (0.270)
	Small	0.591 (0.265)	0.632 (0.320)	0.530 (0.293)	0.519 (0.333)
	None	-- --	0.532 (0.315)	0.400 (0.211)	0.462 (0.256)
15%	Large	0.552 (0.316)	0.447 (0.337)	0.522 (0.275)	0.497 (0.286)
	Medium	0.379 (0.279)	0.554 (0.302)	0.552 (0.226)	0.348 (0.139)
	Small	0.458 (0.237)	0.552 (0.328)	0.458 (0.268)	0.645 (0.245)
	None	0.628 (0.326)	0.575 (0.250)	0.538 (0.294)	0.477 (0.324)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.386 (0.225)	0.415 (0.285)	0.555 (0.310)	0.478 (0.291)
	Medium	0.427 (0.274)	0.541 (0.284)	0.464 (0.290)	0.538 (0.283)
	Small	0.447 (0.289)	0.673 (0.272)	0.671 (0.260)	0.506 (0.367)
	None	-- --	0.477 (0.296)	0.586 (0.287)	0.563 (0.274)
30%	Large	0.577 (0.222)	0.495 (0.320)	0.494 (0.275)	0.497 (0.229)
	Medium	0.540 (0.254)	0.587 (0.304)	0.508 (0.278)	0.427 (0.294)
	Small	0.585 (0.267)	0.626 (0.332)	0.519 (0.287)	0.512 (0.330)
	None	-- --	0.539 (0.310)	0.402 (0.212)	0.492 (0.262)
15%	Large	0.546 (0.318)	0.442 (0.335)	0.523 (0.273)	0.506 (0.294)
	Medium	0.386 (0.293)	0.550 (0.296)	0.547 (0.219)	0.349 (0.136)
	Small	0.459 (0.248)	0.527 (0.328)	0.458 (0.261)	0.638 (0.246)
	None	0.626 (0.324)	0.567 (0.249)	0.539 (0.300)	0.473 (0.316)

APPENDIX L

INTERCEPT-SLOPE COVARIANCE PARAMETER ESTIMATE

First-Order Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.050 (0.036)	-0.063 (0.035)	-0.064 (0.028)	-0.071 (0.029)
	Medium	-0.056 (0.033)	-0.058 (0.034)	-0.063 (0.029)	-0.070 (0.034)
	Small	-0.040 (0.026)	-0.061 (0.034)	-0.056 (0.031)	-0.069 (0.025)
	None	-- --	-0.052 (0.030)	-0.057 (0.025)	-0.070 (0.033)
30%	Large	-0.054 (0.033)	-0.056 (0.030)	-0.062 (0.042)	-0.075 (0.031)
	Medium	-0.047 (0.032)	-0.054 (0.030)	-0.052 (0.035)	-0.054 (0.034)
	Small	-0.044 (0.033)	-0.047 (0.037)	-0.056 (0.027)	-0.074 (0.032)
	None	-- --	-0.060 (0.027)	-0.053 (0.028)	-0.054 (0.030)
15%	Large	-0.048 (0.039)	-0.058 (0.038)	-0.056 (0.035)	-0.051 (0.031)
	Medium	-0.046 (0.032)	-0.051 (0.034)	-0.054 (0.031)	-0.042 (0.027)
	Small	-0.051 (0.034)	-0.048 (0.034)	-0.048 (0.028)	-0.051 (0.030)
	None	-0.053 (0.029)	-0.049 (0.032)	-0.049 (0.037)	-0.056 (0.032)

Second-Order Constrained Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.087 (0.029)	-0.088 (0.032)	-0.088 (0.031)	-0.095 (0.029)
	Medium	-0.096 (0.024)	-0.083 (0.032)	-0.086 (0.025)	-0.088 (0.032)
	Small	-0.071 (0.030)	-0.086 (0.031)	-0.078 (0.026)	-0.083 (0.023)
	None	-- --	-0.076 (0.029)	-0.074 (0.033)	-0.073 (0.029)
30%	Large	-0.086 (0.033)	-0.084 (0.030)	-0.094 (0.038)	-0.097 (0.032)
	Medium	-0.083 (0.033)	-0.086 (0.027)	-0.078 (0.032)	-0.081 (0.031)
	Small	-0.077 (0.030)	-0.080 (0.027)	-0.080 (0.029)	-0.092 (0.031)
	None	-- --	-0.086 (0.031)	-0.071 (0.026)	-0.074 (0.028)
15%	Large	-0.081 (0.037)	-0.091 (0.041)	-0.088 (0.036)	-0.078 (0.033)
	Medium	-0.077 (0.026)	-0.080 (0.028)	-0.086 (0.029)	-0.069 (0.029)
	Small	-0.082 (0.030)	-0.084 (0.045)	-0.080 (0.033)	-0.074 (0.027)
	None	-0.088 (0.030)	-0.074 (0.030)	-0.077 (0.035)	-0.084 (0.033)

Second-Order Free Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.096 (0.028)	-0.098 (0.032)	-0.096 (0.028)	-0.104 (0.029)
	Medium	-0.101 (0.024)	-0.088 (0.031)	-0.090 (0.024)	-0.093 (0.031)
	Small	-0.073 (0.029)	-0.088 (0.030)	-0.080 (0.026)	-0.086 (0.022)
	None	-- --	-0.073 (0.028)	-0.072 (0.032)	-0.072 (0.029)
30%	Large	-0.089 (0.033)	-0.086 (0.030)	-0.096 (0.038)	-0.099 (0.031)
	Medium	-0.085 (0.033)	-0.088 (0.028)	-0.079 (0.032)	-0.082 (0.031)
	Small	-0.077 (0.029)	-0.080 (0.026)	-0.081 (0.028)	-0.092 (0.031)
	None	-- --	-0.084 (0.030)	-0.070 (0.026)	-0.073 (0.028)
15%	Large	-0.081 (0.037)	-0.091 (0.041)	-0.088 (0.036)	-0.078 (0.033)
	Medium	-0.077 (0.026)	-0.080 (0.028)	-0.086 (0.029)	-0.070 (0.029)
	Small	-0.082 (0.030)	-0.084 (0.045)	-0.079 (0.033)	-0.074 (0.026)
	None	-0.088 (0.030)	-0.074 (0.030)	-0.077 (0.035)	-0.084 (0.033)

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.096 (0.161)	-0.025 (0.160)	-0.029 (0.115)	-0.084 (0.151)
	Medium	-0.074 (0.234)	-0.017 (0.141)	-0.081 (0.177)	0.034 (0.183)
	Small	-0.091 (0.197)	-0.001 (0.145)	0.045 (0.122)	-0.011 (0.191)
	None	-- --	-0.045 (0.163)	0.019 (0.159)	0.004 (0.145)
30%	Large	-0.059 (0.098)	-0.070 (0.153)	-0.020 (0.115)	-0.037 (0.103)
	Medium	-0.023 (0.171)	-0.050 (0.165)	-0.006 (0.163)	-0.068 (0.200)
	Small	-0.029 (0.174)	0.009 (0.128)	-0.027 (0.174)	-0.097 (0.165)
	None	-- --	-0.029 (0.171)	-0.076 (0.141)	-0.078 (0.131)
15%	Large	-0.016 (0.165)	-0.080 (0.177)	-0.026 (0.176)	-0.060 (0.156)
	Medium	-0.129 (0.147)	-0.072 (0.150)	0.017 (0.152)	-0.073 (0.161)
	Small	-0.057 (0.225)	-0.036 (0.177)	-0.012 (0.170)	0.064 (0.111)
	None	-0.077 (0.204)	-0.008 (0.128)	-0.039 (0.163)	-0.108 (0.169)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.222	-0.177	-0.127	-0.175
		(0.184)	(0.171)	(0.140)	(0.203)
	Medium	-0.243	-0.108	-0.162	-0.063
		(0.264)	(0.164)	(0.184)	(0.222)
	Small	-0.177	-0.085	-0.018	-0.156
		(0.230)	(0.114)	(0.106)	(0.210)
	None	--	-0.123	-0.097	-0.110
		--	(0.218)	(0.173)	(0.167)
30%	Large	-0.110	-0.143	-0.128	-0.140
		(0.093)	(0.173)	(0.125)	(0.088)
	Medium	-0.132	-0.148	-0.158	-0.198
		(0.160)	(0.194)	(0.122)	(0.245)
	Small	-0.130	-0.114	-0.144	-0.209
		(0.179)	(0.122)	(0.221)	(0.231)
	None	--	-0.151	-0.206	-0.182
		--	(0.182)	(0.126)	(0.151)
15%	Large	-0.118	-0.204	-0.109	-0.136
		(0.166)	(0.216)	(0.187)	(0.162)
	Medium	-0.240	-0.155	-0.105	-0.201
		(0.175)	(0.146)	(0.190)	(0.184)
	Small	-0.195	-0.132	-0.110	-0.028
		(0.256)	(0.210)	(0.190)	(0.125)
	None	-0.144	-0.057	-0.149	-0.201
		(0.212)	(0.158)	(0.193)	(0.196)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.220 (0.179)	-0.205 (0.166)	-0.137 (0.138)	-0.194 (0.195)
	Medium	-0.251 (0.250)	-0.124 (0.163)	-0.185 (0.191)	-0.079 (0.219)
	Small	-0.179 (0.204)	-0.093 (0.106)	-0.031 (0.111)	-0.173 (0.219)
	None	-- --	-0.128 (0.209)	-0.104 (0.178)	-0.111 (0.170)
30%	Large	-0.117 (0.094)	-0.153 (0.171)	-0.143 (0.131)	-0.154 (0.088)
	Medium	-0.135 (0.160)	-0.151 (0.186)	-0.162 (0.121)	-0.201 (0.247)
	Small	-0.132 (0.172)	-0.122 (0.125)	-0.149 (0.221)	-0.209 (0.221)
	None	-- --	-0.148 (0.180)	-0.206 (0.126)	-0.171 (0.152)
15%	Large	-0.121 (0.168)	-0.205 (0.216)	-0.111 (0.185)	-0.137 (0.163)
	Medium	-0.241 (0.173)	-0.154 (0.141)	-0.107 (0.190)	-0.202 (0.182)
	Small	-0.197 (0.254)	-0.143 (0.207)	-0.111 (0.190)	-0.026 (0.125)
	None	-0.141 (0.207)	-0.059 (0.162)	-0.150 (0.197)	-0.202 (0.199)

APPENDIX M

INTERCEPT-QUADRATIC SLOPE COVARIANCE PARAMETER *P*-VALUE

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.583 (0.257)	0.517 (0.296)	0.603 (0.299)	0.594 (0.279)
	Medium	0.511 (0.291)	0.442 (0.217)	0.520 (0.296)	0.277 (0.256)
	Small	0.477 (0.285)	0.433 (0.217)	0.475 (0.298)	0.406 (0.269)
	None	-- --	0.504 (0.280)	0.453 (0.305)	0.412 (0.203)
30%	Large	0.704 (0.218)	0.498 (0.240)	0.565 (0.245)	0.698 (0.289)
	Medium	0.557 (0.305)	0.536 (0.268)	0.539 (0.304)	0.454 (0.289)
	Small	0.573 (0.268)	0.518 (0.225)	0.505 (0.309)	0.588 (0.265)
	None	-- --	0.476 (0.316)	0.636 (0.283)	0.613 (0.280)
15%	Large	0.514 (0.310)	0.468 (0.227)	0.475 (0.261)	0.490 (0.283)
	Medium	0.534 (0.177)	0.629 (0.221)	0.447 (0.267)	0.606 (0.270)
	Small	0.563 (0.291)	0.481 (0.273)	0.478 (0.319)	0.443 (0.284)
	None	0.528 (0.279)	0.503 (0.271)	0.520 (0.251)	0.538 (0.231)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.557 (0.281)	0.624 (0.317)	0.574 (0.196)	0.583 (0.265)
	Medium	0.597 (0.311)	0.530 (0.248)	0.521 (0.261)	0.390 (0.254)
	Small	0.446 (0.333)	0.721 (0.215)	0.576 (0.214)	0.445 (0.272)
	None	-- --	0.427 (0.227)	0.603 (0.256)	0.623 (0.298)
30%	Large	0.768 (0.151)	0.462 (0.186)	0.595 (0.197)	0.722 (0.185)
	Medium	0.632 (0.236)	0.625 (0.260)	0.636 (0.223)	0.499 (0.327)
	Small	0.631 (0.273)	0.635 (0.212)	0.531 (0.255)	0.597 (0.278)
	None	-- --	0.545 (0.250)	0.517 (0.212)	0.579 (0.322)
15%	Large	0.569 (0.257)	0.423 (0.251)	0.577 (0.284)	0.492 (0.243)
	Medium	0.466 (0.243)	0.683 (0.280)	0.624 (0.387)	0.451 (0.255)
	Small	0.532 (0.259)	0.453 (0.216)	0.520 (0.314)	0.580 (0.217)
	None	0.654 (0.233)	0.497 (0.218)	0.566 (0.276)	0.540 (0.274)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.602 (0.272)	0.578 (0.297)	0.592 (0.203)	0.587 (0.262)
	Medium	0.598 (0.311)	0.540 (0.244)	0.529 (0.285)	0.418 (0.272)
	Small	0.528 (0.361)	0.735 (0.190)	0.561 (0.187)	0.434 (0.274)
	None	-- --	0.452 (0.227)	0.590 (0.259)	0.630 (0.312)
30%	Large	0.763 (0.144)	0.470 (0.182)	0.591 (0.196)	0.708 (0.193)
	Medium	0.653 (0.257)	0.636 (0.250)	0.650 (0.238)	0.500 (0.338)
	Small	0.631 (0.267)	0.632 (0.213)	0.530 (0.258)	0.594 (0.262)
	None	-- --	0.551 (0.244)	0.520 (0.210)	0.618 (0.330)
15%	Large	0.567 (0.261)	0.422 (0.257)	0.585 (0.288)	0.491 (0.234)
	Medium	0.471 (0.239)	0.684 (0.276)	0.618 (0.381)	0.451 (0.253)
	Small	0.524 (0.254)	0.451 (0.210)	0.516 (0.307)	0.580 (0.231)
	None	0.641 (0.230)	0.495 (0.224)	0.555 (0.267)	0.539 (0.274)

APPENDIX N

INTERCEPT-QUADRATIC SLOPE COVARIANCE PARAMETER ESTIMATE

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.003 (0.035)	-0.013 (0.036)	-0.015 (0.025)	-0.001 (0.034)
	Medium	0.001 (0.055)	-0.018 (0.034)	-0.001 (0.039)	-0.030 (0.041)
	Small	0.008 (0.046)	-0.016 (0.036)	-0.027 (0.027)	-0.017 (0.046)
	None	-- --	-0.004 (0.038)	-0.022 (0.037)	-0.021 (0.034)
30%	Large	-0.004 (0.021)	-0.002 (0.034)	-0.014 (0.027)	-0.011 (0.024)
	Medium	-0.011 (0.036)	-0.004 (0.037)	-0.015 (0.033)	-0.002 (0.048)
	Small	-0.005 (0.041)	-0.017 (0.029)	-0.011 (0.041)	-0.001 (0.036)
	None	-- --	-0.012 (0.039)	0.003 (0.031)	0.000 (0.033)
15%	Large	-0.012 (0.036)	0.004 (0.043)	-0.013 (0.038)	-0.002 (0.039)
	Medium	0.015 (0.036)	0.001 (0.031)	-0.018 (0.037)	0.005 (0.039)
	Small	0.002 (0.054)	-0.008 (0.041)	-0.012 (0.038)	-0.029 (0.023)
	None	0.000 (0.046)	-0.015 (0.031)	-0.004 (0.037)	0.007 (0.038)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.028 (0.040)	0.021 (0.039)	0.005 (0.032)	0.020 (0.045)
	Medium	0.037 (0.063)	0.002 (0.039)	0.018 (0.043)	-0.006 (0.051)
	Small	0.025 (0.055)	0.000 (0.024)	-0.013 (0.025)	0.019 (0.053)
	None	-- --	0.013 (0.050)	0.007 (0.039)	0.010 (0.040)
30%	Large	0.004 (0.018)	0.015 (0.041)	0.008 (0.031)	0.016 (0.017)
	Medium	0.007 (0.034)	0.014 (0.043)	0.017 (0.029)	0.026 (0.057)
	Small	0.015 (0.043)	0.007 (0.030)	0.014 (0.053)	0.027 (0.050)
	None	-- --	0.011 (0.041)	0.033 (0.029)	0.027 (0.040)
15%	Large	0.009 (0.037)	0.030 (0.053)	0.001 (0.041)	0.013 (0.041)
	Medium	0.038 (0.043)	0.018 (0.033)	0.005 (0.047)	0.031 (0.045)
	Small	0.028 (0.061)	0.012 (0.050)	0.008 (0.044)	-0.009 (0.028)
	None	0.013 (0.049)	-0.006 (0.037)	0.019 (0.042)	0.027 (0.048)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.024 (0.037)	0.024 (0.038)	0.004 (0.031)	0.022 (0.043)
	Medium	0.037 (0.059)	0.004 (0.037)	0.021 (0.044)	-0.005 (0.049)
	Small	0.024 (0.049)	0.001 (0.022)	-0.011 (0.026)	0.021 (0.055)
	None	-- --	0.013 (0.048)	0.008 (0.040)	0.009 (0.041)
30%	Large	0.004 (0.018)	0.016 (0.040)	0.012 (0.030)	0.017 (0.017)
	Medium	0.007 (0.034)	0.014 (0.041)	0.017 (0.028)	0.025 (0.057)
	Small	0.015 (0.042)	0.008 (0.030)	0.015 (0.052)	0.027 (0.048)
	None	-- --	0.011 (0.040)	0.033 (0.029)	0.024 (0.041)
15%	Large	0.009 (0.037)	0.031 (0.053)	0.001 (0.041)	0.013 (0.041)
	Medium	0.038 (0.042)	0.018 (0.033)	0.005 (0.047)	0.031 (0.045)
	Small	0.028 (0.060)	0.014 (0.048)	0.008 (0.044)	-0.009 (0.028)
	None	0.012 (0.048)	-0.005 (0.038)	0.019 (0.042)	0.027 (0.048)

APPENDIX O

SLOPE-QUADRATIC SLOPE COVARIANCE PARAMETER *P*-VALUE

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.255 (0.199)	0.418 (0.261)	0.315 (0.156)	0.286 (0.258)
	Medium	0.393 (0.246)	0.411 (0.310)	0.221 (0.244)	0.291 (0.158)
	Small	0.367 (0.294)	0.397 (0.255)	0.431 (0.229)	0.446 (0.293)
	None	-- --	0.294 (0.180)	0.298 (0.229)	0.396 (0.262)
30%	Large	0.348 (0.240)	0.256 (0.208)	0.375 (0.241)	0.383 (0.298)
	Medium	0.453 (0.241)	0.289 (0.169)	0.337 (0.278)	0.299 (0.191)
	Small	0.274 (0.213)	0.399 (0.281)	0.381 (0.283)	0.218 (0.223)
	None	-- --	0.348 (0.297)	0.228 (0.222)	0.276 (0.202)
15%	Large	0.350 (0.236)	0.332 (0.256)	0.401 (0.297)	0.251 (0.211)
	Medium	0.232 (0.223)	0.236 (0.152)	0.459 (0.237)	0.226 (0.147)
	Small	0.265 (0.275)	0.332 (0.250)	0.404 (0.282)	0.501 (0.251)
	None	0.370 (0.248)	0.406 (0.283)	0.372 (0.269)	0.265 (0.214)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.252 (0.180)	0.373 (0.235)	0.441 (0.215)	0.348 (0.267)
	Medium	0.374 (0.258)	0.461 (0.338)	0.328 (0.292)	0.440 (0.238)
	Small	0.447 (0.370)	0.430 (0.135)	0.517 (0.279)	0.515 (0.366)
	None	-- --	0.437 (0.295)	0.394 (0.260)	0.429 (0.232)
30%	Large	0.372 (0.209)	0.353 (0.296)	0.462 (0.267)	0.314 (0.209)
	Medium	0.475 (0.189)	0.372 (0.207)	0.280 (0.104)	0.367 (0.213)
	Small	0.297 (0.202)	0.390 (0.305)	0.433 (0.292)	0.254 (0.203)
	None	-- --	0.381 (0.304)	0.243 (0.185)	0.313 (0.256)
15%	Large	0.373 (0.281)	0.317 (0.229)	0.482 (0.325)	0.312 (0.260)
	Medium	0.218 (0.202)	0.321 (0.195)	0.429 (0.260)	0.261 (0.170)
	Small	0.266 (0.233)	0.379 (0.293)	0.440 (0.287)	0.593 (0.289)
	None	0.424 (0.263)	0.525 (0.325)	0.396 (0.279)	0.351 (0.284)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.248 (0.179)	0.317 (0.214)	0.420 (0.209)	0.325 (0.246)
	Medium	0.367 (0.249)	0.457 (0.344)	0.322 (0.299)	0.421 (0.240)
	Small	0.467 (0.344)	0.420 (0.133)	0.510 (0.280)	0.505 (0.368)
	None	-- --	0.431 (0.278)	0.359 (0.229)	0.427 (0.222)
30%	Large	0.363 (0.207)	0.347 (0.291)	0.428 (0.264)	0.302 (0.201)
	Medium	0.468 (0.184)	0.375 (0.205)	0.275 (0.097)	0.374 (0.211)
	Small	0.294 (0.198)	0.386 (0.299)	0.430 (0.288)	0.264 (0.201)
	None	-- --	0.389 (0.310)	0.245 (0.186)	0.331 (0.249)
15%	Large	0.371 (0.282)	0.315 (0.225)	0.479 (0.324)	0.313 (0.260)
	Medium	0.216 (0.197)	0.321 (0.195)	0.430 (0.263)	0.261 (0.170)
	Small	0.265 (0.231)	0.367 (0.286)	0.440 (0.286)	0.594 (0.288)
	None	0.423 (0.262)	0.525 (0.328)	0.395 (0.280)	0.351 (0.285)

APPENDIX P

SLOPE-QUADRATIC SLOPE COVARIANCE PARAMETER ESTIMATE

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.087 (0.043)	-0.060 (0.042)	-0.071 (0.039)	-0.088 (0.057)
	Medium	-0.073 (0.066)	-0.065 (0.050)	-0.103 (0.055)	-0.072 (0.037)
	Small	-0.079 (0.061)	-0.060 (0.035)	-0.054 (0.032)	-0.064 (0.061)
	None	-- --	-0.083 (0.055)	-0.083 (0.056)	-0.062 (0.039)
30%	Large	-0.066 (0.035)	-0.084 (0.043)	-0.069 (0.047)	-0.066 (0.046)
	Medium	-0.053 (0.030)	-0.083 (0.057)	-0.072 (0.036)	-0.085 (0.058)
	Small	-0.086 (0.053)	-0.068 (0.048)	-0.075 (0.068)	-0.105 (0.060)
	None	-- --	-0.079 (0.054)	-0.097 (0.048)	-0.088 (0.054)
15%	Large	-0.073 (0.050)	-0.077 (0.045)	-0.074 (0.060)	-0.095 (0.054)
	Medium	-0.099 (0.050)	-0.088 (0.043)	-0.052 (0.036)	-0.094 (0.045)
	Small	-0.102 (0.070)	-0.080 (0.056)	-0.067 (0.052)	-0.047 (0.034)
	None	-0.071 (0.045)	-0.067 (0.050)	-0.074 (0.055)	-0.096 (0.059)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.094	-0.072	-0.059	-0.083
		(0.042)	(0.046)	(0.043)	(0.059)
	Medium	-0.091	-0.064	-0.087	-0.060
		(0.086)	(0.053)	(0.057)	(0.044)
	Small	-0.077	-0.054	-0.049	-0.067
		(0.069)	(0.022)	(0.041)	(0.073)
	None	--	-0.073	-0.072	-0.060
		--	(0.065)	(0.053)	(0.033)
30%	Large	-0.063	-0.081	-0.062	-0.077
		(0.027)	(0.056)	(0.049)	(0.037)
	Medium	-0.052	-0.075	-0.080	-0.080
		(0.025)	(0.056)	(0.024)	(0.059)
	Small	-0.084	-0.074	-0.072	-0.106
		(0.047)	(0.053)	(0.069)	(0.065)
	None	--	-0.078	-0.097	-0.089
		--	(0.056)	(0.042)	(0.054)
15%	Large	-0.076	-0.086	-0.065	-0.090
		(0.051)	(0.048)	(0.068)	(0.057)
	Medium	-0.106	-0.082	-0.063	-0.094
		(0.049)	(0.048)	(0.046)	(0.048)
	Small	-0.106	-0.082	-0.067	-0.041
		(0.071)	(0.068)	(0.056)	(0.038)
	None	-0.067	-0.057	-0.075	-0.091
		(0.051)	(0.059)	(0.055)	(0.071)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.092 (0.043)	-0.077 (0.045)	-0.060 (0.042)	-0.082 (0.056)
	Medium	-0.089 (0.082)	-0.064 (0.052)	-0.088 (0.058)	-0.062 (0.044)
	Small	-0.070 (0.065)	-0.055 (0.021)	-0.050 (0.043)	-0.068 (0.072)
	None	-- --	-0.073 (0.062)	-0.077 (0.052)	-0.060 (0.033)
30%	Large	-0.063 (0.027)	-0.081 (0.056)	-0.067 (0.050)	-0.079 (0.037)
	Medium	-0.052 (0.024)	-0.074 (0.055)	-0.080 (0.023)	-0.079 (0.061)
	Small	-0.085 (0.047)	-0.074 (0.053)	-0.072 (0.068)	-0.103 (0.063)
	None	-- --	-0.077 (0.056)	-0.097 (0.043)	-0.085 (0.053)
15%	Large	-0.076 (0.052)	-0.085 (0.048)	-0.065 (0.066)	-0.090 (0.056)
	Medium	-0.106 (0.049)	-0.081 (0.047)	-0.063 (0.046)	-0.094 (0.048)
	Small	-0.106 (0.071)	-0.083 (0.065)	-0.067 (0.056)	-0.041 (0.038)
	None	-0.067 (0.050)	-0.058 (0.060)	-0.076 (0.056)	-0.092 (0.071)

APPENDIX Q

QUADRATIC SLOPE PARAMETER *P*-VALUE

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.140 (0.277)	0.184 (0.292)	0.061 (0.126)	0.019 (0.043)
	Medium	0.236 (0.246)	0.074 (0.170)	0.167 (0.256)	0.153 (0.295)
	Small	0.412 (0.440)	0.222 (0.278)	0.148 (0.246)	0.169 (0.277)
	None	-- --	0.232 (0.332)	0.130 (0.251)	0.107 (0.189)
30%	Large	0.231 (0.295)	0.143 (0.230)	0.120 (0.216)	0.082 (0.185)
	Medium	0.361 (0.311)	0.210 (0.266)	0.275 (0.336)	0.173 (0.260)
	Small	0.243 (0.241)	0.454 (0.322)	0.300 (0.294)	0.075 (0.159)
	None	-- --	0.281 (0.354)	0.144 (0.273)	0.197 (0.206)
15%	Large	0.163 (0.157)	0.239 (0.298)	0.297 (0.345)	0.254 (0.257)
	Medium	0.270 (0.338)	0.200 (0.241)	0.184 (0.248)	0.276 (0.335)
	Small	0.526 (0.293)	0.355 (0.341)	0.257 (0.285)	0.254 (0.326)
	None	0.268 (0.259)	0.140 (0.185)	0.282 (0.306)	0.270 (0.282)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.033 (0.070)	0.052 (0.055)	0.022 (0.053)	0.004 (0.006)
	Medium	0.215 (0.306)	0.040 (0.085)	0.041 (0.095)	0.016 (0.035)
	Small	0.138 (0.160)	0.087 (0.111)	0.024 (0.042)	0.072 (0.148)
	None	-- --	0.125 (0.191)	0.024 (0.058)	0.014 (0.026)
30%	Large	0.177 (0.315)	0.061 (0.109)	0.024 (0.048)	0.006 (0.015)
	Medium	0.259 (0.274)	0.101 (0.226)	0.104 (0.155)	0.040 (0.091)
	Small	0.097 (0.168)	0.251 (0.335)	0.154 (0.228)	0.014 (0.031)
	None	-- --	0.170 (0.283)	0.092 (0.196)	0.087 (0.163)
15%	Large	0.171 (0.229)	0.036 (0.060)	0.062 (0.073)	0.138 (0.220)
	Medium	0.095 (0.130)	0.135 (0.255)	0.020 (0.026)	0.035 (0.034)
	Small	0.306 (0.271)	0.152 (0.265)	0.152 (0.212)	0.163 (0.282)
	None	0.171 (0.287)	0.049 (0.052)	0.200 (0.310)	0.194 (0.241)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.021 (0.041)	0.005 (0.007)	0.002 (0.008)	0.000 (0.000)
	Medium	0.125 (0.170)	0.009 (0.021)	0.006 (0.012)	0.000 (0.001)
	Small	0.078 (0.085)	0.040 (0.066)	0.003 (0.004)	0.010 (0.018)
	None	-- --	0.120 (0.218)	0.015 (0.022)	0.003 (0.005)
30%	Large	0.109 (0.174)	0.045 (0.073)	0.021 (0.051)	0.002 (0.006)
	Medium	0.239 (0.292)	0.080 (0.224)	0.049 (0.061)	0.023 (0.042)
	Small	0.071 (0.080)	0.218 (0.335)	0.109 (0.221)	0.012 (0.031)
	None	-- --	0.126 (0.171)	0.081 (0.175)	0.088 (0.166)
15%	Large	0.154 (0.196)	0.036 (0.061)	0.086 (0.120)	0.127 (0.246)
	Medium	0.102 (0.145)	0.116 (0.196)	0.016 (0.020)	0.044 (0.048)
	Small	0.281 (0.237)	0.133 (0.217)	0.156 (0.245)	0.159 (0.280)
	None	0.175 (0.245)	0.051 (0.070)	0.201 (0.335)	0.159 (0.199)

APPENDIX R

QUADRATIC SLOPE PARAMETER ESTIMATE

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.028 (0.015)	-0.024 (0.016)	-0.032 (0.016)	-0.041 (0.015)
	Medium	-0.016 (0.016)	-0.028 (0.011)	-0.026 (0.016)	-0.031 (0.020)
	Small	-0.014 (0.013)	-0.021 (0.016)	-0.024 (0.013)	-0.025 (0.018)
	None	-- --	-0.021 (0.014)	-0.025 (0.020)	-0.030 (0.015)
30%	Large	-0.018 (0.014)	-0.029 (0.020)	-0.022 (0.015)	-0.038 (0.018)
	Medium	-0.007 (0.020)	-0.019 (0.015)	-0.017 (0.020)	-0.023 (0.012)
	Small	-0.018 (0.012)	-0.007 (0.013)	-0.016 (0.017)	-0.033 (0.014)
	None	-- --	-0.016 (0.016)	-0.024 (0.019)	-0.022 (0.018)
15%	Large	-0.014 (0.012)	-0.022 (0.015)	-0.016 (0.012)	-0.018 (0.015)
	Medium	-0.018 (0.018)	-0.020 (0.021)	-0.023 (0.013)	-0.022 (0.019)
	Small	-0.003 (0.011)	-0.014 (0.012)	-0.018 (0.017)	-0.020 (0.018)
	None	-0.015 (0.012)	-0.020 (0.015)	-0.009 (0.018)	-0.017 (0.017)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.036 (0.012)	-0.032 (0.016)	-0.036 (0.011)	-0.043 (0.011)
	Medium	-0.019 (0.009)	-0.030 (0.010)	-0.034 (0.012)	-0.038 (0.012)
	Small	-0.022 (0.009)	-0.027 (0.013)	-0.035 (0.012)	-0.030 (0.011)
	None	-- --	-0.028 (0.015)	-0.035 (0.012)	-0.040 (0.014)
30%	Large	-0.025 (0.013)	-0.034 (0.016)	-0.034 (0.009)	-0.045 (0.013)
	Medium	-0.020 (0.014)	-0.028 (0.011)	-0.028 (0.013)	-0.033 (0.010)
	Small	-0.027 (0.011)	-0.020 (0.013)	-0.024 (0.013)	-0.039 (0.010)
	None	-- --	-0.026 (0.014)	-0.033 (0.013)	-0.033 (0.016)
15%	Large	-0.021 (0.009)	-0.034 (0.013)	-0.026 (0.008)	-0.024 (0.012)
	Medium	-0.030 (0.014)	-0.028 (0.015)	-0.032 (0.008)	-0.034 (0.016)
	Small	-0.014 (0.007)	-0.024 (0.011)	-0.025 (0.014)	-0.027 (0.015)
	None	-0.023 (0.012)	-0.028 (0.010)	-0.024 (0.014)	-0.023 (0.014)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	-0.042	-0.044	-0.055	-0.067
		(0.015)	(0.015)	(0.016)	(0.014)
	Medium	-0.024	-0.040	-0.046	-0.055
		(0.010)	(0.011)	(0.015)	(0.012)
	Small	-0.025	-0.035	-0.047	-0.043
		(0.007)	(0.017)	(0.012)	(0.012)
	None	--	-0.032	-0.039	-0.047
		--	(0.016)	(0.013)	(0.014)
30%	Large	-0.027	-0.039	-0.037	-0.050
		(0.011)	(0.019)	(0.011)	(0.012)
	Medium	-0.022	-0.030	-0.031	-0.037
		(0.015)	(0.011)	(0.012)	(0.012)
	Small	-0.027	-0.023	-0.027	-0.045
		(0.010)	(0.014)	(0.014)	(0.012)
	None	--	-0.028	-0.036	-0.035
		--	(0.014)	(0.015)	(0.018)
15%	Large	-0.022	-0.035	-0.027	-0.026
		(0.010)	(0.013)	(0.011)	(0.012)
	Medium	-0.030	-0.028	-0.033	-0.033
		(0.014)	(0.014)	(0.007)	(0.017)
	Small	-0.015	-0.024	-0.027	-0.028
		(0.008)	(0.011)	(0.015)	(0.016)
	None	-0.023	-0.029	-0.025	-0.025
		(0.012)	(0.011)	(0.014)	(0.013)

APPENDIX S

QUADRATIC SLOPE VARIANCE PARAMETER *P*-VALUE

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.298 (0.238)	0.403 (0.270)	0.343 (0.161)	0.303 (0.270)
	Medium	0.383 (0.259)	0.393 (0.284)	0.227 (0.195)	0.252 (0.190)
	Small	0.327 (0.204)	0.390 (0.231)	0.395 (0.255)	0.421 (0.252)
	None	-- --	0.283 (0.182)	0.245 (0.210)	0.381 (0.233)
30%	Large	0.371 (0.226)	0.261 (0.196)	0.406 (0.248)	0.417 (0.289)
	Medium	0.426 (0.197)	0.280 (0.174)	0.337 (0.225)	0.318 (0.221)
	Small	0.261 (0.219)	0.383 (0.282)	0.350 (0.230)	0.248 (0.249)
	None	-- --	0.311 (0.225)	0.236 (0.249)	0.340 (0.237)
15%	Large	0.338 (0.251)	0.330 (0.266)	0.385 (0.261)	0.238 (0.207)
	Medium	0.279 (0.220)	0.235 (0.143)	0.446 (0.265)	0.236 (0.191)
	Small	0.221 (0.241)	0.340 (0.257)	0.361 (0.243)	0.379 (0.246)
	None	0.352 (0.219)	0.354 (0.219)	0.343 (0.257)	0.286 (0.211)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.339 (0.212)	0.418 (0.252)	0.524 (0.203)	0.364 (0.262)
	Medium	0.430 (0.314)	0.485 (0.285)	0.351 (0.282)	0.367 (0.117)
	Small	0.439 (0.282)	0.498 (0.144)	0.440 (0.246)	0.569 (0.359)
	None	-- --	0.434 (0.287)	0.427 (0.323)	0.483 (0.209)
30%	Large	0.402 (0.210)	0.394 (0.292)	0.520 (0.276)	0.399 (0.250)
	Medium	0.535 (0.135)	0.433 (0.221)	0.395 (0.154)	0.438 (0.219)
	Small	0.329 (0.190)	0.376 (0.259)	0.481 (0.257)	0.308 (0.239)
	None	-- --	0.414 (0.280)	0.298 (0.210)	0.396 (0.268)
15%	Large	0.401 (0.275)	0.323 (0.198)	0.493 (0.288)	0.333 (0.250)
	Medium	0.304 (0.179)	0.372 (0.205)	0.443 (0.229)	0.291 (0.191)
	Small	0.294 (0.227)	0.401 (0.293)	0.440 (0.266)	0.530 (0.312)
	None	0.439 (0.266)	0.486 (0.276)	0.409 (0.264)	0.385 (0.226)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.330 (0.219)	0.354 (0.207)	0.513 (0.201)	0.361 (0.260)
	Medium	0.429 (0.314)	0.487 (0.288)	0.353 (0.286)	0.363 (0.118)
	Small	0.504 (0.319)	0.495 (0.148)	0.443 (0.243)	0.571 (0.357)
	None	-- --	0.448 (0.280)	0.395 (0.304)	0.482 (0.210)
30%	Large	0.398 (0.209)	0.398 (0.296)	0.476 (0.257)	0.399 (0.252)
	Medium	0.531 (0.136)	0.448 (0.224)	0.395 (0.155)	0.435 (0.227)
	Small	0.329 (0.192)	0.380 (0.261)	0.483 (0.254)	0.331 (0.247)
	None	-- --	0.417 (0.286)	0.298 (0.213)	0.410 (0.275)
15%	Large	0.402 (0.274)	0.323 (0.198)	0.492 (0.288)	0.333 (0.249)
	Medium	0.303 (0.180)	0.372 (0.206)	0.446 (0.233)	0.291 (0.191)
	Small	0.294 (0.226)	0.407 (0.281)	0.440 (0.267)	0.528 (0.313)
	None	0.436 (0.263)	0.485 (0.274)	0.407 (0.261)	0.385 (0.226)

APPENDIX T

QUADRATIC SLOPE VARIANCE PARAMETER ESTIMATE

First-Order Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.025 (0.015)	0.018 (0.011)	0.020 (0.010)	0.026 (0.018)
	Medium	0.022 (0.016)	0.020 (0.014)	0.030 (0.016)	0.024 (0.010)
	Small	0.024 (0.017)	0.018 (0.009)	0.018 (0.011)	0.019 (0.016)
	None	-- --	0.027 (0.018)	0.029 (0.017)	0.019 (0.010)
30%	Large	0.019 (0.010)	0.024 (0.011)	0.021 (0.016)	0.018 (0.013)
	Medium	0.017 (0.009)	0.025 (0.015)	0.022 (0.011)	0.025 (0.016)
	Small	0.027 (0.016)	0.021 (0.014)	0.024 (0.020)	0.031 (0.020)
	None	-- --	0.025 (0.015)	0.030 (0.016)	0.024 (0.017)
15%	Large	0.023 (0.015)	0.024 (0.014)	0.024 (0.020)	0.030 (0.017)
	Medium	0.026 (0.012)	0.027 (0.013)	0.017 (0.010)	0.030 (0.016)
	Small	0.033 (0.018)	0.024 (0.016)	0.021 (0.012)	0.019 (0.011)
	None	0.022 (0.013)	0.022 (0.016)	0.024 (0.016)	0.027 (0.015)

Second-Order Constrained Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.023 (0.012)	0.018 (0.010)	0.014 (0.010)	0.023 (0.016)
	Medium	0.023 (0.021)	0.017 (0.013)	0.024 (0.015)	0.019 (0.008)
	Small	0.021 (0.017)	0.014 (0.005)	0.017 (0.011)	0.017 (0.020)
	None	-- --	0.022 (0.019)	0.021 (0.016)	0.015 (0.008)
30%	Large	0.018 (0.008)	0.021 (0.014)	0.017 (0.016)	0.019 (0.011)
	Medium	0.013 (0.005)	0.019 (0.014)	0.019 (0.008)	0.019 (0.014)
	Small	0.023 (0.012)	0.021 (0.012)	0.019 (0.019)	0.028 (0.019)
	None	-- --	0.020 (0.013)	0.026 (0.013)	0.022 (0.016)
15%	Large	0.021 (0.015)	0.023 (0.010)	0.020 (0.020)	0.026 (0.017)
	Medium	0.025 (0.011)	0.022 (0.014)	0.018 (0.011)	0.026 (0.014)
	Small	0.028 (0.017)	0.022 (0.017)	0.018 (0.013)	0.014 (0.012)
	None	0.019 (0.013)	0.018 (0.016)	0.021 (0.015)	0.024 (0.017)

Second-Order Free Non-Linear Latent Growth Model					
Percent of DIF	Loading DIF	Intercept DIF			
		None	Small	Medium	Large
45%	Large	0.023 (0.012)	0.020 (0.010)	0.014 (0.010)	0.022 (0.015)
	Medium	0.022 (0.020)	0.017 (0.013)	0.023 (0.015)	0.018 (0.008)
	Small	0.018 (0.017)	0.014 (0.005)	0.016 (0.011)	0.017 (0.020)
	None	-- --	0.021 (0.019)	0.022 (0.016)	0.015 (0.008)
30%	Large	0.018 (0.008)	0.021 (0.014)	0.018 (0.016)	0.019 (0.011)
	Medium	0.013 (0.005)	0.018 (0.014)	0.019 (0.008)	0.019 (0.014)
	Small	0.023 (0.012)	0.021 (0.012)	0.019 (0.019)	0.027 (0.019)
	None	-- --	0.020 (0.014)	0.026 (0.013)	0.022 (0.017)
15%	Large	0.021 (0.015)	0.023 (0.010)	0.020 (0.020)	0.026 (0.017)
	Medium	0.025 (0.011)	0.022 (0.014)	0.017 (0.011)	0.026 (0.014)
	Small	0.028 (0.017)	0.022 (0.017)	0.018 (0.013)	0.014 (0.012)
	None	0.019 (0.012)	0.018 (0.016)	0.021 (0.015)	0.024 (0.017)

APPENDIX U

TEST LENGTH RESULTS COMPARISON FOR THE SECOND-ORDER CONSTRAINED LATENT GROWTH MODEL

Second-Order Constrained Latent Growth Model					
	Percent of DIF	20 Item		40 Item*	
		Mean	*SD*	*Mean*	*SD*
Model Fit					
Chi-squared *p*-value	30%	0.000	0.000	0.000	0.000
	15%	0.000	0.000	0.000	0.000
RMSEA	30%	0.009	0.001	0.005	0.001
	15%	0.008	0.001	0.005	0.001
CFI	30%	0.998	0.001	0.999	0.000
	15%	0.998	0.000	0.999	0.000
Parameter Recovery					
Slope Relative Bias	30%	0.119	0.036	0.122	0.051
	15%	0.038	0.036	0.028	0.056
Slope RMSE	30%	0.059	0.018	0.061	0.026
	15%	0.021	0.015	0.024	0.017
Slope Variance Relative Bias	30%	-0.179	0.118	-0.193	0.025
	15%	-0.149	0.120	-0.179	0.108
Slope Variance RMSE	30%	0.039	0.018	0.039	0.005
	15%	0.032	0.021	0.036	0.022
Intercept Variance Relative Bias	30%	0.099	0.086	0.127	0.072
	15%	0.128	0.100	0.066	0.115
Intercept Variance RMSE	30%	0.103	0.081	0.127	0.072
	15%	0.128	0.100	0.086	0.096
Incorrectly Specified Growth Parameters					
Covariance *p*-value	30%	0.047	0.065	0.040	0.063
	15%	0.038	0.055	0.097	0.094
Covariance Estimate	30%	-0.074	0.028	-0.071	0.019
	15%	-0.084	0.033	-0.066	0.038

* 40-item condition based on 5 replications

APPENDIX V

TEST LENGTH RESULTS COMPARISON FOR THE SECOND-ORDER FREE LATENT GROWTH MODEL

Second-Order Free Latent Growth Model					
	Percent of DIF	20 Item		40 Item*	
		Mean	*SD*	*Mean*	*SD*
Model Fit					
Chi-squared *p*-value	30%	0.000	0.000	0.000	0.000
	15%	0.000	0.000	0.000	0.001
RMSEA	30%	0.008	0.001	0.005	0.001
	15%	0.008	0.001	0.004	0.001
CFI	30%	0.998	0.001	0.999	0.000
	15%	0.998	0.000	0.999	0.000
Parameter Recovery					
Slope Relative Bias	30%	0.186	0.046	0.178	0.048
	15%	0.053	0.041	0.043	0.062
Slope RMSE	30%	0.093	0.023	0.089	0.024
	15%	0.029	0.017	0.028	0.023
Slope Variance Relative Bias	30%	-0.184	0.120	-0.199	0.023
	15%	-0.151	0.118	-0.180	0.110
Slope Variance RMSE	30%	0.040	0.018	0.040	0.005
	15%	0.032	0.021	0.036	0.022
Intercept Variance Relative Bias	30%	0.096	0.084	0.120	0.072
	15%	0.127	0.099	0.065	0.115
Intercept Variance RMSE	30%	0.100	0.080	0.120	0.072
	15%	0.127	0.099	0.087	0.095
Incorrectly Specified Growth Parameters					
Covariance *p*-value	30%	0.046	0.066	0.043	0.068
	15%	0.035	0.051	0.097	0.094
Covariance Estimate	30%	-0.073	0.028	-0.070	0.019
	15%	-0.084	0.033	-0.066	0.038

* 40-item condition based on 5 replications