The Performance of a Higher-Order Invariance Idiographic Filter in Confirmatory Factor Analysis of Cross Sectional and Longitudinal Data

Geneva Taylor

Philadelphia, PA

B.A., LaSalle University, 2006
M.A., University of Virginia, 2012

A Dissertation Submitted to the Graduate Faculty of
the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Psychology
University of Virginia
August, 2016

Steven M. Boker (Chair)

Melvin N. Wilson

Timo Von Oertzen

Daniel M. Keenan
(Department of Statistics)

Abstract

This work evaluates the performance of a methodology for analyzing cross sectional and longitudinal latent factor data featuring measurement non-invariance. An overview of the historical context and current methodologies for establishing factorial invariance is presented, and the higher-order invariance idiographic filter (HOIIF) method is proposed as a candidate for establishing factorial invariance in conditions of measurement non-invariance. HOIIF allows for flexibility in the measurement structure of a latent factor model, while constraining invariance on a higher factor order. However, the method is currently in a nascent stage and the performance of the HOIIF method in conditions of varying sample size, measurement model, and higher-order equivalence has not yet been established. Therefore, the current study implements a Monte Carlo simulation of A) cross-sectional and B) longitudinal data sets in order to evaluate HOIIF based on recovery of the first-order factor loadings and expected error, ratios of rejected to non-rejected first-order constrained models, ratios of rejected to non-rejected second-order constrained models, and model convergence. Following the Monte Carlo simulations, a practical application of HOIIF to empirical data is employed. Results indicate that the HOIIF method is sufficiently sensitive (fails to reject correct models) and specific (rejects incorrect models), and can produce accurate parameter estimates. The study highlights the importance of following the procedure in order, correct model specification, and consideration of sample size and model complexity.

Table of Contents

1. Introduction

1.1 Overview

Invariance is a foundational concept in striving to understand performance between individuals or groups. Specifically, psychology is typically concerned with understanding differences (variation), for example, whether scores are different for individuals who received a treatment and those who did not, or whether level of education can predict scores on a reading test. In order to understand those differences and make comparisons, there must be aspects of the assessment that are invariant across people–if the assessment were different for everyone, separating variation due to the variable of interest and variation due to the differences in the assessment becomes a difficult task. Consider a researcher who is interested in comparing an individual's weight before and after a diet change. The researcher assesses weight using an electronic scale before the diet change and finds that the person weighs 150 lbs. After the diet change, the researcher uses a balance to find that the person weighs 25 mallard ducks. Did the person gain or lose weight? Without knowing how the assessments relate to each other (being able to transform lbs to ducks or ducks to lbs), it is unclear, because the assessment itself was not invariant. This issue becomes even more complex in the context of assessing latent factors, as the variable of interest itself is not directly observable. In the following sections, I will provide a brief overview of the background and aims of the current study. Following the overview, I will present important literature that provides a foundation for issues that will be addressed by the current study, including the history of invariance, modern invariance concerns, and how invariance is addressed in practice. After the review of important literature, I will present the methodology for the current study, including the analysis of simulated as well as observed data, followed by the results and indications for moving forward.

Psychology frequently concerns itself with that which is not directly observable (e.g. *spatial ability*; in comparison, *heart rate* is an example of a variable that is directly observable), also called latent factors. In order to assess these latent factors, Spearman (1904a,b) showed that the correlation matrix for a variety of observed items theorized to assess a latent factor could be decomposed into two parts–a shared, common variance across all of the observations, and a unique variance that was specific to each variable. Spearman's model eventually became known as the common factor model, which theorizes that a latent, or unobservable, factor can account for variation in a set of observed variables across a sample of observations. This work grew into the method currently known as exploratory factor analysis (EFA). At the time, Spearman did not address how this factor might vary depending on the sample selected (Millsap and Meredith, 2007).

In the field of psychology, notions of invariance grew out of selection theory from Pearson (1903); Aitken (1935) and Lawley (1944). The Pearson-Aitken selection formula (Pearson, 1903; Aitken, 1935), later generalized to multiple factors (Thurstone, 1931, 1947; Thomson and Lederunn, 1939; Ahmavaara, 1954; Meredith, 1964a), was concerned with how selection affects model parameters and established that selection can operate in such a way that factor loadings can be transformed to be identical, or *invariant*. Confirmatory factor analysis (CFA Jöreskog, 1969; Jöreskog, 1970) soon emerged as a superior method of investigating invariance between samples and groups. CFA models expanded to include higher-order models in investigations of invariance (Marsh and Hocevar, 1985; Marsh, 1985; Thurstone, 1931, 1947), with a suggestion that they be considered as plausible alternatives to the model of interest. Invariance in these models holds the same meaning and is established in the same method (Marsh and Hocevar, 1985).

Comparing scores between groups is currently seen as untenable in the absence of invariant measurement (Cheung and Rensvold, 1999; Schmitt and Kuljanin, 2008; Widaman and Reise, 1997). Measurement invariance is, therefore, considered a crucial step in cross-cultural research, as instruments can vary in their measurement qualities and validity across cultures. Configural invariance may even hold while measurement is non-invariant across groups (Byrne, 2003). Byrne (2003) points to bias as one of the primary concerns in cross-cultural research. In particular, construct-, instrument-, and item-derived bias are not uncommon concerns in cross cultural research (Byrne, 2003) and these biases can result in measurement non-invariance. Likewise, longitudinal research across age has shown levels of non-invariance (e.g. Horn and McArdle, 1992; Meredith and Horn, 2001). For example, the way in which an individual interprets items on a scale can vary as they age. In a longitudinal study measuring psychological distress, Drapeau et al. (2010) were able to establish configural invariance–invariance in the location of zero- and non-zero-loadings–but could not establish complete metric invariance. Drapeau et al. found that a number of factor loadings had to be unconstrained at various time points in order to establish a model with suitable fit. Yet methods for dealing with this type of non-invariance have been theorized (e.g. Nesselroade et al., 2007), but not fully investigated under varying sample characteristics.

Thus, a gap exists in establishing factor invariance in conditions of non-invariance across groups. When measurement is non-invariant, as is the case when assessments must be modified for children of different ages or when item biases affect how the questions are answered by participants, there are few current methods for being able to still establish invariance in the factor being assessed. Methods such as the idiographic filter (Nesselroade et al., 2007), also called a higher-order invariance idiographic filter (HOIIF Zhang et al., 2011), have been advanced but not fully developed.

A HOIIF confirmatory factor model involves imposing invariance at a higher-order than in a traditional confirmatory factory model. It is a method of filtering idiosyncratic differences while still establishing invariant relationships across individuals. In HOIIF, invariance is tested via an invariant second-order factor model, applied at the level of factor intercorrelations, while the first-order factor measurement model is allowed to vary. As an example of why it might be necessary to employ such a method, Chen et al. (2005) noted that in comparing the heights of Europeans and Americans, despite measurement non-invariance due to a difference in measurement units (centimeters vs inches), the construct being measured is invariant across individuals. To extend this example, in its essence a HOIIF would allow comparisons of height based on the relationship between height and weight, rather than directly from the observed values due to the non-invariance of the measurement.

The current study is an exploration of HOIIF. I used a pair of Monte Carlo simulation studies to investigate the performance of the HOIIF method for multiple-group latent factor methods, specifically in cross-sectional and longitudinal data. For example, the cross-sectional data were simulated to feature nine manifest ("observed") variables, three first-order factors, one second-order factor, and two groups (e.g. Figure 3.3). I compared samples with differing sample size, type of first order (non-)invariance, and presence or absence of second-order invariance. Using simulated data allows for direct comparisons of the known parameters of the data to the model-estimated parameters.

Following this, I applied the method to one cross-sectional and one longitudinal empirical data set as practical examples of implementation and results from the HOIIF method. The overall purpose of the current study is to investigate the utility and performance of a higher-order invariance idiographic filter for testing across-group invariance-based equivalence in latent factor methods.

In demonstrating the performance of the HOIIF in establishing factorial invariance in conditions of non-invariant measurement, I provide evidence for the feasibility of the method in not uncommon research circumstances–namely, in cross-sectional and longitudinal studies with non-invariant measurement. Until now, performance of the HOIIF method had not been fully explored. The results of the current study indicate that this method is a viable option for establishing factorial invariance, providing strong evidence for the application of the HOIIF as an option for investigating group differences on latent factors in conditions of measurement non-invariance. As such, in cases where higher-order factor structure is invariant despite non-invariant first-order measurement, the HOIIF could open a previously inaccessible route to make cross-cultural comparisons on latent factors, to investigate the effectiveness of interventions in varying age groups for which the measurement instrument is different by design, and other analogous situations.

With the results for the simulations, I address the following considerations:

- How well does this method recover parameters? In other words, how often and how closely do the point estimates match the parameters used to create the data?
- How much bias is present in the estimates?
- How specific is it? How often does it reject the models it should reject and fail to reject the models it should fail to reject?
- In what ways do the relevant fit statistics vary relative to accuracy of point estimates, bias, and specificity?
- How does the structure of the measurement model affect the previous considerations?
- How does sample size affect the previous considerations?
- How often does the method fail to converge?
- Do convergence issues depend on sample size, measurement model, or the second-order variance condition? Does varying the starting values affect convergence?

Subsequently, by applying the method to empirical data I address possible concerns regarding practical implementation (e.g. starting values, integrating theory) and how to interpret results from applying the method.

1.2 Important Terms

In the next sections, I will be discussing the history and usage of invariance and methods for determining invariance. Some of the terms that are germane to this discussion have multiple definitions depending on context or otherwise can be unclear; therefore, I have highlighted and provided brief definitions for a few important terms below. These terms will be further clarified as they arise throughout the subsequent sections.

1. CFA: Confirmatory factor analysis

2. Construct Validity: That the latent construct or factor being assessed is that which is intended to be assessed

3. EFA: Exploratory factor analysis

4. Factor(ial) Invariance: Equivalence of latent constructs across observations. Involves some level of statistical invariance, traditionally imposed at the level of factor measurement (see Measurement Invariance) but potentially imposed at a higher-order, as in HOIIF

5. Higher-Order Factor Model: Factor model that contains at least one latent factor that predicts/is measured by/loads onto other latent factors that are measured by manifest variables

6. Higher-Order Invariance Idiographic Filtering (HOIIF): A method in which invariance is imposed on either higher-order factor structure or higher-order factor dynamics, rather than on the first-order measurement structure

7. Idiographic Filter: *See HOIIF*

8. Invariance/Non-Invariance: Equivalence across observations/lack of equivalence across observations

9. Measurement Invariance (MI): Equivalence of assessment and assessment parameters across observations, at minimum includes factor loadings

10. Metric Invariance: Equivalence of factor loading values

1.3 Invariance

The concept of invariance emerged out of selection theory in factor analysis in the early 20th century (Pearson, 1903; Aitken, 1935; Thurstone, 1931, 1947; Thomson and Lederunn, 1939; Ahmavaara, 1954; Meredith, 1964a), and continues to be an important aspect of factor analysis to this day. In multiple group analysis, whether comparing across time or varying cultures or groups based on some other variable, measurement invariance has been considered necessary in order to establish factorial invariance (Cheung and Rensvold, 1999; Schmitt and Kuljanin, 2008; Vandenberg and Lance, 2000; Widaman and Reise, 1997). Occasions exist in which strict measurement invariance is not possible. For example, the Child Behavior Checklist (Achenbach and Rescorla, 2000, 2001) is a tool used to assess behavioral problems in children and exists in two versions–one for preschool-aged children and one for older, school-aged children–because of the extensive developmental changes that occur throughout childhood. Sometimes measurement is only partially invariant. For example, as one of the goals for their study, Schlotz et al. (2011) investigated measurement invariance across gender of a scale for assessing stress reactivity. The researchers found evidence that while some item factor loadings could be constrained to invariance across gender, other items were non-invariant across genders. In this case, Schlotz et al. (2011) interpret this as a need for gender-specific norms.

Formal methods for establishing factorial invariance in conditions of measurement non-invariance are still in development, and in the current study I sought to further develop one such method–the higher order invariance idiographic filter (HOIIF). I used simulated data to investigate the performance of the HOIIF method, and followed with an example of the method in one empirical cross-sectional data set as well as one empirical longitudinal data set. The next sections review the background literature related to invariance, confirmatory factor analysis, and cross-group comparisons.

2. Literature Review

Invariance was born out of selection theory, with the idea that there exists some rotation of parameters that will fit all subgroups of a population if a simple structure exists in any randomly selected subgroup (Meredith, 1964a; Thurstone, 1931, 1947). After establishing that there can exist some rotation for all randomly selected subgroups of a population, researchers found many empirical examples of evidence of factorial non-invariance across groups (e.g. Berger, 1968; Eysenck and Eysenck, 1969). From this, an interest in establishing invariance across groups grew while recognition that gender, culture, and other demographic variables can contribute to not just variation, but *bias* in studies also increased. This growing attention to group differences was bolstered by advances in computing that permitted more sophisticated analyses. Out of this, confirmatory factor analysis (CFA) emerged as a primary method for testing group invariance.

Because psychologists frequently seek to study hypothetical constructs that cannot be directly observed, researchers in psychology sought appropriate methods for assessing such unobservable, or latent, constructs. Spearman (1904a,b) produced seminal work on a method of using correlation to differentiate shared variance–which he hypothesized derived from an unobservable common factor–from unique variance in scores across a sample. This model eventually became known as the common factor model. The common factor model theorizes that a latent factor will account for variation in an set of observed variables across a sample:

$$Y = T + \Lambda F + E \tag{2.1}$$

In this model, Y is a $p \times n$ matrix of observations for p variables on n people, T is a $p \times n$ matrix consisting of a $p \times 1$ vector of manifest variable intercepts repeated n times, Λ is a

$p \times k$ matrix of loadings for k factors on p variables, F is a $k \times p$ matrix of common factor scores, and E is a $p \times n$ matrix of unique factor scores for each nth person on each pth variable. The covariance matrix Σ for this model can be expressed as

$$\Sigma = \Lambda\Psi\Lambda' + \Theta \tag{2.2}$$

where Σ is a $p \times p$ matrix of manifest variable variances and covariances, Ψ is a $k \times k$ matrix of common factor variances and covariances and Θ is a $p \times p$ diagonal matrix of variances of the unique factors.

This model, originally a decomposition of the covariance matrix of Y, indicates that an individual Y observation is a linear combination of (an intercept, along with) weighted score(s) on some common factor(s) and some amount of error. Invariance as a concern was not a consideration for Spearman (Millsap and Meredith, 2007). In the early years of factor analysis, the discipline of psychology tended to focus on generalization; thus, the focus was on measuring latent constructs that were invariant across people with little attention devoted to the implications of group differences (Millsap and Meredith, 2007). However, consideration was given to the notion that sampling individuals out of a population implied some process of *selection* having occurred to produce said sample, and researchers like Aitken (1935), Thomson and Lederunn (1939), Lawley (1944), and later Ahmavaara (1954), were interested in what effects this selection had on the parameters of the factor model.

In that groups are selected from a parent population based on some variable(s), the notion of invariance is rooted in selection. Therefore, the question arose as to whether parameters of interest were invariant across the groups. One of the earliest considerations of the effects of selection was first formulated by Pearson (1903) and refined by Aitken (1935) into what is now known as the Pearson-Aitken selection formulas, which demonstrates the effects of selection on group means and variances/covariances assuming the variables are normally distributed. Specifically, given random vectors a and b with means μ_a and μ_b and variance-covariance matrix

$$Cov_{ab} = \begin{pmatrix} V_a & C_{ab} \\ C_{ba} & V_b \end{pmatrix} \tag{2.3}$$

the Pearson-Aitken selection formula states that selection on a, changing μ_a to $\tilde{\mu_a}$, correspondingly changes the mean vector of μ_b to $\tilde{\mu_b}$ where $\tilde{\mu_b}$ equals

$$\tilde{\mu_b} = \mu_b + C_{ba}V_a^{-1}(\tilde{\mu_a} - \mu_a) \tag{2.4}$$

and changes the variance-covariance matrix to

$$\tilde{Cov}_{ab} = \begin{pmatrix} \tilde{V_a} & \tilde{V_a}V_a^{-1}C_{ab} \\ C_{ba}V_a^{-1}\tilde{V_a} & V_b - C_{ba}(V_a^{-1} - V_a^{-1}\tilde{V_a}V_a^{-1})C_{ab} \end{pmatrix} \tag{2.5}$$

The formulae show that for correlated variables, selection on one variable correspondingly affects the mean and covariance of the other variable such that an identical factor structure can be found for both.

Lawley (1944) generalized this formula to show that these selection effects hold as long as the regression is linear and homoscedastic. Building directly on the work of Lawley (1944), as well as the work by Ahmavaara (1954) on the effects of selection on simple structure of factor loadings, Meredith (1964a,b) demonstrated the invariance of simple structure in subpopulations and methods for establishing such invariance in the framework of factor rotation. Importantly, Meredith (1964a) noted that the selection variable need not be known in order for these theorems to hold; we can select random samples and expect to find an invariant solution across groups if a simple structure exists in any one sample.

Exploratory factor analysis (EFA) as a method preceded CFA; the relevant groundwork for invariance was, therefore, first considered within the framework of EFA (e.g. Ahmavaara, 1954; Meredith, 1964a,b). Establishing invariance via EFA involves attempts to rotate factor loadings towards a target that is invariant across groups, but does not allow for explicitly constraining parameters to invariance (Estabrook, 2012; Thompson and Daniel, 1996). In 1980, Darton (1980) explicated contemporary methods of rotation of what we now consider exploratory factor models for confirmatory purposes. However, concerns regarding EFA will not be considered further in this paper, as EFA is not adequate for investigating issues of invariance while CFA is a preferable method for addressing invariance concerns (Alwin and Jackson, 1981; Schmitt and Kuljanin, 2008).

2.1 Invariance Concerns

Factorial invariance, and therefore measurement invariance, is a necessary condition to allow group comparisons (e.g. Cheung and Rensvold, 1999; Meredith, 1993; Schmitt and Kuljanin, 2008; Vandenberg and Lance, 2000). Factorial invariance is rooted in construct validity, in that invariance of latent factors is a crucial step in establishing construct validity–that one is measuring what one purports to measure–across groups. This definition makes clear that construct validity is vital in making group comparisons, as there is no possible direct comparison if we cannot establish that we have measured the same, relevant construct in both groups.

In their 1955 article, Cronbach and Meehl synthesized recent APA developments that had been made in psychological testing, including the genesis of the term "construct validity," and wrote that identification of the latent construct rests in the interpretation of the relationships between the manifest variables. Factorial invariance is identified as a set of constraints on those relationships; it is specific to factor models in that it is used to denote invariance of parameters in a factor model. Measurement invariance, on the other hand, is the invariance of the relationships between manifest variables and the latent factors that they measure. Measurement invariance has been considered to be requisite in order to establish factorial invariance, in that factors are measured through manifest variables and therefore the measurement of a factor must be invariant in order for the factor itself to be considered invariant (e.g. Gregorich, 2006; Bollen and Lennox, 1991). Metric invariance refers to equivalence of parameter values, typically factor loadings, which is a feature of strict invariance (e.g. Borsboom, 2006; Dimitrov, 2010; Estabrook, 2012; Meredith, 1993; Meredith and Teresi, 2006).

Cronbach and Meehl (1955) stressed the importance of model falsifiability in what was ultimately the conceptual precursor to confirmatory factor analysis (CFA). The work of Jöreskog in 1970 proposed a method of testing constraints on variance-covariance matrices, laying the groundwork for modern structural equation modeling which permits a more controlled test for factorial invariance via CFA. In CFA, the theoretical/hypothetical model of interest is only one of many possible models (Thompson and Daniel, 1996). I will not discuss falsifiability in EFA, as rotation creates specific falsifiability concerns. However, in CFA, testing rival hypotheses has historically been an overlooked step (Schmitt and Kuljanin, 2008; Vandenberg and Lance, 2000), yet is crucial in establishing the validity of the model of interest and any related inferences (Thompson and Daniel, 1996). Researchers are recommended to carefully consider and test other plausible models beyond null models, which are sufficient for establishing fit but often are not sufficiently plausible rivals to the theoretical model of interest. Model selection, including testing of rival models, should be guided and informed by theoretical concerns and not solely the statistical results (e.g. Jöreskog, 1971; Schmitt and Kuljanin, 2008; Thompson and Daniel, 1996).

Thompson and Daniel (1996) note that model invariance should not be taken to mean that all parameters must be invariant. The researchers suggest factor correlations as an example of parameters that need not be invariant across samples because most theories do not require that factor covariances be invariant. They suggest that if a test consistently "marks" the constructs of interest, then the expectation of invariance of factor covariances would only exist if it were required by the theory or by the score use. This can be ambiguous, but is testable. Current methods allow for factorial invariance and group comparisons when factor covariances are non-invariant; methods for instances where factor covariances are invariant while the factor measurement is non-invariant have not yet been fully developed, and are the motivation for the current study.

Full measurement invariance (MI) is, however, uncommon, and many models identify some level of partial invariance (e.g. Schmitt and Kuljanin, 2008; Byrne, 2003). Researchers generally strive to fit a fully measurement invariant model, find some item measurement non-invariance, and then continue with their analyses under partial measurement invariance (e.g. Byrne, 2003; Drapeau et al., 2010; Schlotz et al., 2011). Based on the finding of Millsap and Kwok (2004) that selection based on manifest variables as an approximation for latent factor scores resulted in only minor effects on selection errors might suggest that such a course of action could be sufficient. However, in cross-cultural studies, selection is often on extraneous factors that could be related to the factors or items or instrument, each potentially introducing accompanying bias.

2.2 Issues in Practice

Comparing scores between groups, regardless of whether the groups are longitudinally or cross-sectionally derived, requires establishing measurement invariance (Cheung and Rensvold, 1999; Schmitt and Kuljanin, 2008; Widaman and Reise, 1997; Steenkamp and Baumgartner, 1998). Measurement invariance is, therefore, a crucial step in cross-cultural research, as instruments can vary in measurement qualities and validity across cultures. Strict measurement invariance involves establishing configural invariance, and so the location of zero loadings is invariant across the groups. However, strict measurement invariance is not always useful in allowing cross-cultural comparisons. Configural invariance may hold despite non-invariant items, or measurement non-invariance (MNI), across groups (Byrne, 2003; Steenkamp and Baumgartner, 1998).

Among others, concerns in multigroup research can include issues of bias (Byrne, 2003; Bollen and Lennox, 1991). Byrne (2003) further divide such bias into three sources,

(a) construct, (b) method, and (c) item. Construct-derived bias comes from a difference in meaning of the construct for the different groups, which researchers like Byrne (2003) and Pedraza and Mungas (2008) suggest can result from differences in what is considered appropriate manifestations of the construct and failure of the instrument to address the construct appropriately.

Method bias refers to bias in how the group comparisons are made, and includes instrument bias. Instrument bias occurs when group differences on the manifest variables result from the instrument used to perform the assessment, rather than underlying differences on the construct of interest. As an example, Byrne (2003) describe a computer-based assessment comparing a group whose members tend to have extensive experience with computer usage and testing to a group whose members have little to no experience with computers could produce biased results related to the computer experience rather than to the latent construct (this specific type of instrument bias is called stimulus familiarity).

The third bias type–item bias–encompasses extraneous group differences on items. Similar to construct bias, item bias is often a result of differences in interpretation due to cultural differences. However, item bias refers specifically to differences on individual items–one item can be biased while the rest of the items are not.

Construct, instrument, and item bias are of critical importance in considerations of invariance, as they can directly affect whether or not measurement will be invariant between the groups. However, suggestions for managing these types of variance involve study design; no methods are described for establishing factorial invariance despite bias in the measurement. A strong method for establishing factorial invariance should allow for non-invariant measurement while still distinguishing between construct bias or a lack thereof in the study.

Horn and McArdle (1992) present one of the earlier instances of investigating factor and measurement invariance across age. Consistent with a parameter trade-off between a model with configural invariance and fewer factors compared to one with metric invariance but more factors (e.g. Horn and McArdle, 1992; Jöreskog, 1971), the authors found that measurement was not invariant across ages when they fit simpler one- or two-factor models; however, they found that factor means were non-invariant with age when they implemented a more complex two-factor model. Horn and McArdle note that metric invariance is an ideal but unlikely standard and configural invariance is a tenable but somewhat ambiguous standard. However, they suggest that unity-specified factor patterns could be a sufficiently tenable but simultaneously sufficiently rigorous invariance test.

CFA can be used to examine more complicated models that feature higher-order factors, (e.g. Golay and Lecerf, 2011; Norton et al., 2013). A higher-order factor is a latent variable that loads (regresses) onto another latent variable. For example in Figure 3.1, $F2$ is a second-order factor that loads onto the three first order factors $F1a$, $F1b$, and $F1c$. The fit of a higher-order factor model suggests that the covariances between the first-order factors share some structure that can be represented by another unobserved variable.

The concept of higher-order factor models goes as far back as Thurstone (1931, 1947), in which Thurstone discusses second-order factor analysis, showing that if the first-order factor intercorrelation matrix is linearly dependent (unit rank), the structure may require a second-order factor. Despite this, higher-order factor models are generally less common in the literature, and therefore are less frequently seen in the context of invariance. However, Marsh has devoted time to examining invariance in instruments that do suggest an underlying second-order factor structure, and therefore included higher-order factor models as plausible alternative models for model comparisons (e.g. Marsh and Hocevar, 1985; Marsh, 1985). In these analyses, Marsh (1985) highlights a few important considerations for higher-order models. Higher-order models represent constraints on the covariance matrix of lower-order factors, and so the absolute fit indices for the higher-order model cannot be better than the fit for the lower-order model. Because of this, the higher-order model must be compared with the lower-order model as part of the evaluation of the goodness of fit for the higher-order model Marsh (1985). Factorial invariance still holds the same meaning, with the minimal condition for any level of factorial invariance being (measurement) invariance–invariance of the factor loadings.

Suggestions for analysis of invariance sometimes differ slightly in language, but generally proceed similarly to the pattern established by Jöreskog (1971). Vandenberg and Lance (2000) and subsequently Schmitt and Kuljanin (2008) provide comprehensive overviews of typical procedures and relative differences in procedures. Researchers have made various suggestions for identifying measurement invariance and its effects: e.g. McDonald (2011) who distinguishes between measurement and prediction and proposes distribution-free approach to measurement, and Merkle and Zeileis (2013) who propose likelihood-derivative-based methods for identifying measurement invariance. On the other hand, establishing factorial invariance in conditions of measurement non-invariance has been seen far less attempts in the literature to establish viable methodologies.

In the event that a researcher expects traditional measurement invariance might not be tenable, Nesselroade et al. (2007) suggested a higher-order invariance idiographic filtering (HOIIF) as a method of filtering idiosyncratic differences while still establishing invariant relationships across individuals. In HOIIF, invariance is tested via an invariant second-order factor model, applied at the level of factor intercorrelations, while the first-order factor measurement model is allowed to vary.

Chen et al. (2005) provided examples of procedures for testing measurement invariance in second-order models. As an example of why this might be necessary, they noted that in comparing the heights of Europeans and Americans, despite measurement non-invariance due to a difference in measurement units (centimeters vs inches), the construct being measured is invariant across individuals. In the end, Chen et al. expressed concern about a) choosing a fit statistic and b) what to do when measurement invariance fails. Nesselroade et al. (2007) showed that in situations lacking measurement invariance, HOIIF is a method in which generalizable abstract (latent) constructs can be established within the confines of invariance. In HOIIF, factor invariance is established on the level of factor intercorrelations. This invariance allows for interpretation of the factors as the same variables across individuals, while allowing for individual variation in how the factors are measured. In order to compare relationships of latent variables across groups, invariant relationships must be established between variables across individuals. However, they suggested that this invariance does not need to rely on invariant measurement, despite current tradition. For example, Nesselroade et al. pointed out that despite the fact that the area of a rectangle is measured as width $\times$ height and the area of a triangle is πr^2, the relationship between area and volume is consistent as area $\times$ length (measurement of third dimension). They noted concerns about falsifiability–that, with enough tailoring of first-order factor loadings, researchers could conceivably always find a HOIIF model that fits across all individuals. Therefore, Nesselroade et al. reiterated that in order to implement a HOIIF model, the researcher will need a strong theoretical basis for supporting identification of the latent variables.

Zhang et al. (2011) further investigated the use of HOIIF in factor analysis. Some researchers will want to be able to test for the same latent constructs even when different measurements or instruments are used, and Zhang et al. provided the example of measuring pigeons pecking compared to rats pressing a bar in the study of reinforcement in operant conditioning. In this case, measurement is not invariant, but the reinforcement and extinction reflected are the same latent concepts. Zhang et al. noted that HOIIF is a potential method for dealing with such situations. Importantly, Zhang et al. (2011) noted that first-order factor loadings must be rotated in some meaningful way because there may exist an arbitrary rotation that will satisfy the requirement of an invariant first-order factor correlation matrix. For this reason, the authors suggested use of one or more marker variables to identify each factor. The marker variables would uniquely identify each factor and must be common across individuals, but need not have the same loading value for each individual. All other variables could be freely estimated. The authors recommended that marker variables load greater than 0.5, and substantial loadings on marker variables would support identification of the factors, while the loadings of the remaining variables would help the researcher to understand individual differences.

In order to determine model fit, Zhang et al. (2011) tested a non-invariant model against a saturated model, and then an invariant model was tested against the non-invariant model. The saturated model is a model of the data covariance matrix and means, the non-invariant model is a first-order factor model with non-invariant factor correlation matrices, and the invariant model is a second-order factor model with non-invariant first-order factor loadings and an invariant second-order factor model replacing the first-order factor correlations. Zhang et al. found that the non-marker variables helped with the accuracy of the factor correlation estimates to some extent; however, they found diminishing returns after a point. They also found that the level of communalities of marker variables is important in the accuracy of the factor correlation estimates, and that a likelihood ratio test is sufficient to detect whether a HOIIF model is, in fact, invariant.

Important steps have been taken in examining the viability of the HOIIF method. For instance, Molenaar and Nesselroade (2012) applied HOIIF in the study of processes using dynamic factor analysis. However, to date, the literature on the application of the HOIIF method has not established how well the method recovers first-order factor loadings in a confirmatory factor analysis framework. Likewise, there has not been an examination of the performance of the method in different levels of first-order configural and metric (in)variance. In the current study, I take steps to address these gaps.

2.3 Summary

Factorial invariance emerged in early considerations of measuring latent factors and methods of investigating it continue to be refined. The importance of factorial invariance in group comparisons, such as in cross-cultural and longitudinal studies, has been linked to measurement invariance; however, these are particular conditions in which measurement invariance may not be tenable. Methods for establishing factorial invariance in conditions of measurement non-invariance are currently still in development, and the current study further develops one such method–the higher order invariance idiographic filter (HOIIF). I use simulated data to investigate the performance of the HOIIF method, and follow with an example of the method in one empirical cross-sectional data set as well as one empirical longitudinal data set.

3. Simulations

3.1 Introduction

In order to investigate the utility and performance of a higher-order invariance idiographic filter (HOIIF) in multiple-group latent factor methods, I analyzed both simulated as well as observed data that result from two conditions (cross-sectional vs. longitudinal) in which measurement non-invariance can arise. A HOIIF confirmatory factor model involves imposing invariance at a higher-order than in a traditional confirmatory factory model. Suppose we have *p* persons measured on *n* variables and *k* common factors. Consider the common factor model,

$$Y = \Lambda F + E \tag{3.1}$$

where Y is an n $\times$ p matrix of observations, Λ is an n $\times$ k matrix of factor loadings, F is a k $\times$ p matrix of factor scores, and E is an n $\times$ p matrix of errors. We can replace the factor scores, F, with Y^* where

$$Y^* = \Lambda^* F^* + E^* \tag{3.2}$$

in which Y^* is a k $\times$ p matrix of first-order factor scores, Λ^* is a k $\times$ k* matrix of (second-order) factor loadings, F^* is a k$^*\times$ p matrix of (second order) factor scores, and E^* is a k $\times$ p matrix of (second order) errors. The first order model now contains the second order model:

$$Y = \Lambda Y^* + E \tag{3.3}$$

With higher-order invariance, invariance is imposed on Y^* instead of Λ, allowing the loadings, Λ, to vary across groups.

In the current studies, I applied a HOIIF to confirmatory factor models in order to probe the performance of the method; the analyses for each data type (simulated vs observed) are presented as two separate studies. In the first study, the performance of a HOIIF in cross-sectional and longitudinal multiple-group latent variable data, the performance of the HOIIF was investigated by implementing a Monte Carlo simulation analysis of both cross sectional and longitudinal simulated data. This study was followed by applications of the HOIIF to cross sectional as well as longitudinal observed data.

3.2 Methodology

While mathematical proofs provide the groundwork for bringing a method from theory to application, the actual implementation of a method relies on some type of estimator in order to obtain parameter values. Although the estimator might have a known asymptotically normal distribution, the finite sampling distribution of the estimator and its related parameters may be unknown. In the current study, full information maximum likelihood (FIML) estimation was used to estimate confirmatory factor models of varying structures. FIML uses raw data in addition to the observed covariance matrix, and is therefore a preferred estimator for data sets with missing values. While the data simulated in these studies do not have missing values, many observed data sets in psychology–including one of the data sets used following the simulations–contain missing values, for which reason FIML estimation would be preferred. Because the finite sampling distribution for FIML in this context is unknown and the method is still in its early stages, the performance of a HOIIF method in conditions of varying sample size, measurement model, and higher-order equivalence requires investigation. Therefore, I simulated data sets with different population parameter conditions in order to analyze the performance of a higher-order invariance idiographic filter (HOIIF) in two multiple group, latent variable data types: 1) cross-sectional data and 2) longitudinal data.

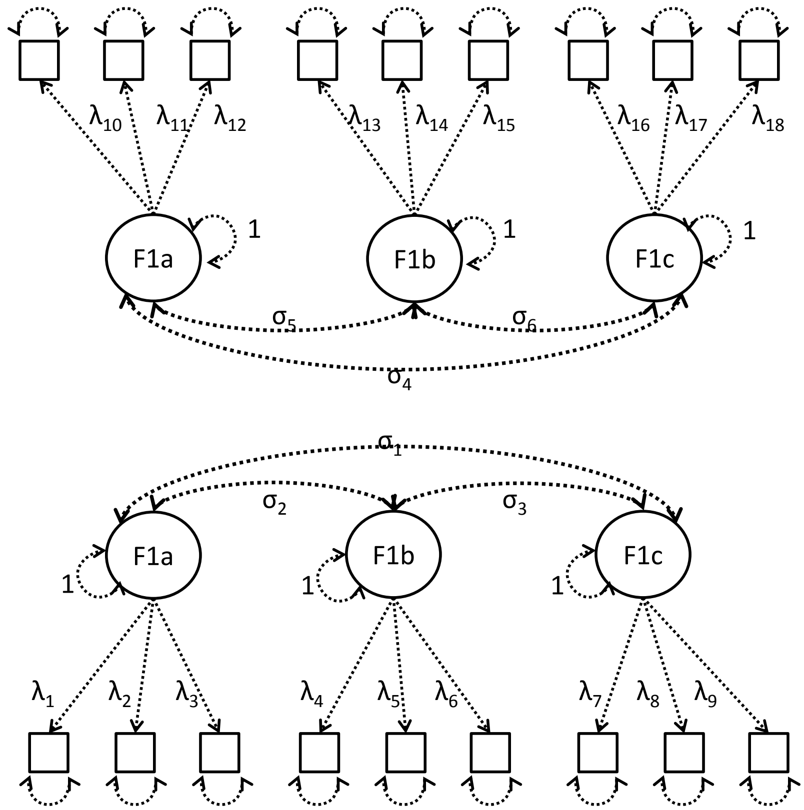
In order to implement a HOIIF in a confirmatory factor model in two group, cross-sectional data, I followed the stepwise procedure identified in Zhang et al., 2011 and utilized in Dodson et al., 2014. Model fit was assessed via a likelihood ratio test comparison with a null model of the data variances, covariances, and means, which is the best fitting model of the data (see Bentler and Bonett, 1980, for a review of saturated and nested model comparisons). Although the stepwise implementation of the higher-order invariance method requires assessment of fit at each step in order to determine whether or not to proceed, I proceeded through all steps regardless of fit for the sake of gathering a maximum of data regarding the method's performance. The specific model varied between the cross-sectional and longitudinal data styles; however, the basic procedure was as follows:

1. In the first step, I fitted a three-factor, first order model that had no across-group constraints on the estimation of the parameters (Figure 3.1). Factor variances were fixed to one in order to facilitate comparison of the estimated loadings with the values that were used to create the simulated data.

2. Next, I fitted a model in which the (first-order) factor covariance matrices were constrained to equality across groups (Figure 3.2). This model featured no across-group constraints on the first-order factor loadings nor on the errors of the manifest variables.

3. In the next step, I moved from a first-order model to a second-order model. In this model, covariances between the first-order factors were removed; instead, one second-order factor loaded onto the three first-order factors (see Figure 3.3). Therefore, this model has a first-order measurement model, in which the first-order factors load onto the manifest variables, and a second-order "measurement" model, in which the second-order factors load onto the first-order factors. The first-order factors were

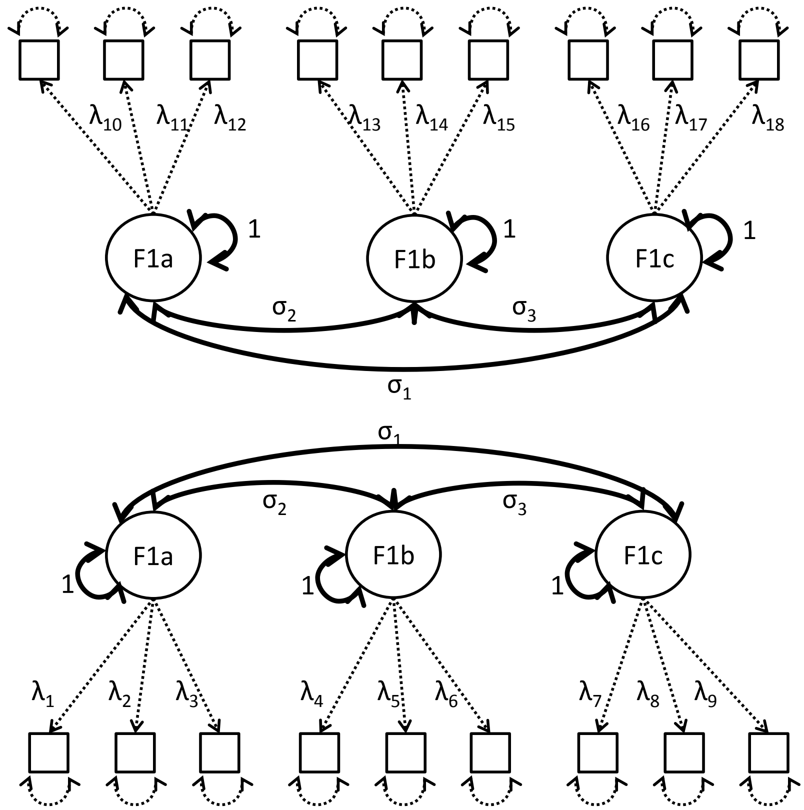
Figure 3.1. Three-factor, first-order model that has no across-group constraints on the estimation of the parameters.



scaled by fixing one loading on each factor to its population value, as fixing the first-order factor variances to 1 would be a constraint on the "item" errors of the second-order "measurement" model. This second-order model imposes configural invariance in the second-order "measurement" model, but is otherwise unconstrained across groups.

4. From the configural-invariance second-order model, I proceeded to fit nested models with increasing equality constraints imposed across groups on the second-order "measurement" model. At each level, constraints from the previous level were maintained and additional constraints were imposed. From the original second-order model, featuring configural invariance in the second-order "measurement" model, additional constraints were added in the following order:

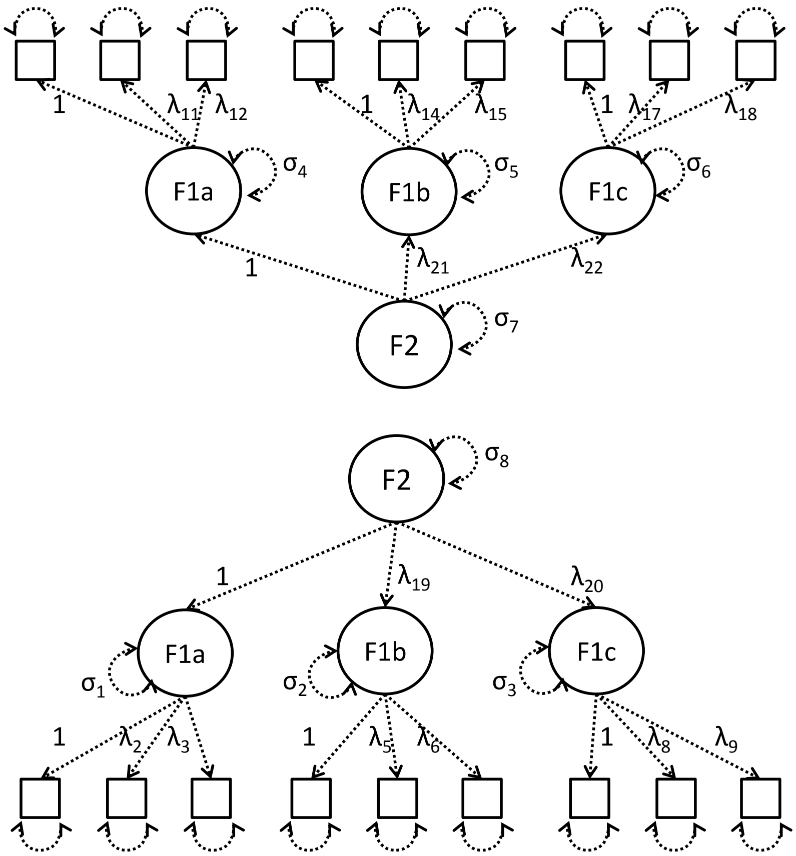
Figure 3.2. Three-factor, first-order model with constrained factor covariance matrix.



- second-order factor loadings (Figure 3.4)
- first-order factor variances (Figure 3.5)
- second-order factor variances and covariance (Figure 3.6)

I limited the complexity of the analyses to sample size, type of first order (non-) invariance, and higher order invariance vs non-invariance (detailed in the next section). From the analyses, I present information about the parameter estimates and fit statistics, and the Expected Average Squared Error (EASE, Zhang et al., 2011) for the estimated parameters. The results of the simulations are discussed based on their recovery of the first-order factor loadings and expected error, their ratios of rejected to non-rejected first-order constrained models, and their ratios of rejected to non-rejected higher-order constrained models. Specifically, with these results for the simulations, I address the following consid-

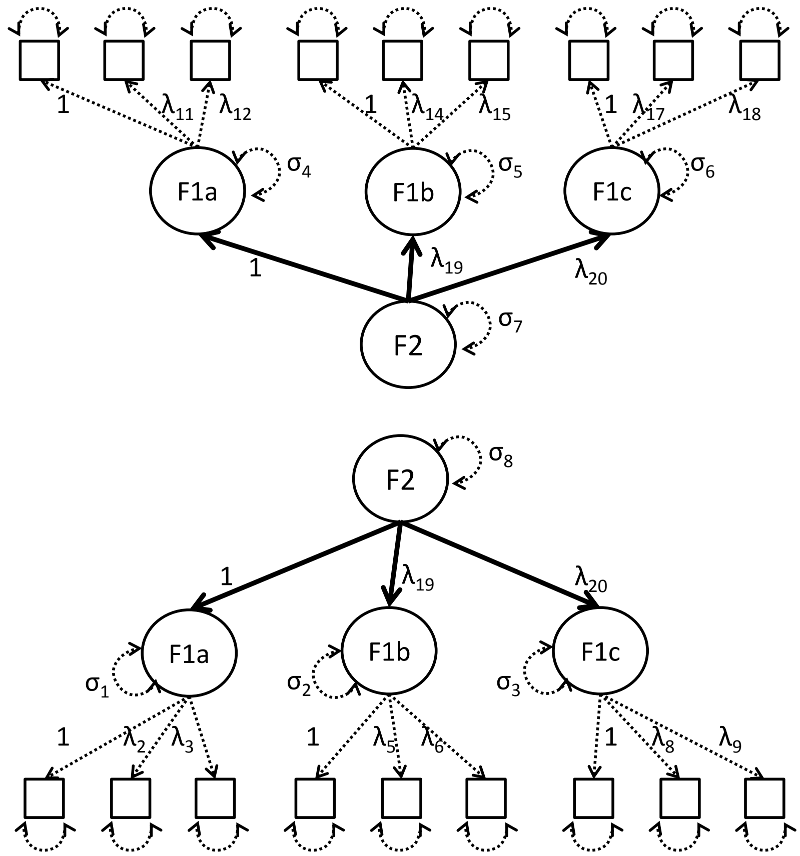
Figure 3.3. Three-factor, second-order model with second-order configural invariance and no across-group constraints.



erations:

- How well does this method recover parameters? In other words, how often and how closely do the point estimates match the parameters used to create the data?
- How much bias is present in the estimates?
- How specific is it? How often does it reject the models it should reject and fail to reject the models it should fail to reject?
- In what ways do the relevant fit statistics vary relative to accuracy of point estimates, bias, and specificity?
- How does the structure of the measurement model affect the previous considerations?

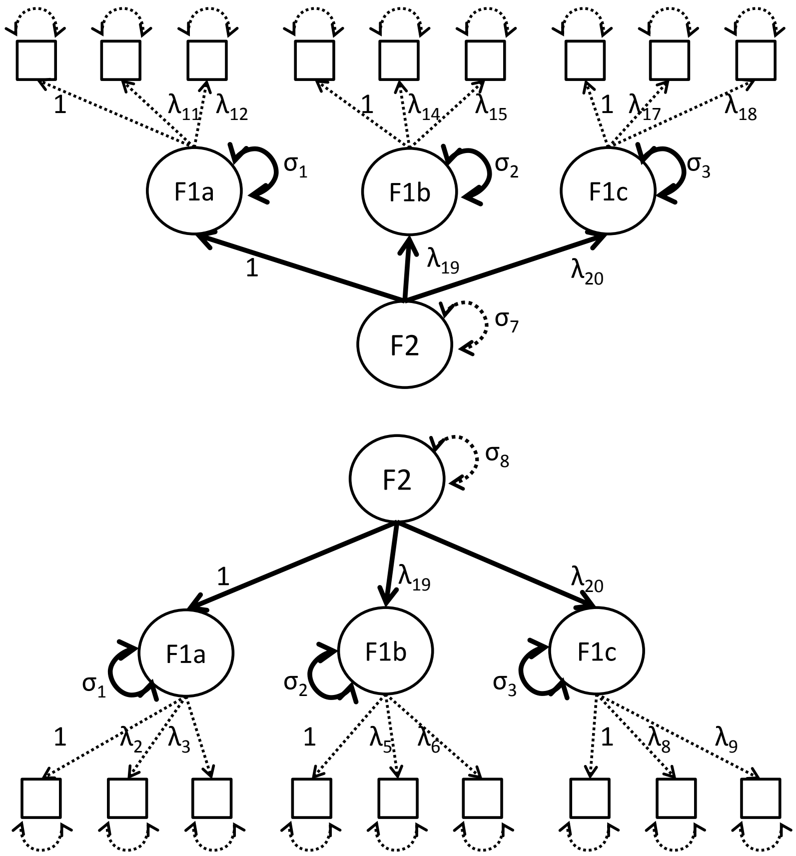
Figure 3.4. Three-factor, second-order model with constrained second-order factor loadings.



- How does sample size affect the previous considerations?
- How often does the method fail to converge?
- Do convergence issues depend on sample size, measurement model, or the second-order variance condition?
- Does varying the starting values affect convergence?

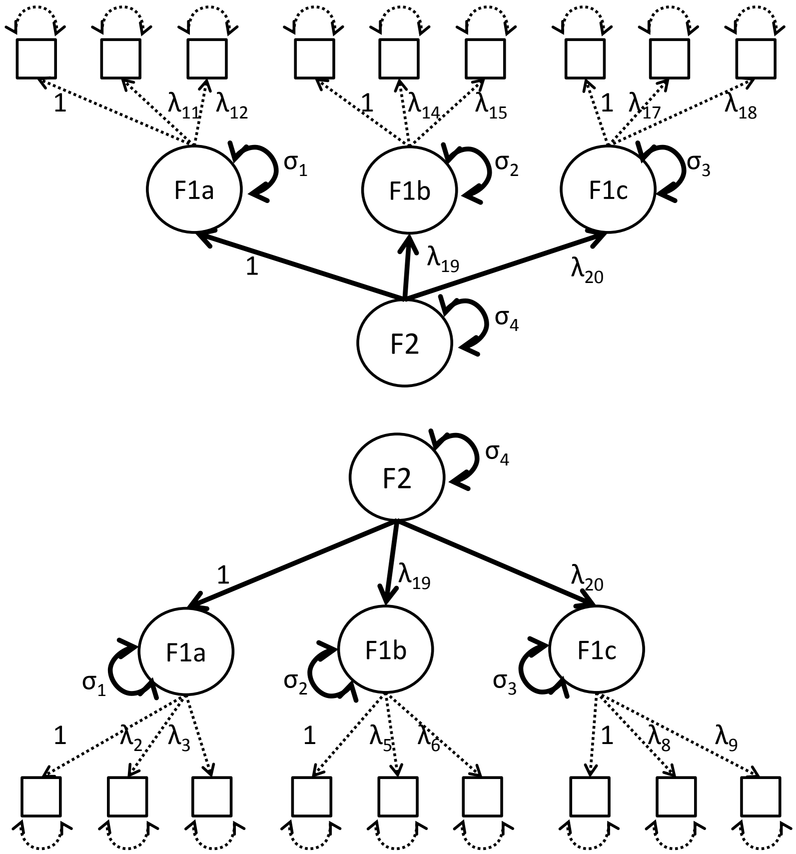
With these considerations in mind, following the analysis of the simulated data sets, I applied the method to two observed data sets as an example application of the method. The decision of how to proceed was based on the sensitivity and specificity of the method, as well as the interpretation of the results. Adequate sensitivity/specificity was defined by

Figure 3.5. Three-factor, second-order model with constrained second-order factor loadings and first-order factor variances (which serve as second-order "measurement errors").



two conditions, 1) the model fails to be rejected in 75% of the instances where the second-order model is invariant, and 2) the model is rejected in 75% of the instances where the second-order model is non-invariant. I then chose from one of the following mutually exclusive possibilities to continue the analysis (see 3.1): 1) If the results showed adequate specificity and sensitivity, I would use the HOIIF in the applied data and will test a traditional method as a plausible rival hypothesis. 2) If the results showed adequate specificity but inadequate senstitivity, I would compare the HOIIF with a traditional method in order to illustrate the strengths, weaknesses, and implications of both methods. 3) If the results showed inadequate specificity but adequate sensitivity, I would again compare the HOIIF with a traditional method. 4) If the results showed inadequate specificity and inadequate senstitivity as well, I would attempt to determine whether this reflects a problem of the

Figure 3.6. Three-factor, second-order model with constrained second-order factor variance.

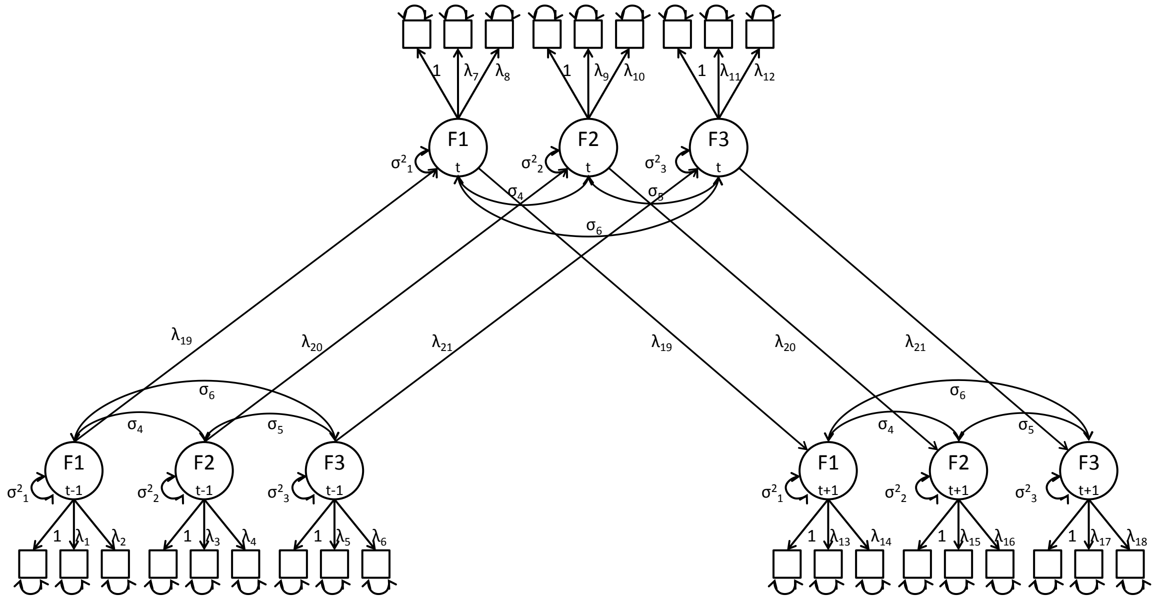


theory, the data, or some other issue. Note that this last outcome was unlikely, based on prior research (Zhang et al., 2011; Dodson et al., 2014).

Table 3.1. Decision for applied analysis, based on simulation-indicated sensitivity and specificity of the HOIIF

	Sensitive	
Specific	*Yes*	*No*
Yes	Use, plausible rival	Use, compare
No	Use, compare	Troubleshoot, discuss

Figure 3.7. Three-factor, longitudinal model. For HOIIF model, across-time longitudinal regressions are constrained to equality across groups while the first order factor model is free to vary.



3.2.1 Cross-Sectional Simulation

The power of a HOIIF to identify varying second-order factor structures in cross sectional data was investigated with a Monte Carlo simulation study. The data were simulated to feature nine manifest (“observed”) variables, three first-order factors, one second-order factor, and two subgroups (e.g. Figure 3.3). In the simulations, the sample size (3 levels: small–$N = 100$, medium–$N = 500$, large–$N = 1500$), measurement model (2 levels of difference between groups: metric, configural), and second-order factor model (2 levels: metrically invariant second-order model, metrically non-invariant second-order model) differed between two groups. This set resulted in 12 different data set types, identified as subsets of their measurement models (see Table 3.2): 1 - 3) metrically non-invariant with non-invariant second-order model (small, medium and large sample size), 4 - 6) metrically non-invariant with invariant second-order model (small, medium, and large sample size), 7 - 9) configurally non-invariant with non-invariant second-order model (small, medium, and large sample size), and 10 - 12) configurally non-invariant with invariant second-order model (small, medium, and large sample size). For each data set type, the data were simulated and analyzed in 1000 replications.

Table 3.2. Cross-Sectional Data Set Types

Measurement Model	S-O Model	Sample Size
Metrically Non-Invariant	Non-Invariant	1. Small (100)
		2. Medium (500)
		3. Large (1500)
	Invariant	4. Small (100)
		5. Medium (500)
		6. Large (1500)
Configurally Non-Invariant	Non-Invariant	7. Small (100)
		8. Medium (500)
		9. Large (1500)
	Invariant	10. Small (100)
		11. Medium (500)
		12. Large (1500)

3.2.2 Longitudinal Simulation

Unlike cross-sectional data, in longitudinal data, a latent variable can have a directional effect on itself (over time). Such models can be critical in properly representing the dynamics of human development in latent constructs; however, in such a longitudinal context, the relationship between latent variables can be more complicated than in a cross-sectional context. Because these models are important but represent a more complicated model of relationships between latent variables, I directly investigated the performance of the HOIIF in a longitudinal framework as well.

As in the cross-sectional simulation, the power of a HOIIF to identify varying second-order factor structures in longitudinal data was investigated with a Monte Carlo simulation study. The data again were simulated to feature nine manifest ("observed") variables, three first-order factors, and two subgroups (see Figure 3.7).

The degree to which the model differed between groups was varied. In the simulations, the sample size (3 levels: small–$N = 100$, medium–$N = 500$, large–$N =$

1500), measurement model (2 levels of variance between groups: metric, configural), and second-order factor model (3 levels: metrically non-invariant "second-order" model–autoregressions and factor covariances are non-invariant across time, metrically invariant "second-order" model–autoregressions and factor covariances are invariant across time, and a mixed "second-order" model–covariances are non-invariant while autoregressions are invariant across time) differed between the two groups. As in the cross-sectional simulation, this set of data type variables resulted in 18 different data set types, identified as subsets of their measurement models (see Table 3.3): 1 - 3) metrically non-invariant with non-invariant second-order model (small, medium and large sample size), 4 - 6) metrically non-invariant with invariant second-order model (small, medium, and large sample size), 7 - 9) metrically non-invariant with mixed second-order model (small, medium, and large sample size), 10 - 12) configurally non-invariant with non-invariant second-order model (small, medium, and large sample size), 13 - 15) configurally non-invariant with invariant second-order model (small, medium, and large sample size), and 16 - 18) configurally non-invariant with mixed second-order model (small, medium, and large sample size).

For each data set type, the data were simulated and analyzed in 1000 replications. The data were analyzed with a confirmatory dynamic factor model (Molenaar and Nesselroade, 2012; Nesselroade, 2001; Nesselroade and Molenaar, 2003), from which the distribution of parameter estimates and fit statistics, and the Expected Average Squared Error (EASE, Zhang et al., 2011) for the estimated parameters are presented in the same format as in the case of the cross-sectional data.

I calculated an average squared error (ASE) and estimated an expected average squared error (EASE) based on the formulas used by Zhang et al. (2011): I calculated an ASE for the factor loadings of each replication, with k estimated factor loadings in each replication, λ_{ij}^{r} denoting the i-th loading on the j-th factor for the r-th replication while λ_{ij}^{o} denotes the i-th loading on the j-th factor in the population values that were used to create

Table 3.3. Longitudinal Data Set Types

Measurement Model	S-O Model	Sample Size
Metrically Non-Invariant	Non-Invariant	1. Small (100)
		2. Medium (500)
		3. Large (1500)
	Invariant	4. Small (100)
		5. Medium (500)
		6. Large (1500)
	Mixed	7. Small (100)
		8. Medium (500)
		9. Large (1500)
Configurally Non-Invariant	Non-Invariant	10. Small (100)
		11. Medium (500)
		12. Large (1500)
	Invariant	13. Small (100)
		14. Medium (500)
		15. Large (1500)
	Mixed	16. Small (100)
		17. Medium (500)
		18. Large (1500)

the data, so that

$$ASE_r = \frac{1}{k(p)} \sum_{i=1}^{k} \sum_{j=1}^{p} (\lambda_{ij}^r - \lambda_{ij}^0)^2 \tag{3.4}$$

and

$$EASE = \frac{1}{R} \sum_{r=1}^{R} ASE_r \tag{3.5}$$

Here, the EASE is the average of all of the ASEs.

To control for the influence of outliers, as an average of 9% of models still suffered from convergence issues and gave spuriously extreme results, I based model fits on the middle 95% of the data and excluded the lower and upper 2.5%.

3.3 Results

3.3.1 Cross-Sectional Model Parameter Estimation

First-order parameters

Mean first-order factor loading estimates, EASEs, and EASE 95% confidence intervals for the cross-sectional data are presented by group in appendix A Tables A.1 and A.2 for the first-order model with no constraints across the groups (FO Unconstrained), and Tables A.3 and A.4 for the final second-order model with constrained second-order factor loadings and variance and constrained first-order factor variances (SO Constrained). Because in the second-order models, first-order measurement was scaled by fixing one first-order factor loading to a value of 1 for each factor, the estimated loading values needed to be standardized after model fitting in order to make them comparable to the population values used to simulate the data. The first-order factor loading values presented in appendix A Tables A.3 and A.4 are standardized values based on the estimated loading-scaled values, factor variances, and item variances.

For all data types and in both the FO Unconstrained as well as the SO Constrained models, error in estimated first-order loading values decreased as sample size increased, as evidenced by the decrease in the EASE across sample sizes. The EASE is in variance units; therefore, an EASE of 0.1 indicates an average deviation of approximately 0.32, an EASE of 0.01 indicates an average deviation of 0.1, and an EASE of 0.001 indicates an average deviation of approximately 0.03. Most of the data types in both FO Unconstrained as well as SO Constrained models had an EASE of close to 0.01 at sample size 500, so at 500 observations, loading estimates for the HOIIF average were within approximately +/- 0.1 of the population values (see Tables 3.4, 3.5, and 3.6 for population loading values). On the other hand, at 1500 observations the EASEs range from 0.013 to 0.002 for the FO Unconstrained models, which indicates average deviations of +/- 0.11 to 0.04, and from 0.007 to 0.003 for the SO Constrained models, which indicates average deviations of +/-

0.08 to 0.05. EASE 95% CIs were smaller on average for the SO Unconstrained models than the FO Constrained models.

Table 3.4. Cross-sectional Simulations: Population factor loading values for group 1, in all data types

	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}
x1	0.37		
x2	0.48		
x3	0.64		
x4		0.59	
x5		0.34	
x6		0.81	
x7			0.46
x8			0.31
x9			0.60

Table 3.5. Cross-sectional Simulations: Population factor loading values for group 2, in data with metrically invariant first-order measurement models

	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}
x1	0.66		
x2	0.78		
x3	0.30		
x4		0.66	
x5		0.60	
x6		0.51	
x7			0.70
x8			0.80
x9			0.36

Second-order parameters

For the second-order "measurement" model in the second-order model with constrained second-order factor loadings and variance, and constrained first-order factor variances (SO Constrained), mean second-order factor loading estimates, EASEs, and EASE 95% confidence intervals for the cross-sectional data are presented in appendix A Table A.5. Because the second-order models were scaled by fixing one factor loading to a

Table 3.6. Cross-sectional Simulations: Population factor loading values for group 2, in data with configurally invariant first-order measurement models

	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}
x1	0.66		
x2		0.78	
x3			0.30
x4	0.66		
x5		0.60	
x6			0.51
x7	0.70		
x8		0.80	
x9			0.36

value of 1 for each factor, the estimated loading values needed to be standardized after model fitting in order to make them comparable to the population values used to simulate the data. Therefore, the factor loading values presented in appendix A Table A.5 are standardized values based on the estimated loading-scaled values, second-order factor variance, and first-order factor variance. The data with invariant second-order models have EASE and 95% CI relative to only the single set of population values, while the data with non-invariant second-order models have EASE and 95% CI to both the group 1 and group 2 population values.

For all data types, error in estimated loading values decreased as sample size increased, as evidenced by the decrease in the EASE across sample sizes. While mean estimates deviated less than 0.05 at sample size 100 for models of data with invariant second-order population structure, at sample sizes 500 and 1500, mean estimates were within 0.02 of the population values and had EASEs of 0.007 - 0.002. On the other hand, unsurprisingly, mean estimates for data with non-invariant second-order population structures were biased towards the mean of the two population values. At the large sample size, the EASE for these estimates approached 0.01 but were not less than 0.01, and the values remained biased.

3.3.2 Cross-Sectional Model Fit Information

Table 3.7 presents mode (50th percentile) p values and skew of the p values for three model types–the first-order model with no constraints (FO Unconstrained), the first-order model with factor variance/covariance matrix constrained across groups (FO Constrained), and the second-order model with constraints on the second-order factor loadings, second-order factor variance, and first-order factor variances (SO Constrained). The p values result from χ^2 likelihood ratio tests comparing the model of interest to either a) a model of variances, covariances, and means for the data set (referred to as the "null" model comparison), or to b) the previous model, which has fewer constraints/more degrees of freedom (referred to as the "previous" or "prev" model comparison). In the case of the FO Constrained model (a first-order model with the factor variance/covariance matrix constrained across groups), the associated "prev" model is the FO Unconstrained model which has no across-group constraints. In the case of the SO Constrained model (a second-order model with constrained second-order factor loadings and variance, and constrained first-order factor variances), the associated "prev" model is the second-order model with constraints on only the second-order factor loadings and the first-order factor variances. Thus the model comparisons displayed are as follows:

- The first-order factor model with no constraints (FO Unconstrained) against the null variance/covariance model
- The first-order factor model with constraints on the factor variances and covariances (FO Constrained) against the previous F-O Unconstrained model in a nested model comparison
- The FO Constrained model against the null variance/covariance model
- The final second-order factor model with constraints on the second-order factor loadings and variance, and first-order factor variances (SO Constrained) against the pre-

vious second order model with constraints on only the second-order factor loadings and first-order factor variances in a nested model comparison

- The SO Constrained model against the null variance/covariance model

Table 3.7. Cross-sectional Simulations, model fit information: 50th percentiles (i.e. modes) and associated distributional skews for p values of likelihood ratio tests. Rows are the target data types and columns are the relevant model comparisons. Square brackets "[ ]" contain expected mode given data type.

Size	1. FO Unconstr$_{null}$	2. FO Constr$_{prev}$	3. FO Constr$_{null}$	4. SO Constr$_{prev}$	5. SO Constr$_{null}$
Measurement Model: metrically non-invariant, S-O Model: non-invariant [$< .05$]					
100	0.118 (1.578)	0.322 (1.285)	0.132 (1.452)	0.383 (0.354)	0.167 (1.148)
500	0.373 (0.612)	0.05 (1.703)	0.202 (1.129)	0.368 (0.377)	0.054 (1.252)
1500	0.425 (0.564)	0 (0.999)	0.029 (1.607)	0.156 (1.347)	0 (0.601)
Measurement Model: metrically non-invariant, S-O Model: invariant [$> .05$]					
100	0.22 (1.029)	0.535 (0.246)	0.249 (0.878)	0.307 (0.649)	0.225 (0.822)
500	0.484 (0.098)	0.478 (0.232)	0.472 (0.214)	0.159 (1.315)	0.142 (1.146)
1500	0.478 (0.167)	0.403 (0.948)	0.365 (1.041)	0.019 (1.327)	0.003 (1.171)
Measurement Model: configurally non-invariant, S-O Model: non-invariant [$< .05$]					
100	0.063 (2.095)	0.537 (0.348)	0.169 (1.517)	0.323 (0.669)	0.207 (0.938)
500	0.33 (0.891)	0.133 (1.781)	0.261 (0.681)	0.289 (0.834)	0.036 (1.26)
1500	0.429 (0.253)	0 (1.242)	0.041 (1.622)	0.224 (1.092)	0 (0.49)
Measurement Model: configurally non-invariant, S-O Model: invariant [$> .05$]					
100	0.204 (1.133)	0.565 (0.106)	0.25 (0.862)	0.333 (0.603)	0.202 (0.941)
500	0.373 (0.469)	0.513 (0.07)	0.395 (0.334)	0.129 (1.321)	0.108 (1.229)
1500	0.486 (0.079)	0.435 (0.535)	0.462 (0.292)	0.005 (1.116)	0.001 (0.925)

For the same models and model comparisons, Table 3.8 presents the percent of analyses in which the likelihood ratio p value was non-significant (where $p > 0.05$), indicating that the model of interest did not fit significantly worse than the comparison model. For these calculations, the distribution of p values was reduced to the middle 95% in order to control for spurious extremes. Model comparisons displayed were the same comparisons listed above.

Table 3.8. Cross-sectional Simulations, model fit information: Percent of models with non-significant likelihood ratio test comparisons ($p > .05$). Rows are the target data types and columns are the relevant model comparisons. Square brackets "[ ]" contain expected percent given data type.

Size	1. FO Unconstr$_{null}$	2. FO Constr$_{prev}$	3. FO Constr$_{null}$	4. SO Constr$_{prev}$	5. SO Constr$_{null}$
Measurement Model: metrically non-invariant, S-O Model: non-invariant [$\% \to 0$]					
100	0.64	0.70	0.63	0.90	0.77
500	0.86	0.51	0.75	0.89	0.51
1500	0.90	0.11	0.41	0.69	0.00
Measurement Model: metrically non-invariant, S-O Model: invariant [$\% \to 1$]					
100	0.75	0.83	0.78	0.86	0.81
500	0.94	0.95	0.97	0.71	0.71
1500	0.93	0.79	0.78	0.33	0.13
Measurement Model: configurally non-invariant, S-O Model: non-invariant [$\% \to 0$]					
100	0.54	0.76	0.66	0.87	0.79
500	0.75	0.66	0.85	0.83	0.44
1500	0.86	0.15	0.46	0.78	0.00
Measurement Model: configurally non-invariant, S-O Model: invariant [$\% \to 1$]					
100	0.72	0.86	0.82	0.87	0.82
500	0.84	0.88	0.85	0.65	0.66
1500	0.97	0.92	0.93	0.21	0.06

Mode p values for the FO Unconstrained null-model comparisons ranged from 0.063 to as high as 0.486. The lowest mode p value of 0.063, close to the α level of 0.05, was found for the data types with configurally non-invariant first-order measurement structure and non-invariant second-order structure. However, all data types had mode p values greater than 0.4 for large sample sizes in the FO Unconstrained null-model comparison. Reflecting this result, between 54% and 97% of the FO Unconstrained p values were greater than 0.05, meaning that more than half of all of the FO Unconstrained models would have been evaluated as fitting the data not statistically differently from the null, and therefore would not be rejected, regardless of the underlying population structures.

For both the FO Constrained previous-model comparison and the FO Constrained null-model comparison, at each sample size mode p values were lower for the data types with populations featuring non-invariant second-order structures than they were for the corresponding data featuring invariant second-order structures; however, they were only at a level that would result in rejecting models for data with non-invariant second-order population structure and failing to reject models with invariant second-order population structure for the large sample sizes. At a small sample size, the mode p values were above 0.3 for all data types in the previous-model comparisons, and were between 0.132 and 0.250 for null-model comparisons. At a medium sample size in the previous-model comparisons, the mode p values were further apart although still not discriminant–modes of 0.05 (meeting α level) and 0.133 were found for data with populations featuring non-invariant second-order structure and first-order measurement that was either metrically non-invariant or configurally non-invariant, respectively. On the other hand, in the previous-model comparisons at a medium sample size, the mode p values for data with populations featuring invariant second-order structure and first-order measurement that was either metrically or configurally non-invariant were 0.478 and 0.513, respectively–well over the α significance level of 0.05. While for the null-model comparisons at a medium sample size, the mode p values ranged from 0.202 to 0.472–values were lower for the data with non-invariant second-order populations than those with invariant second-order populations, but were still well over 0.05 in both cases.

The previous-model comparison of the FO Constrained model at a large sample size was the most discriminant, based on mode p values alone. Both data types featuring populations with non-invariant second-order structures had mode p values of 0, while both data types featuring populations with invariant second-order structure had mode p values of over 0.4. The corresponding null-model comparisons followed similarly, with p values of less than 0.05 for the non-invariant second-order populations and p values of greater than 0.36 for the invariant second-order populations.

Consistent with the results of the mode p values, the percent of p values greater than 0.05 indicate that the FO Constrained previous-model comparison is a more discriminant test and is best at large sample sizes. In the large data size, the FO Constrained previous-model comparison resulted in correctly failing to reject 79% and 92% of data sets with invariant second-order population structure and either metrically or configurally non-invariant first-order factor measurement in the population, respectively. For large sized data sets with non-invariant second-order population structure and metrically or configurally non-invariant first-order measurement, the FO Constrained previous-model comparison resulted in (incorrectly) failing to reject only 11% and 15% of the data sets, or correctly rejecting 89% and 85%, of the data sets respectively. In terms of both specificity and sensitivity, this surpasses the 75% benchmark. The FO Constrained null-model comparison was less discriminant, incorrectly failing to reject 41% and 46% of such models.

Previous-model comparisons for the SO Constrained models were neither sufficiently discriminant for the small nor medium sample sizes, for which the mode p values ranged from 0.307 to 0.383 and from 0.129 to 0.368, respectively. In the large sample size and contrary to original expectations, mode p values were well above α for the data with populations featuring non-invariant second-order structure while mode p values were well below α for data with populations featuring invariant second-order structure. The percent of p values greater than 0.05 reflected the same trend, with small to no difference between data types for the small and medium sample sizes, and only 33% and 21% of models meeting the $p > 0.05$ criterion for the samples with invariant second-order population structures with metrically and configurally non-invariant measurement structures, respectively, while 69% and 78% of models met the $p > 0.05$ criterion in samples with non-invariant second-order population structure and metrically and configurally non-invariant measurement models, respectively. However, re-examination of my procedure revealed that I was inadvertently imposing non-invariance in the model; once corrected, it appears that the models very much fit as expected (see tables 3.17 and 3.18 and the cross-sectional section of chapter 3–the Discussion of Simulation Models sub-section–for details).

Null-model comparisons for the SO Constrained models likewise were not discriminant at a small sample size, for which mode p values ranged from 0.167 to 0.225. However, at the medium sample size, null-model comparisons of the SO Constrained models resulted in mode p values of 0.054 and 0.036, near or below the α significance value, for data with populations featuring non-invariant second-order structure and either metrically or configurally non-invariant first-order factor measurement, respectively. On the other hand, for the corresponding data with populations featuring invariant second-order structure, mode p values were 0.142 and 0.108 respectively. At the large sample size for the SO Constrained null-model comparisons, mode p values for all data types were well below the α of 0.05, ranging from 0 to 0.003.

The percent of SO Constrained models with p values greater than 0.05 reflected the trends seen in the corresponding mode p values. Similar to the mode p values, at the small sample size, the percent of SO Constrained models with p values greater than 0.05 was not discriminant with a range of 77% to 82%. At the medium sample size, the SO Constrained null-model comparisons did result in slightly lower percentages for data with non-invariant second-order population structure than those with invariant second-order population structure. However, unlike the mode p values, at the medium sample size the SO Constrained null-model comparison was not sufficiently discriminant based on the percentage of p values greater than 0.05. Percents of p values greater than 0.05 ranged from 44% to 71% for the medium sample size. At the large sample size, SO Constrained null-model comparisons resulted in similar percentages across data types, ranging from 0% to 13%.

The SO Constrained model had 54 DF, compared to the 0 DF of the null model. The values presented are mean values drawn from the middle 95% of the distributions, in order to control for outliers. Additional fit statistics for the null-model and previous-model comparisons for only the SO Constrained model (see table B.1 of appendix B), match the results of the mode p values and p values greater than 0.05 in that they indicate no discrimination between second-order population structure in the null-model comparisons and the unexpected result of more frequent rejection of the models with invariant second-order population structure in the previous-model comparisons. The SO Constrained RMSEA values were all low enough to indicate a good fit, despite populations that should not fit the model well for the data with non-invariant second-order population structures, indicating that the RMSEA value is not sufficiently discriminant.

3.3.3 Longitudinal Model Parameter Estimation

First-order parameters

Mean factor loading estimates, EASEs, and EASE 95% confidence intervals for the longitudinal data are presented by group in appendix A Tables A.8, A.9, A.10, A.11, A.12, and A.13 for the first-order model with no constraints across the times (FO Unconstrained), and Tables A.14, A.15, A.16, A.17, A.18, and A.19 for the final "second-order" or autoregressive model with constrained factor autoregressions, variances, and covariances (AR Constrained).

For all data types and in both the FO Unconstrained as well as the AR Constrained models, error in estimated loading values decreased as sample size increased, as evidenced by the decrease in the EASE from small to large sample sizes. The EASE was 0.011 or lower for all metrically non-invariant models, even at the small sample size, while it was unacceptably large–as high as 0.124–in the small sample size for the configurally non-invariant models.

The data types featuring metric non-invariance in the population first-order measurement had small EASEs of 0.002 or less at the medium and large sample sizes in the FO Unconstrained models, while those with configural non-invariance had EASEs of as large as 0.101 in the FO Unconstrained. In the AR Constrained models, although both the metrically non-invariant as well as the configurally non-invariant data sets had larger EASEs, the EASEs for the configurally non-invariant data types were unacceptably large even in the medium and large sample sizes, as much as 0.092, while the EASEs for the metrically non-invariant data types were as large as 0.028 for the medium and large sample sizes.

Overall, although the errors are somewhat larger (the estimates are less precise) for the AR Constrained models than for the FO Unconstrained model, the mean estimates are more accurate for the AR Constrained model (see Tables 3.9, 3.10, 3.11, 3.12, and 3.13 for the population first-order factor loading values). The FO Unconstrained estimates are more precise, but are biased.

Table 3.9. Longitudinal Simulations: Population factor loading values for for time 1, in data with metrically invariant first-order measurement models

	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}
1	0.41		
2	0.58		
3	0.70		
4		0.47	
5		0.37	
6		0.69	
7			0.52
8			0.79
9			0.53

Table 3.10. Longitudinal Simulations: Population factor loading values for for time 2, in data with metrically invariant first-order measurement models

	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}
1	0.59		
2	0.53		
3	0.33		
4		0.66	
5		0.84	
6		0.77	
7			0.39
8			0.60
9			0.64

Table 3.11. Longitudinal Simulations: Population factor loading values for for time 3, in data with metrically invariant first-order measurement models

	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}
1	0.81		
2	0.30		
3	0.31		
4		0.40	
5		0.80	
6		0.71	
7			0.32
8			0.82
9			0.77

Table 3.12. Longitudinal Simulations: Population factor loading values for for time 2, in data with configurally invariant first-order measurement models

	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}
1	0.59		
2		0.53	
3			0.33
4	0.66		
5		0.84	
6			0.77
7	0.39		
8		0.60	
9			0.64

Table 3.13. Longitudinal Simulations: Population factor loading values for for time 3, in data with configurally invariant first-order measurement models

	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}
1	0.81		
2		0.30	
3			0.31
4			0.40
5	0.80		
6		0.71	
7		0.32	
8			0.82
9	0.77		

“Second-order” autoregression parameters

For the longitudinal model, error in measuring the “second-order” parameters also decreased as sample size increased (see appendix A Tables A.20 and A.21). At the small sample size, EASEs ranged from 0.016 to 0.021, while at the large sample size they ranged from 0.001 to 0.008. There was no appreciable difference in estimates and EASEs based on the measurement model in the population. Estimation of the autoregression was not affected by the type of structure and invariance in the first-order factor loadings. Again, average estimates for the data with non-invariant autoregression (in which the population autoregression values were 0.3 for time 1 to time 2 and 0.15 for time 2 to time 3) were biased towards the mean of the two values, while the average estimates for the data with invariant autoregression (in which the population autoregression values were 0.3) were unbiased.

3.3.4 Longitudinal Model Fit Information

Table 3.14 presents mode (50th percentile) p values and skew of the p values for three model types–the first-order model with no constraints (FO Unconstrained), the first-order model with factor variance/covariance matrix constrained across groups (FO Constrained), and the second-order model with constraints on the "second-order" factor autoregressions and first-order factor variances and covariances (AR Constrained). The p values result from χ^2 likelihood ratio tests comparing the model of interest to either a) a null model of variances, covariances, and means for the data set (referred to as the "null" model comparison), or to b) the previous model, which has fewer constraints/more degrees of freedom (referred to as the "previous" or "prev" model comparison). Thus the comparisons are as follows:

- The first-order factor model with no constraints (FO Unconstrained) against the null variance/covariance model
- The first-order factor model with constraints on the factor variances and covariances (FO Constrained) against the previous F-O Unconstrained model in a nested model comparison
- The FO Constrained model against the null variance/covariance model
- The final second-order factor model with constraints on the factor variances, covariances, and autoregression (AR Constrained) against the previous second order model with constraints on only the factor variances and covariances in a nested model comparison
- The AR Constrained model against the null variance/covariance model

In the case of the FO Constrained model (a first-order model with the factor variance/covariance matrix constrained across groups), the associated "prev" model is the FO Unconstrained model which has no across-group constraints. In the case of the AR Constrained model (a second-order model with constrained autoregressions and constrained first-order factor variances and covariances), the associated "prev" model is the second-order model with constraints on only the first-order factor variances and covariances.

Table 3.14. Longitudinal Simulations, model fit information: 50th percentiles (i.e. modes) and associated distributional skews for p values of likelihood ratio tests. Rows are the target data types and columns are the relevant model comparisons. Square brackets "[ ]" contain expected mode given data type.

Size	1. FO Unconstr$_{null}$	2. FO Constr$_{prev}$	3. FO Constr$_{null}$	4. AR Constr$_{prev}$	5. AR Constr$_{null}$
Measurement Model: metrically non-invariant, S-O Model: non-invariant [$< .05$]					
100	0.025 (1.274)	0.177 (1.094)	0.019 (1.26)	0.257 (0.981)	0.047 (1.168)
500	0.005 (1.135)	0.007 (1.172)	0.001 (0.992)	0.046 (1.338)	0.152 (1.031)
1500	0 (0.338)	0 (0.327)	0 (0.131)	0 (0.696)	0.011 (1.242)
Measurement Model: metrically non-invariant, S-O Model: invariant [$> .05$]					
100	0.015 (1.185)	0.386 (0.371)	0.014 (1.223)	0.425 (0.072)	0.059 (1.221)
500	0 (0.738)	0.483 (0.208)	0 (0.831)	0.48 (-0.034)	0.404 (0.148)
1500	0 (0.13)	0.507 (-0.035)	0 (0.182)	0.493 (-0.024)	0.481 (0.031)
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg [$< .05$]					
100	0.012 (1.297)	0.18 (1.11)	0.009 (1.298)	0.417 (0.244)	0.047 (1.366)
500	0 (0.637)	0.009 (1.258)	0 (0.525)	0.484 (0.166)	0.22 (0.91)
1500	0 (0.19)	0 (0.253)	0 (0.174)	0.488 (0.072)	0.044 (1.21)
Measurement Model: configurally non-invariant, S-O Model: non-invariant [$< .05$]					
100	0.021 (1.296)	0.208 (0.953)	0.017 (1.252)	0.282 (0.709)	0.039 (1.312)
500	0.005 (1.244)	0.004 (1.069)	0.001 (1.085)	0.047 (1.376)	0.135 (1.119)
1500	0 (0.319)	0 (0.424)	0 (0.262)	0 (0.671)	0.012 (1.308)
Measurement Model: configurally non-invariant, S-O Model: invariant [$> .05$]					
100	0.012 (1.219)	0.39 (0.44)	0.011 (1.222)	0.405 (0.208)	0.055 (1.304)
500	0 (0.581)	0.492 (-0.146)	0 (0.555)	0.481 (0.054)	0.351 (0.549)
1500	0 (0.205)	0.511 (-0.065)	0 (0.192)	0.493 (0.045)	0.466 (0.125)
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg [$< .05$]					
100	0.01 (1.314)	0.19 (0.993)	0.007 (1.307)	0.427 (0.21)	0.041 (1.418)
500	0 (0.785)	0.008 (1.208)	0 (0.565)	0.474 (0.235)	0.208 (0.736)
1500	0 (0.237)	0 (0.274)	0 (0.163)	0.48 (0.038)	0.043 (1.125)

For the same models and model comparisons, Table 3.15 presents the percent of analyses in which the likelihood ratio p value was non-significant (where $p > 0.05$), indicating that the model of interest did not fit significantly worse than the comparison model. For these calculations, the distribution of p values was reduced to the middle 95% in order to control for spurious extremes.

Mode p values for the FO Unconstrained null-model comparisons ranged from 0 to 0.025 across all sample sizes, and the distribution skew decreased with increased sample size. Consistent with this result, only 35% or fewer of the FO Unconstrained models were not rejected, with 0% of models failing to be rejected in the large sample size regardless of underlying population structures.

In the FO Constrained previous-model comparison, mode p values decreased as sample size increased in data with non-invariance in their "second-order" population models, while mode p values increased as sample size increased in data with invariant "second-order" population structures, regardless of their first-order measurement. Skews were relatively larger for the data with non-invariance than those with invariance as well. In the small sample size, all mode p values were above 0.05, ranging from 0.177 to 0.390. However, in the medium and large sample sizes, mode p values ranged from 0 to 0.009 for the second-order non-invariant data, and from 0.483 to 0.511 for the second-order invariant data. The percentages of models not rejected at α = 0.05 were consistent with these results for the mode p values. While from 67% to 92% of FO Constrained models were not rejected in the small sample size across all data types, in data with non-invariant second-order structures from 15% to 24% and 0% of models were not rejected in the medium and large sample sizes, respectively. On the other hand, 88% to 98% and 97% to 98% of models were not rejected in the medium and large sample sizes, respectively, for data with invariant second-order population structure. Despite the FO Unconstrained model comparison being indiscriminating in terms of the underlying population structures, the FO Constrained model comparison test is discriminant between the fully invariant second-order populations

Table 3.15. Longitudinal Simulations, model fit information: Percent of models with non-significant likelihood ratio test ($p > .05$). Rows are the target data types and columns are the relevant model comparisons. Square brackets "[ ]" contain expected percent given data type.

Size	1. FO Unconstr$_{null}$	2. FO Constr$_{prev}$	3. FO Constr$_{null}$	4. AR Constr$_{prev}$	5. AR Constr$_{null}$
Measurement Model: metrically non-invariant, S-O Model: non-invariant [$\% \to 0$]					
100	0.35	0.76	0.30	0.81	0.49
500	0.19	0.15	0.07	0.49	0.72
1500	0.00	0.00	0.00	0.02	0.27
Measurement Model: metrically non-invariant, S-O Model: invariant [$\% \to 1$]					
100	0.26	0.83	0.25	0.82	0.48
500	0.00	0.88	0.00	0.88	0.85
1500	0.00	0.97	0.00	0.97	0.97
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg [$\% \to 0$]					
100	0.23	0.67	0.20	0.84	0.44
500	0.00	0.22	0.00	0.87	0.73
1500	0.00	0.00	0.00	0.97	0.47
Measurement Model: configurally non-invariant, S-O Model: non-invariant [$\% \to 0$]					
100	0.33	0.76	0.28	0.84	0.45
500	0.20	0.16	0.10	0.49	0.69
1500	0.00	0.00	0.00	0.02	0.26
Measurement Model: configurally non-invariant, S-O Model: invariant [$\% \to 1$]					
100	0.26	0.92	0.25	0.92	0.52
500	0.00	0.98	0.01	0.96	0.92
1500	0.00	0.98	0.00	0.97	0.97
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg [$\% \to 0$]					
100	0.24	0.76	0.22	0.94	0.46
500	0.00	0.24	0.00	0.98	0.81
1500	0.00	0.00	0.00	0.97	0.46

and the second-order populations with non-invariance in their first-order factor covariances.

The FO Constrained null-model comparison, in contrast to the previous-model comparison, was not discriminant. Mode p values were well below α, ranging from 0 to 0.019 across all of the data types and sample sizes. Likewise, only from 20% to 30%, 0 to 10%, and 0% of models were not rejected in the small, medium, and large sample sizes, respectively.

The AR Constrained previous-model comparison test discriminated between the data with invariant autoregression and those without. While the mode p values for all of the data types were well above α in the small sample size (ranging from 0.257 to 0.427), they increased as sample size increased in the data with invariant population autoregression (ranging from 0.474 to 0.493 across the medium and large sizes) but decreased as sample size increased in the data with non-invariant population autoregression (ranging from 0 to .047 across the medium and large sizes). Likewise, the percent of models not rejected went from over 80% to 2% as the sample size increased for data with non-invariant autoregressions, but increased from over 80% to 97% in data with invariant autoregressions.

Finally, at the large sample size, the AR Constrained null-model comparison was discriminant in terms of different results for data with invariant factor covariances as well as autoregressions and data with non-invariant factor covariances regardless of the autoregression. At the small sample size, mode p values were small, ranging from 0.039 to 0.059 and between 44% and 52% of models failing to be rejected, across all data types. At the medium sample size, the mode p values were slightly higher for the data with an invariant second-order population (0.404 and 0.351 for the first-order metric and configural invariance conditions, respectively) than those with non-invariant factor covariances (ranging from 0.135 to 0.22). In terms of the percentages of models not rejected, rates for the models with an invariant second-order population were again slightly higher relative to those with non-invariance in the factor covariances (85% compared to 72% and 73% for the second-order invariant compared to the second-order non-invariant and the second-order mixed when first-order is metrically non-invariant, 92% compared to 69% and 81% for the same second-order types when first-order is configurally non-invariant), but the differences are small. However, in the large sample size, mode p values were 0.011 and 0.012 for the data featuring a second-order non-invariant population structure with first-order metric and configural non-invariance, respectively, 0.044 and 0.043 for data featuring non-invariant factor covariances but invariant autoregression in the population with first-order metric and configural non-invariance, respectively, and 0.481 and 0.466 for data featuring an invariant second-order population structure, again respectively depending on whether they featured metric or configural non-invariance. Similarly, the percent of models not rejected was less than 30% for the data with fully non-invariant second order population structure, and 97% for data with fully invariant second-order population structure.

The AR Constrained model had 345 DF, compared to the 0 DF of the null model. Note that the values presented are drawn from the middle 95% of the distributions, in order to control for outliers. As Table 3.16 presents, AIC for the previous-model comparisons for only the AR Constrained model are consistent with the results of the mode p values and p values greater than 0.05 in that the AICs for the data with invariant population autoregression are lower than the comparison AICs and would not be rejected, while the AICs for the data with second-order population non-invariance in the autoregressions are higher than the AICs for the comparison model and would be rejected. The difference in log likelihoods reflected the AIC differences, while the RMSEAs for the models were all small, approximately 0 at the medium and large sample sizes, and did not vary between the data types.

Likewise, the differences in log likelihood and AIC from the null-model comparisons are large and positive, consistent with the results of the null-model χ^2 comparisons, which suggest rejecting the models regardless of data type and sample size. Again, all RMSEA values were low enough to indicate a good fit, regardless of whether or not the data had populations that should not fit the model well, such as the data with non-invariant second-order population structures, indicating that the RMSEA value is not sufficiently discriminant.

3.4 Discussion of Simulation Models

3.4.1 Cross-Sectional

Initially, the cross-sectional model results appeared to be contrary to expectations; ideally the HOIIF would reject models featuring populations with non-invariant second-order structure at the step of constraining the first-order factor covariances and in constraining the second-order model. However, because the HOIIF steps involve nested comparisons against a previous model, if the previous model did not fit the data well because the model constrains parameters to invariance that are not invariant in the population, then additional constraints might not make the model fit significantly worse than a model without those constraints. For example, if a second-order model with across-group constraints on first-order factor variances and second-order factor loadings does not fit the data well, then the fit of the next model, in which I add a constraint on the second-order factor variance in addition to the previous model's constraints on the first-order factor variances and the second-order factor loadings, might not substantially reduce the already poor fit. Reviewing the practical steps that one would take in order to implement this method does reveal that there are points at which the majority of the non-invariant models would be rejected– namely 1) at the step of constraining the first-order factor variance/covariance matrix and comparing this model to the the previous, unconstrained first-order model, and 2) in a comparison of the final second-order, constrained model to the null model of variances and covariances.

On the other hand, the high rejection rate of the models of data with invariant second-order structure was, at first, perplexing. These data had invariant second- order "measurement" models, which was a match to the structure created in the analysis and

Table 3.16. Longitudinal Simulations, model fit information: Difference in DF and mean likelihood difference between the null variance/covariance model with constrained "second-order" model, mean likelihood difference of previous model (constrained loadings and errors) with constrained "second-order" model, AIC difference, mean RMSEA (SD), RMSEA skew

Sample Size	Diff df	Diff_{null} -2LL	Diff_{null} AIC	Diff_{prev} -2LL	Diff_{prev} AIC	RMSEA	RMSEA skew
Measurement Model: metrically non-invariant, S-O Model: non-invariant							
100	345	391.97	76112.26	4.77	-1.04	0.001 (0.001)	0.23
500	345	373.46	54493.66	8.90	3.15	0 (0)	0.18
1500	345	407.98	528.24	20.42	14.68	0 (0)	0.04
Measurement Model: metrically non-invariant, S-O Model: invariant							
100	345	387.35	76107.52	3.40	-2.42	0.001 (0.001)	0.06
500	345	351.56	54471.73	2.94	-2.92	0 (0)	0.10
1500	345	346.28	466.61	2.88	-2.96	0 (0)	0.10
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg							
100	345	391.15	76111.35	3.37	-2.48	0.001 (0.001)	0.11
500	345	365.97	54486.25	2.94	-2.91	0 (0)	0.12
1500	345	391.23	511.42	2.90	-2.96	0 (0)	0.04
Measurement Model: configurally non-invariant, S-O Model: non-invariant							
100	345	393.07	76113.40	4.50	-1.34	0.001 (0.001)	0.05
500	345	374.93	54495.08	8.62	2.92	0 (0)	0.09
1500	345	407.63	527.94	19.50	13.63	0 (0)	0.05
Measurement Model: configurally non-invariant, S-O Model: invariant							
100	345	388.47	76108.62	3.52	-2.34	0.001 (0.001)	0.05
500	345	354.40	54474.38	2.97	-2.90	0 (0)	-0.01
1500	345	346.49	466.66	2.88	-2.94	0 (0)	-0.01
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg							
100	345	392.25	76112.44	3.38	-2.41	0.001 (0.001)	0.05
500	345	367.01	54487.25	2.88	-2.99	0 (0)	0.12
1500	345	392.89	513.10	2.93	-2.93	0 (0)	0.16

therefore should not have been rejected. Yet they were fitting the data at the unconstrained stage and being rejected as they were constrained. Upon closer consideration of the model constraints, I determined that I was forcing the second-order factor model into different measurement scales between the groups by scaling the first-order factors by the manifest variable loadings. As a brief confirmation of this, I ran a small number of simulations of the four medium sample size data sets (data types 2, 5, 8, and 11), and scaled the first-order factor model by the first-order factor variances. Although this creates a built-in constraint on the second-order "measurement" model, it is necessary in order for the HOIIF method to work. Indeed, a post-hoc small scale simulation study revealed that when the models were appropriately scaled, the models featuring invariance of the second-order factor structure failed to be rejected while the models featuring non-invariant second-order factor structure were rejected (see tables 3.17, 3.18). In this small simulation, the HOIIF appears to meet the criterion for sensitivity (failing to reject for data with invariant SO structures); however, it does not appear to meet the criterion for specificity (rejecting models with data featuring non-invariant SO structures).

3.4.2 Longitudinal

Unsurprisingly, the small sample size was not sufficient for this type of modeling. In the small sample size I found substantially more spurious results. There were more unrealistically extreme parameter estimates and more instances of type I error. Implementing a HOIIF requires a sufficient number of parameters to create a structure with at least two-orders; the method requires many free parameters and so the sample size must be sufficient to estimate.

The model fits suggest that the HOIIF first-order, unconstrained model is sensitive to the effect of the autoregression on the parameters. Without modeling the autoregression, between 80% and 100% of models are rejected in the medium and large sample sizes, and 65% - 77% are rejected in the small sample size. Even the effect of constraining the model

Table 3.17. Cross-Sectional Simulations (RE-SCALED MODELS), model fit information: Percent of models with non-significant likelihood ratio test ($p > .05$). Rows are the target data types and columns are the relevant model comparisons. Square brackets "[ ]" contain expected percent given data type.

Size	1. FO Unconstr$_{null}$	2. FO Constr$_{prev}$	3. FO Constr$_{null}$	4. SO Constr$_{prev}$	5. SO Constr$_{null}$
Measurement Model: metrically non-invariant, S-O Model: non-invariant [$\% \to 0$]					
500	0.87	0.44	0.76	0.35	0.65
Measurement Model: metrically non-invariant, S-O Model: invariant [$\% \to 1$]					
500	1.00	0.91	0.89	1.00	1.00
Measurement Model: configurally non-invariant, S-O Model: non-invariant [$\% \to 0$]					
500	0.93	0.35	0.74	0.42	0.78
Measurement Model: configurally non-invariant, S-O Model: invariant [$\% \to 1$]					
500	0.96	0.96	1.00	0.94	1.00

Table 3.18. Cross-sectional Simulations, model fit information (RE-SCALED MODELS): Df difference, mean likelihood difference of the null variance/covariance model with constrained second-order model, mean likelihood difference of previous model (constrained loadings and errors) with constrained second-order model, AIC difference, mean RMSEA (SD), RMSEA skew

Sample Size	Diff df	Diff$_{null}$ -2LL	Diff$_{null}$ AIC	Diff$_{prev}$ -2LL	Diff$_{prev}$ AIC	RMSEA	RMSEA skew
Measurement Model: metrically non-invariant, S-O Model: non-invariant							
500	51	71.22	72.23	17.31	14.55	0.001 (0.001)	1.01
Measurement Model: metrically non-invariant, S-O Model: invariant							
500	51	51.86	51.91	0.77	-1.10	0 (0)	0.03
Measurement Model: configurally non-invariant, S-O Model: non-invariant							
500	51	58.45	59.61	6.93	6.03	0 (0)	0.07
Measurement Model: configurally non-invariant, S-O Model: invariant							
500	51	50.05	51.82	0.51	-1.23	0 (0)	0.58

does not fully recover the effect of failing to model the autoregression. If there is invariance in the second-order model, ideally comparing the constrained model to the unconstrained model would allow us to identify whether or not it is present; however, these results suggest that if autoregression exists at the second-order, the model might be rejected should we fail to model the autoregression directly.

The results indicate that the previous-model comparison test for the constrained second-order factor model discriminates between the models with invariant autoregression and those without. This is intuitive, as the difference between this model and the previous model is the constraint of the autoregressions. These models would have been rejected at the FO Constrained previous-model step; thus the combination of those two steps in the HOIIF would provide strong evidence of this underlying structure.

With a large sample size, the null-model comparison for the model with invariant autoregression but non-invariant factor covariance in the population only fits slightly less than half of the time, while in the medium sample, it fits between 73% and 81% of the time. Perhaps this indicates that, while it is critical to have sufficient data to discriminate signal from noise, when there are still true discrepencies in the signal, having more data will result in those differences becoming more salient than the original signal.

3.4.3 Summary and Limitations

Based on the mode p values in the original simulations as well as the small follow-up, we see that the HOIIF functions well in both cross-sectional and longitudinal models. It can discriminate between invariant and non-invariant second-order structure, even when first-order measurement is non-invariant. Likewise, mode p values and percentages of models not rejected are higher as sample size increases. For models with metrically non-invariant measurement models, the HOIIF appears to be powered to detect differences starting at a medium sample size.

The step of constraining the first-order factor covariances appears to be a crucial one in the HOIIF procedure. The previous model comparison at this step is key to filtering out models with second-order non-invariant structures. The method can discriminate between a model with invariant higher-order structure and one without. However, it will be important to determine whether this method can discriminate between different higher-order structures. Constraining the second-order loadings is one test of how this method functions with incorrectly specified models; the current evidence suggests that there is potential for over fitting at the second-order, especially when the first-order measurement is configurally non-invariant. This is why the first-order model is an important cut point–should the model fail at the first constraint of the higher-order solution, then a higher-order solution can be rejected.

Of note, the RMSEAs for all of the models were very small. For all models, the RMSEA was well less than the conventional cutoff value of .05, suggesting that this fit statistic is not sufficient for evaluating these types of models. On the other hand, despite the natural order of and language around the HOIIF steps, there is not true nesting between the second-order models and the first-order models. Rather than simply constraints, the second-order model is a re-parameterization of the first-order model and therefore is not nested within the first-order model. Therefore, in implementing the HOIIF it would not be proper to perform a likelihood ratio test of the second-order and first-order models. However, they are similar models and because they are modeled on the same data sets, they can be compared on AIC values in a model selection framework.

When the second-order model was non-invariant between groups, the HOIIF showed some issues with estimation–particularly in the small sample size. Somewhat surprisingly, this effect was most pronounced in fitting the first-order, unconstrained model (After varying starting values, 57% and 53% of models were still identified as having convergence issues for the metrically and configurally non-invariant models, respectively). Based on this outcome, while it could be considered somewhat informative to find convergence issues in practice as these results would suggest it could be an indication of a second order structure that is non-invariant between groups, this would be speculative and such convergence issues are a hindrance to identifying structure in the data.

4. Observed Data

Following the Monte Carlo analyses, I applied a HOIIF factor analysis in two empirical data sets. The first data set contained cross-sectional cognitive assessment data to which I applied the proposed HOIIF factor analysis procedure that allows for between-group measurement non-invariance while still establishing between-group invariance. The second collected data set contained longitudinal child outcome data, to which I applied a HOIIF dynamic factor analysis to allow for between-occasion measurement non-invariance. I applied the HOIIF following the stepwise procedure identified in Zhang et al., 2011, utilized in Dodson et al., 2014, and detailed above. Model fit was assessed via a likelihood ratio test comparison with the null variance/covariance model or a nested model comparison, as applicable, at each step of the procedure.

4.1 Cross-Sectional Observed Data

4.1.1 Methodology

Participants

The data in this study come from the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) research program at the National Institute on Aging (Evans et al., 2010). HANDLS is a multidisciplinary epidemiological study of health risks and disparities in aging. Participants were randomly sampled from twelve census tracts across Baltimore; the neighborhoods were identified based on their likelihood of providing a representative sample of Baltimore residents. Participants for the present study were black and white adult residents of the city of Baltimore, and we used data from self-reported demographic information and performance on the HANDLS cognitive battery.

I provide a summary of participant demographic information in Table 4.1. The 2030 (892 female, 1138 male) participants ranged in age from 30-64 years (*mean* = 48.03, σ = 9.29), 59.75% reported income above 125% of the poverty level established by the Department of Health and Human Services at the time of testing (Fed, 2006), and scores on the WRAT-3 measure of literacy ranged from 7 to 57 (μ = 42.15, σ = 8.30). The participants had from 1 to 21 years of education (μ = 12.58, σ = 3.66) and were black (*N* = 1175) and white (*N* = 855). In order to investigate cognitive performance while allowing for differences in literacy and socioeconomic status, participants were divided into six groups based on WRAT-3 scores and income:

1) BLL: below–poverty criterion, low–literacy (reported income $< 125\%$ poverty level, WRAT-3 score < 39, *N* = 330)

2) BML: below–poverty criterion, mid–literacy (reported income $< 125\%$ poverty level, 39$\leq$ WRAT-3 score ≤ 47, *N* = 318)

3) BHL: below–poverty criterion, high–literacy (reported income $< 125\%$ poverty level, WRAT-3 score > 47, *N* = 169)

4) ALL: above–poverty criterion, low–literacy (reported income $> 125\%$ poverty level, WRAT-3 score < 39, *N* = 285)

5) AML: above–poverty criterion, mid–literacy (reported income $> 125\%$ poverty level, 39 $\leq$ WRAT-3 score $\leq$ 47 , *N* = 489)

6) AHL: above–poverty criterion high–literacy (reported income $> 125\%$ poverty level, WRAT-3 score 47 , *N* = 439)

Measures

As part of the HANDLS protocol, participants were administered a cognitive battery including the following nine neuropsychological tests: Benton Visual Retention Test, 5th edition (BVR); the Card Rotation Test; the Identical Pictures Test; the California Verbal

Table 4.1. HANDLS demographic means (standard deviation), and percentages, by group

	BLL	BML	BHL	ALL	AML	AHL
Age	47.23	46.92	47.56	49.84	49.18	47.14
	(8.84)	(9.13)	(9.76)	(9.16)	(9.40)	(9.24)
Gender (% F)	0.43	0.39	0.40	0.46	0.43	0.49
Race (%White)	0.24	0.31	0.49	0.35	0.38	0.70
Household Ed	10.68	11.72	13.34	11.32	12.93	14.86
	(2.25)	(2.44)	(2.80)	(5.80)	(2.57)	(3.46)
Literacy	31.47	43.07	50.28	32.47	43.61	51.04
	(5.98)	(2.61)	(1.99)	(5.45)	(2.58)	(2.20)

Learning Test (CVL, current data includes data from two immediate recall trials); forward and backward digit span tests; the animal fluency test; parts A and B of the Trail-making Test (Trails A and B); the Brief Test of Attention (Attn); and a clock drawing test. Additionally, the reading subtest of the Wide Range Achievement Test, 3rd edition (WRAT-3), is used in the current study to assess educational achievement. Trained psychometrists administered the neuropsychological tests; the psychometrists were trained and supervised by a research psychologist throughout the course of the study. A more detailed overview of the HANDLS study, including participants and measures, can be found in *Healthy Aging in Neighborhoods of Diversity Across the Life Span (HANDLS): Overcoming Barriers to Implementing a Longitudinal, Epidemiologic, Urban Study of Health, Race, and Socioeconomic Status* (Evans et al., 2010).

Analyses

The HOIIF method has been proposed to allow estimation of invariant latent factor models across groups, while filtering out configural or metric non-invariance. In this case, the observed variables are identical between groups, therefore configural non-invariance is not necessarily an *a priori* concern (as it is in the case of different observed variables between groups). In prior work with these data (Dodson et al., 2014), I attempted to fit a HOIIF by establishing independent group factor models via exploratory analysis, and then rotating those models towards a theoretical factor structure. In this previous work, I found that a first-order factor model with invariant factor measurement structure was not supported; therefore, I proceeded directly to allowing first-order measurement non-invariance in the current study. For the current analysis, I used exploratory factor analysis independently in each group to determine anchor items–items with a high loading on one factor and negligible loadings on all other factors–that were consistent across groups. After identifying anchor items, I estimated a first-order factor model in a confirmatory framework with the anchor items constrained to load on only their relevant factors, and all other items freely estimated with two exceptions–because the digit span and CVL items loaded uniquely on their factors in the EFA models, they were both restricted to load only on their respective factors in the CFA. In the event that I was unable to identify anchor items satisfying the condition of loading uniquely on one factor consistently across all groups, I selected items that satisfied that condition as closely as possible. In the case that a model with 0-value loadings for anchor items on the non-anchored factors did not hold, I relaxed the restrictions while fixing very small loadings (0.20 or below) to 0 in order to preserve identifiability. I fixed the loadings to 0 from the anchor item to all factors except its relevant factor in the confirmatory analysis. Model fit was based on the $\chi^2 - 2$log likelihood ($-2LL$) and AIC values from a likelihood ratio test against a null model of item variances and covariances or against the previous model, and model RMSEA. Note that different loading values, or even different simple structure, is acceptable in these analyses because the HOIIF can allow us to filter out such first-order non-invariance. Once the final first-order factor model structures were determined, I followed the HOIIF procedure as previously described.

4.1.2 Results

All analyses, including both the traditional confirmation of factorial invariance and the second-order idiographic filter method, were performed using R and the OpenMx program in R (R Core Team, 2012; Boker et al., 2012).

The data were standardized prior to the analyses. I used the scree plot method as well as Horn's Parallel Analysis to identify the number of factors to use for these data

(Cattell, 1966; Horn, 1965). Taking into account the results of the scree test and Horn's Parallel Analysis, I moved forward with a four-factor model with anchor items; this was also based in large part on previous work using the HANDLS population (Beydoun, 2010) and the Cattell-Horn-Carroll Theory of Cognitive Abilities (CHC, Horn and Blankson, 2005). The first latent variable was anchored by the Trail-making A tests and corresponds with a *Processing Speed (Gs)* factor. The second latent variable was anchored by the first CVL test and generally corresponded with a *Long-Term Storage and Retrieval (Glm)* factor. The third latent variable was anchored by the digit span forward test and corresponded with a *Short-Term Memory (Gsm)* factor. The fourth latent variable was identified by the Card Rotation task and corresponded with a *Visual Processing (Gv)* factor. In all models, the high end of the 95% CI for RMSEA was $< .04$. This was in accordance with finding low RMSEAs regardless of underlying data structure in the simulation study, and therefore model RMSEA will not be reported further.

In specifying the anchor model, anchor items were fixed to 1 in their relevant factors and 0 in all other factors, while the remaining eight items were freely estimated. In comparison to a null model of the data variances and covariances ($-2LL = 58088.47, AIC = 13412.47, df = 22338$), the anchor model -2 log likelihood was higher ($-2LL = 58462.66, df = 22523$) and correspondingly had a significant likelihood ratio test ($p < .001, df_{difference} = 185$) and slightly higher AIC ($AIC = 13416.66$; see appendix B table B.4 for model comparisons). However, given that the anchor model had many more constraints and the AIC was close to that of the null model, I used the results of this model to guide the application of constraints in the subsequent model.

The first-order, unconstrained factor model featured additional 0-constraints on the factor loadings (see e.g. the first-order measurement of B.1). The factor models featured two well defined factors (Glm, Gsm) while the remaining two factors were somewhat less consistent between groups (see figure B.1 for final factor structure. Parameters in dashed boxes were constrained to invariance between groups, while underlined parameters were fixed to the values displayed). This first-order, unconstrained factor model showed a significant difference via likelihood ratio test against the null variance-covariance model ($-2LL = 58486.74, df = 22613, df_{difference} = 275, p < .001$) but much lower AIC ($AIC = 13260.74$). The subsequent first-order model with factor covariances constrained across groups likewise showed a significant difference via likelihood ratio test comparison to the previous, first-order unconstrained model ($-2LL = 58534.48, df = 22643, df_{difference} = 30, p = .02$) but lower AIC ($AIC = 13248.48$).

In the second-order, unconstrained factor model, the first-order measurement model is maintained while the first-order factor covariances are replaced by a second-order factor "measurement" model. For the current case, the first-order factors load onto a single second-order general ability factor "G," again in accord with the CHC (Horn and Blankson, 2005). In the second-order unconstrained factor model, the first-order measurement model was scaled by fixing the first-order factor covariances to 1, while the second-order measurement model was scaled by fixing one second-order factor loading to 1, and featured no constraints between groups. As with the previous models, in a likelihood ratio test comparison to the variance-covariance null model, the second-order unconstrained model had a higher -2LL and a p-value indicating statistical significance, but had a lower AIC ($-2LL = 58527.60, df = 22625, df_{difference} = 287, p =< .001, AIC = 13277.6$). The next step, constraining the second-order factor loadings, again resulted in a model with a higher -2LL and statistically significant likelihood ratio test value, but lower AIC in comparison to the previous, second-order unconstrained model ($-2LL = 58555.74, df = 22640, df_{difference} = 15, p = .02, AIC = 13275.74$). The final model, in which the second-order factor variance was constrained across groups (because the second-order "manifest errors"–i.e. the first-order factor variances–were fixed to 1, they were already constrained between groups and this step was skipped) had a slightly higher -2LL than the previous model, but a likelihood ratio test just above the significance cutoff and an AIC not practically different from the previous model, suggesting a reasonable fit given that the model features more constraints ($-2LL = 58566.50, df = 22645, df_{difference} = 5, p = .06, AIC = 13276.50$). This model again had a higher -2LL than the null variance/covariance model but a substantially lower AIC value ($df_{difference} = 307, p < .001$). See appendix B, figures B.1, B.2, B.3, B.4, B.5, B.6 for the final second order models for each group.

4.2 Longitudinal Observed Data

4.2.1 Methodology

Participants

The data for the longitudinal analysis come from the Early Steps Project (ES), a longitudinal research program (see Dishion et al., 2008, for details). ES is an intervention program for conduct problems in children in high risk families. At recruitment, participants were from low-income families with at least one 2 year-old child who was nonclinical but at risk for conduct problems. Participants were drawn from an urban (Pittsburgh, PA), suburban (Eugene, OR), or rural (Charlottesville, VA) location. Families were assessed at ages 2, 3, 4, 5, 7, 8, and 9 via parent, child, teacher, and observational report. 731 families participated in the first wave of the study, while 588 families participated in the 9th wave. I provide a summary of participant demographic information in Table 4.2. 66.5% of the primary caregiving parents in the 731 families (49.5% participating female toddlers at enrollment) reported income of less than $20,000 at enrollment. 65% of the primary caregivers had a high school/GED diploma or less; 35% had above highschool training or education. The primary caregiver identified as black or African American (28%), white or Caucasian American (50%), biracial (13%), or other group (9%); 13% identified as Hispanic or Latino/a.

Table 4.2. ES demographic means (standard deviation), and percentages, by group. 1. Gender of child (% female) 2. Race of child (% White) 3. Primary caregiver (% biological mother) 4. Household Education (mode response) 5. Household income (% with income less than $20,000 annually) 6. Household size (mode number of family members)

	Age 2	Age 3	Age 4	Age 5	Age 7	Age 8	Age 9
1. Gender	49.5						
2. Race	50.1						
3. PC	92.7	95.5	94.7	93.3	92.9	90.8	90.8
4. House Edu	HS/GED	HS/GED	HS/GED	HS/GED& College/ Training	HS/GED	HS/GED	HS/GED
5. House Income	66.5	60.0	54.3	44.6	39.9	43.7	38.6
6. House Size	4	4	4	4	4	4	4

Measures

As part of the ES protocol, families were administered a variety of measures including the Child Behavior Checklist (CBCL, Achenbach and Rescorla, 2000, 2001). The CBCL was, at enrollment, a 99 item parent-rated questionnaire. The items featured a statement and participants were asked to rate whether that statement described their child, from a scale of 0 (not true, as far as you know) to 2 (very true or very often).

Analyses

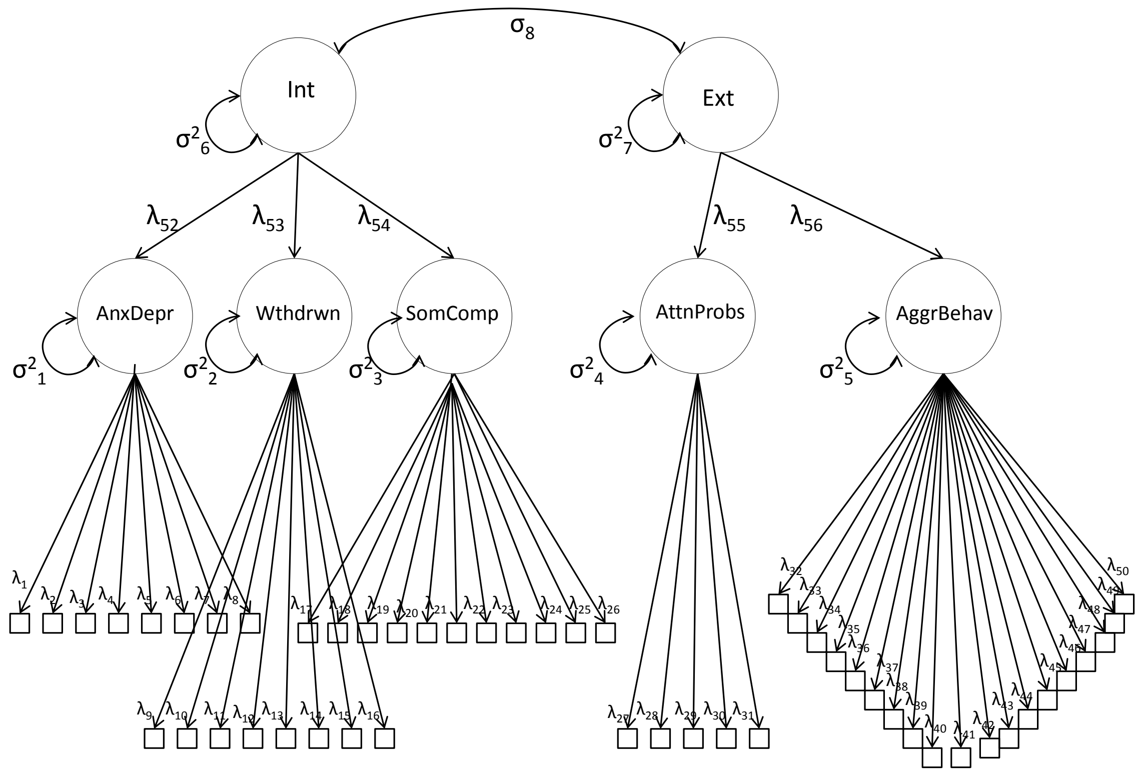
The HOIIF method has been proposed to allow estimation of invariant latent factor models across groups, while filtering out configural or metric non-invariance. In the case of this longitudinal data set, the observed variables change slightly across time, therefore configural non-invariance is necessarily an *a priori* concern. In each wave of measurement, a small number of new items are introduced, a small number of items are dropped, and the wording changes for some items. Because the observed variables change only to a small extent and the scale was designed in a theoretical factor structure that has been validated (Achenbach and Rescorla, 2000, 2001), I did not perform an exploratory analysis. I estimated a first-order factor model in a confirmatory framework based on the theoretical factor structure, and followed the HOIIF procedure as previously described.

4.2.2 Results

Again, because the observed variables change only to a small extent and the scale was designed in a theoretical factor structure that has been validated (Achenbach and Rescorla, 2000, 2001), no exploratory analysis was run. Because of the large number of participants and time points and complexity of the model, I ran the models on the data correlation rather than the raw data. The analysis was limited to the five factors ("Anxious/Depressed," "Withdrawn/Depressed," "Somatic Complaints," "Attention Problems," "Aggressive Behavior") that were identified across ages in the previous work with this scale (Achenbach and Rescorla, 2000, 2001) and therefore the items that loaded onto those five factors; all other factors and items were excluded from consideration (see figure 4.1 for the age 2 factor model). The first-order models were run as a parent model containing separate subgroup models, while the second-order models were run on a full polychoric correlation matrix of all observations across all time points. I estimated a first-order, multiple group factor model in a confirmatory framework based on the theoretical factor structure from the (polychoric) correlation matrices for each age group ($-2LL_{null} = -20049.12, -2LL_{model} = -132044.8, AIC = -125055.9, df = 10442, df_{diffnull} = 10442, p \approx 1$; see appendix B table B.5 for table of model comparisons). In the next model, I constrained the factor variance/covariance matrix of this first model, which did not result in a significant reduction in fit ($-2LL = -138796.7, AIC = -139673.1, df = 10502, df_{difference} = 60, p \approx 1$).

After establishing that constraining the first-order factor variances and covariances between ages did not result in a significant reduction in fit, I fit a second-order model in which the first-order factors loaded onto either the *internalizing* or *externalizing* second-order factors. This model had poor overall fit ($-2LL_{null} = -1133040, -2LL_{model} = 223874.8, AIC = 1203728.4, df_{difference} = 76593, p < .001$). Because 1) this test is increasingly sensitive to difference in degrees of freedom as sample size increases and the current data set was very large, 2) the model is based on a previously validated second-

Figure 4.1. The second-order model for the Early Steps age 2 data. This model features five first-order factors and two second-order factors; although the measurement model varies between ages the second-order "measurement" model is set to be consistent across ages.



order structure, and 3) this exercise is for illustrative purposes, I proceeded with the HOIIF modeling without further modification. As a test of whether or not allowing the dynamic factor structure across ages results in a poorer fitting model, I added parameters to account for the influence of the second-order factors at age $t-1$ on age t; these parameters were free to vary between ages ($-2LL = 223543.5, AIC = 1203421, df = 76581$). This model had fewer degrees of freedom than the previous model and a comparison suggested that constraining the dynamics to zero as in the previous model results in a significant loss of fit ($df_{difference} = 12, p < .001$).

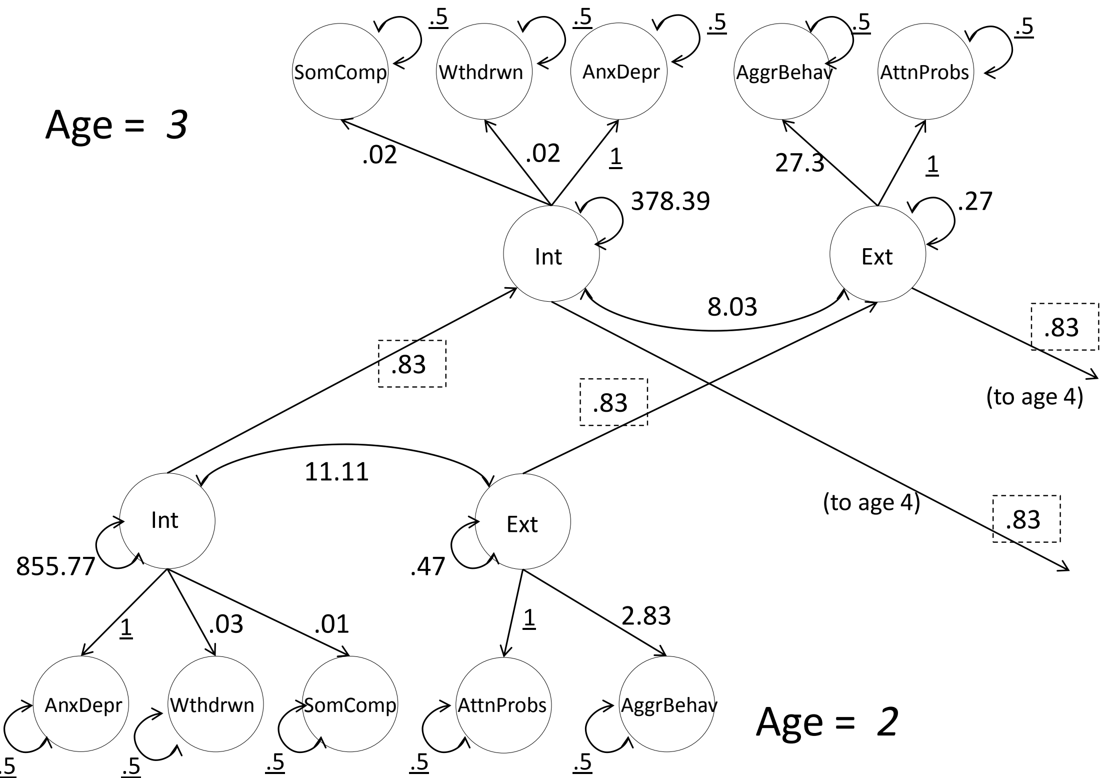
From the model with freely estimated dynamics, I fit a model with the second-order factor dynamics constrained to invariance across groups. Because one year (age 6) was skipped in the study, I did not constrain the autoregressions between the second-order factors at age 5 and age 7 to invariance with the rest of the second-order factor regressions across one year. Therefore, this model estimated two values for the *internalizing* factor autoregression (one for the relationships across one year, and a different estimate for the relationship across 2 years) and likewise two values for the *externalizing* autoregressions. This model did not result in a significant loss of fit ($-2LL = 219163.8, AIC = 1199025, df = 76589, df_{difference} = 8, p \approx 1$).

The next model featured the constrained second-order dynamics with the addition of second-order factor loadings constrained across ages. This model resulted in a significant loss of fit ($-2LL = 219538.5, AIC = 1199364, df = 76607, df_{difference} = 18, p < .001$), therefore the final model contained second-order dynamics constrained across groups without any constraints on the second-order factor loadings (see figures 4.2, 4.3, 4.4, 4.5 for final second-order model structure and coefficients. Parameters in dashed boxes were constrained to invariance between groups, while underlined parameters were fixed to the value displayed.).

4.3 Discussion

The HOIIF allows us to address relationships between factors that otherwise would be virtually inaccessible due to measurement differences. In the empirical applications of the HOIIF, I was able to establish a model of latent cognitive ability that was cohesive with cognitive theory, featured some variation in measurement between different literacy and income levels, but identified the same four factors and single second-order factor between the groups. The results suggest that we are measuring the same latent factors in our groups. Because we have evidence to support that we are measuring the same factors in our different groups, we can see that processing speed (Gs) in particular is more differentiated with increasing literacy level, which is in line with theories about differentiation of cognitive domains as ability increases (e.g. Tucker-Drob, 2009). Here, we have evidence that as

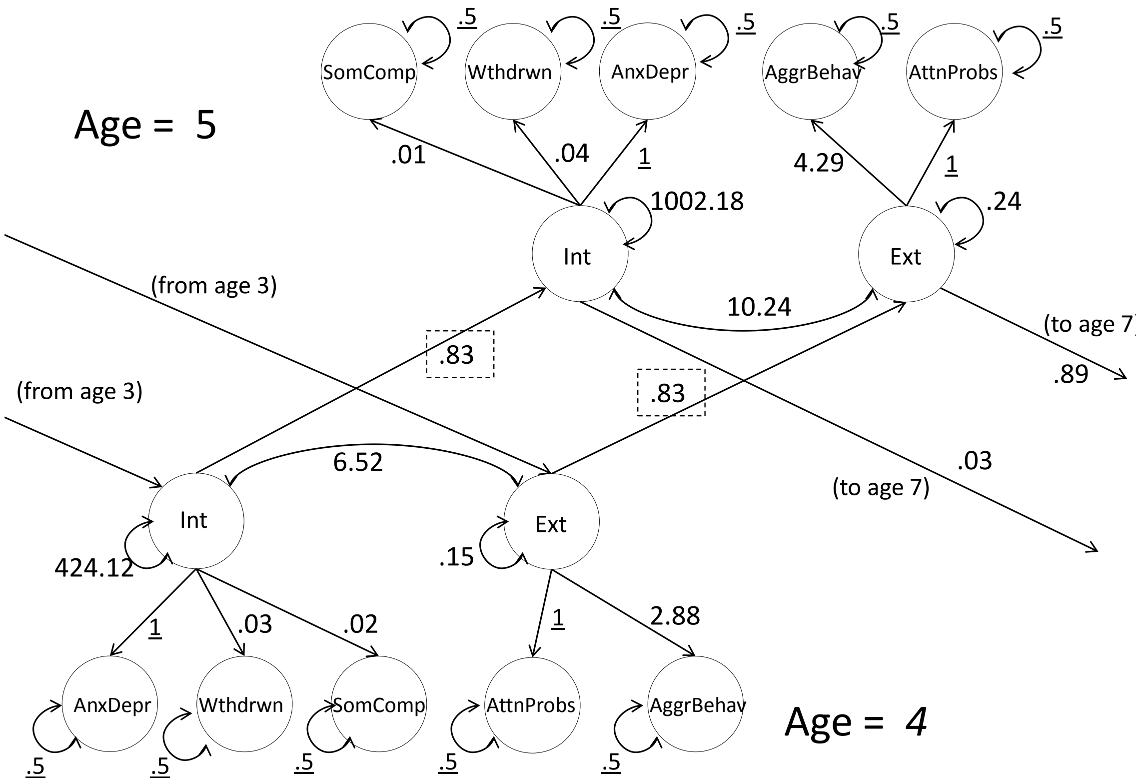
Figure 4.2. Second-order model structure with final model estimates for ages 2 and 3. Underlined values were fixed to the displayed value; non-underlined items were estimated. Values in boxes were constrained to invariance between groups. Because of the large sample size, all estimates were significant.



literacy level increases, fewer scales demand Gs as a resource.

Results for the longitudinal data are more tentative. The results suggest potential estimation/specification problems with the data, most likely owing to either indeterminacy in the data because of low agreement rates for some of the items, and/or–very likely–that the first-order factor variances are not invariant between factors or across time in the data. Therefore, any interpretation of the current analysis is extremely preliminary; proper application of the results would require close scrutiny and modification of the model to correct or eliminate questionable items and perhaps an examination of the model gradients. To the extent that we can interpret the current results, there is some support for autoregression in the factors; however, the results indicate that the second order factor structure might

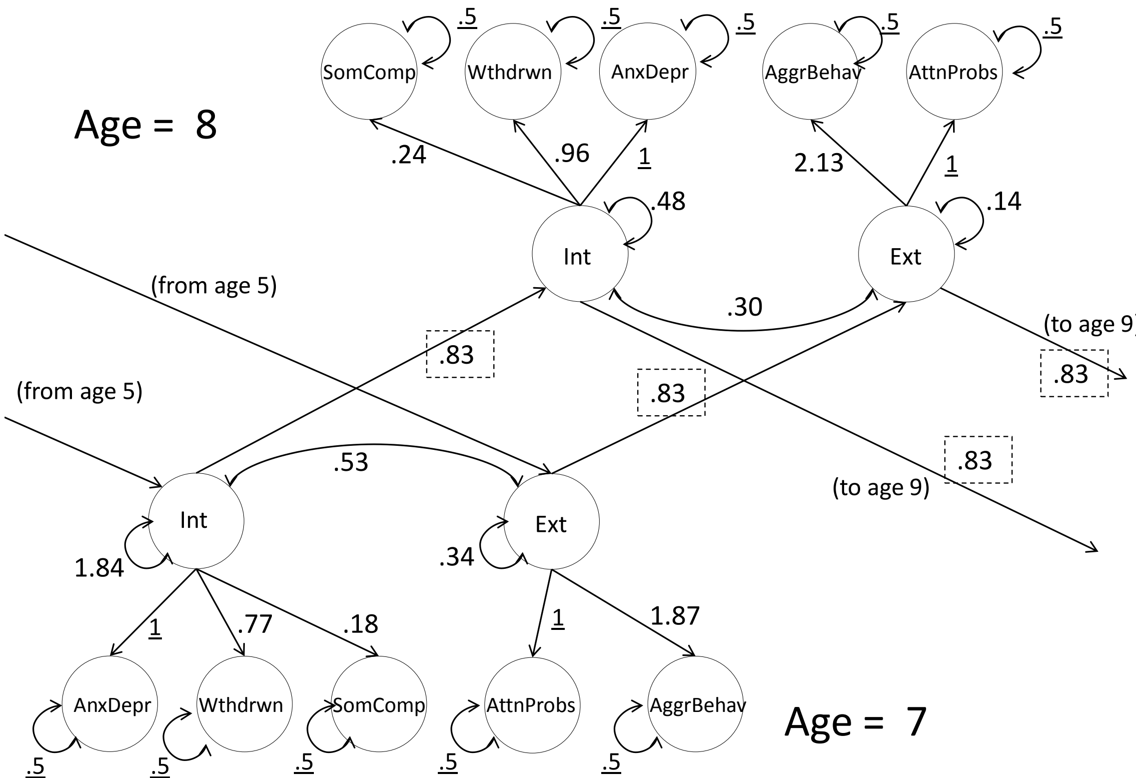
Figure 4.3. Second-order model structure with final model estimates for ages 4 and 5. Underlined values were fixed to the displayed value; non-underlined items were estimated. Values in boxes were constrained to invariance between groups. Because of the large sample size, all estimates were significant.



not be invariant between groups. The results suggest that the internalizing factor measurement strengthens with increasing age; however, because of the estimation indeterminacy this should not be taken as more than an avenue for further investigation. A tangible next step for the Early Steps study might be to free the first-order factor variances between the factors, or to focus on one factor in order to reign in the scope of the analysis and take a deeper look at a subset of the data.

While interpreting the results of the longitudinal model is tentative, the exercise provided important insights for implementing the HOIIF. In terms of the procedure, this exercise reiterated the need to include dynamics as a separate step and interpret model comparison with the first model with no dynamics as a constraint on the dynamic model.

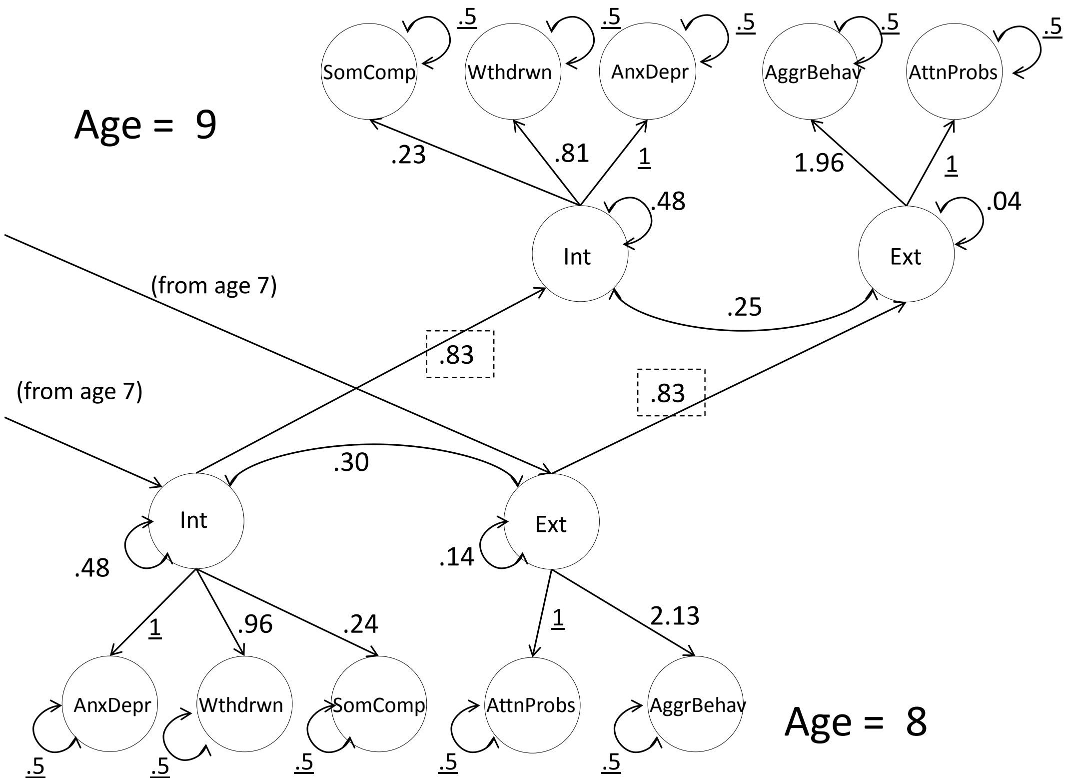
Figure 4.4. Second-order model structure with final model estimates for ages 7 and 8. Underlined values were fixed to the displayed value; non-underlined items were estimated. Values in boxes were constrained to invariance between groups. Because of the large sample size, all estimates were significant.



That the factors autoregress in this way is a testable assumption; future applications of the HOIIF could also test include tests of different models of dynamics. Additionally, this exercise highlighted the interaction of data size and model complexity on model estimation and fit. That model complexity and large data sets can affect estimation and fit statistics is a known issue in SEM (e.g. Gerbing and Anderson, 1985) but does deserve to be specifically noted in HOIIF due to the inherent complexity and potential for misspecification by nature of the HOIIF process. In the same vein, the likelihood ratio test was a stringent standard for the longitudinal model given the sensitivity of χ^2 to large sample size and degrees of freedom difference.

Generally, an avenue of consideration that the HOIIF affords us is in interpreting

Figure 4.5. Second-order model structure with final model estimates for ages 8 and 9. Underlined values were fixed to the displayed value; non-underlined items were estimated. Values in boxes were constrained to invariance between groups. Because of the large sample size, all estimates were significant.



differences in performance between groups. For example, in the cross sectional data an examination of the short term memory (Gsm) factor loadings indicates that in the BHL group, an individual who has high performance on the attention task will have a higher score on Gsm than an individual in the same group who does equally on the digit span forward and back tasks but does more poorly on the attention task. Yet if we were to find two new individuals in the AHL group with the same exact performance as these previous individuals, we would predict that they would have almost equal scores on Gsm.

Stepping through the application of the HOIIF to observed data highlighted a few key considerations for this method. The importance of a strong theory in which to anchor the factor model is evident; without it, the exercise of selecting a factor structure, fitting the model, and interpreting any results becomes challenging and the potential for over fitting (by which I mean capturing unique idiosyncrasies of the particular dataset) increases. For this reason, differences in groups should be based on some evidence, with some understanding of how the underlying structure might vary between groups in order to be able to prevent such erring into overfit.

In addition, the estimation of the anchor model with other loadings free can be a challenge. Because of the number of free parameters, if the latent structure is not crisp then optimization can become a difficult task. This can result in a much longer process, iterating through starting values and constraints in order to support the optimization. In the case of the cross-sectional data set, prior to setting bounds on the factor loading values, some starting values would push the parameter estimates into incredibly large values while others would not.

5. General Discussion

5.1 Summary

Returning to the goals for the simulation studies, I addressed the following considerations:

- How well does this method recover parameters? In other words, how often and how closely do the point estimates match the parameters used to create the data? And how much bias is present in the estimates?
  - The FO Unconstrained models for the longitudinal data did show some slight bias, presumably due to the unmodeled autoregression present in the data. Overall, in both the cross-sectional as well as longitudinal simulations, parameters were consistently well recovered–EASEs indicated deviations of around .05 or smaller from population loading values when sample sizes were large and models were correctly specified (i.e. a constrained second-order model being used in second-order invariant data).
- How specific is it? How often does it reject the models it should reject and fail to reject the models it should fail to reject?
  - The HOIIF appears to have an acceptable rate of rejecting/failing to reject models. For example, at a large sample size in cross-sectional analysis, the HOIIF rejects 85% or more of the models it should reject at the first-order constrained step, while failing to reject 79% or more of the models that it should fail to reject. The performance was even stronger in the longitudinal simulations.

- In what ways do the relevant fit statistics vary relative to accuracy of point estimates, bias, and specificity? How does sample size affect these considerations?
  - As expected, as error in point estimates increases model fit decreases (as does specificity)–for example, by sample size or by underlying structure. Likewise, what bias exists in the estimates decreases with increased sample size and thereby increased fit.
- How does the structure of the measurement model affect the previous considerations?
  - As expected, the metrically non-invariant models fit well when the second-order model is invariant between groups, and fit poorly when the second-order model is non-invariant. The configurally non-invariant models performed similarly, showing better fit and more accurate and less biased estimates when the second-order model is invariant between groups.
- How often does the method fail to converge? Do convergence issues depend on sample size, measurement model, or the second-order variance condition? Does varying the starting values affect convergence?
  - In the simulations, there were a number of models with convergence issues depending on sample size, data type, and model being fit. Issues with convergence were more likely when the sample size was small and when fitting a first-order model in data with a non-invariant second-order model. Varying starting values did not meaningfully help with this particular issue.

5.2 Limitations and Future Directions

One word of caution in implementing this method is that the HOIIF requires the steps to be followed in order, which was made clear in the simulation studies. For example, at a medium sample size, 72% of the longitidinal simulation models with metric and second-order non-invariance failed to be rejected when they should have been via a likelihood ratio comparison with the null variance-covariance model. However, most of these models would have been rejected at the step of constraining the first-order model, at which point only 15% of the models failed to be rejected. Thus, the step-wise structure of the HOIIF method is an important element in encouraging the accuracy of the results.

Along the same lines of considerations in implementation, the first-order model must be scaled by the first-order factor variances. Although this means that it becomes a default constraint on the second-order factor model, without this restriction the first-order factors will be scaled by the manifest loadings. Because the manifest measurement structure is allowed to (presumably *expected* to, even) vary between the groups, this scaling of the factors all but guarantees poor model fit when fixing the factor variance and covariance matrix between groups, regardless of the underlying structure.

On the topic of model fit, the fit statistics can only be as good as the current methods of establishing model fit in SEM. The same limitations that exist in SEM exist here, and given the potential for model complexity in the HOIIF framework, careful consideration must be given to model selection and fit statistics. Given the results of the current study, future implementations of the HOIIF might benefit from inclusion and review of *all* possible fit statistics.

While the present study touches on the issue of comparing between groups, whether or not the HOIIF can be implemented in such a way as to allow direct between-group comparisons on factors remains to be determined. The present work suggests that the HOIIF allows us to compare between groups the nature of differences within groups; however, it does not cover the case of directly comparing factor scores between groups. So while now we might be able to ask to what extent a difference in Internalizing factor scores between at-risk three year olds who received two different interventions varies from the difference in Internalizing factor scores between at-risk 12 year olds, further work needs to be done in order to determine whether we can directly ask whether that sample of at-risk three year olds actually have different scores on Internalizing than the12 year olds.

6. Conclusions

With the empirical studies to bolster understanding of the HOIIF in practice, I have established that we can use the HOIIF method to identify invariant higher-order structure in non-invariant measurement conditions with careful consideration given to model complexity. Because this method allows for first-order measurement non-invariance, the HOIIF permits filtering out biases such as certain method or item biases. For example in the hypothetical computer-based assessment discussed by Byrne (2003), if experience with computers resulted in differences in the first-order measurement model values (factor loadings) but the second-order structure remained consistent between the groups, then the researchers would still be able to identify the latent factors and compare across groups using an HOIIF model.

Thus, the HOIIF is a much needed method to allow measurement non-invariance in the understanding of latent constructs. With the current attention paid to cross-cultural and longitudinal work, the HOIIF broadens our capabilities of understanding and interpreting latent constructs in the context of inescapable differences in manifest measurement, whether they be due to developmental change, method bias, cultural differences, or any other source.

In order to productively use a method such as the HOIIF, we must first accept the notion that while invariance is critical (Cheung and Rensvold, 1999; Schmitt and Kuljanin, 2008; Widaman and Reise, 1997; Steenkamp and Baumgartner, 1998), it need not exist solely on the level of manifest measurement. Building on the groundwork laid by Byrne (2003); Steenkamp and Baumgartner (1998), who discussed the validity of configural invariance even when items are non-invariant, the current work supports that a strong theory with anchors to identify factors can give us access to otherwise unobtainable inquiries.

Until now, performance of the HOIIF method had not been fully explored. The results of the current study indicate that the HOIIF is a viable option for establishing factorial invariance, a key concept in striving to understand performance between individuals or groups. The HOIIF allows us to understand differences and make comparisons with minimal invariance in the manifest aspects of an assessment. Methodologically, we can now ask questions about performance between groups not only when measurement is invariant, but also when there are differences in measurement between the groups. In cases where higher-order factor structure is invariant despite non-invariant first-order measurement, the HOIIF opens a previously inaccessible route to make cross-cultural comparisons on latent factors, to investigate the effectiveness of interventions in varying age groups for which the measurement instrument is different by design, and other analogous situations.

A. Appendix: Simulation Results

Table A.1. Cross-sectional Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for first-order, unconstrained model for group = 1

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.34	0.55	0.42	0.071	[0.068, 0.075]
	0.43	0.31	0.29		
	0.54	0.72	0.54		
500	0.37	0.59	0.45	0.01	[0.009, 0.011]
	0.47	0.33	0.30		
	0.64	0.81	0.59		
1500	0.37	0.59	0.46	0.003	[0.003, 0.004]
	0.48	0.34	0.31		
	0.64	0.81	0.60		
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.35	0.57	0.40	0.071	[0.068, 0.074]
	0.46	0.34	0.27		
	0.57	0.75	0.48		
500	0.37	0.59	0.46	0.011	[0.01, 0.012]
	0.48	0.34	0.31		
	0.64	0.81	0.60		
1500	0.37	0.59	0.46	0.004	[0.004, 0.005]
	0.48	0.34	0.31		
	0.64	0.81	0.59		
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.35	0.56	0.38	0.073	[0.07, 0.077]
	0.44	0.32	0.27		
	0.55	0.72	0.49		
500	0.37	0.59	0.45	0.01	[0.009, 0.011]
	0.48	0.34	0.30		
	0.65	0.81	0.59		
1500	0.37	0.58	0.46	0.003	[0.003, 0.004]
	0.48	0.34	0.31		
	0.64	0.81	0.60		
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.35	0.56	0.40	0.067	[0.063, 0.07]
	0.44	0.33	0.29		
	0.56	0.75	0.51		
500	0.36	0.59	0.45	0.015	[0.014, 0.017]
	0.47	0.34	0.31		
	0.64	0.80	0.59		
1500	0.37	0.59	0.46	0.003	[0.002, 0.003]
	0.48	0.34	0.31		
	0.64	0.81	0.60		

Table A.2. Cross-sectional Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for first-order, unconstrained model for group = 2

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.59	0.55	0.64	0.087	[0.082, 0.091]
	0.64	0.51	0.70		
	0.26	0.44	0.32		
500	0.60	0.65	0.70	0.02	[0.018, 0.023]
	0.71	0.59	0.80		
	0.27	0.50	0.35		
1500	0.63	0.64	0.70	0.013	[0.01, 0.015]
	0.74	0.58	0.80		
	0.28	0.49	0.36		
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.61	0.62	0.68	0.054	[0.05, 0.057]
	0.69	0.57	0.76		
	0.27	0.49	0.34		
500	0.66	0.66	0.70	0.008	[0.007, 0.008]
	0.78	0.60	0.80		
	0.30	0.51	0.36		
1500	0.65	0.66	0.70	0.004	[0.003, 0.005]
	0.78	0.60	0.79		
	0.30	0.51	0.36		
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.39	0.74	0.26	0.118	[0.113, 0.124]
	0.38	0.58	0.37		
	0.41	0.77	0.31		
500	0.57	0.73	0.27	0.043	[0.039, 0.048]
	0.57	0.56	0.49		
	0.60	0.75	0.34		
1500	0.65	0.77	0.28	0.012	[0.01, 0.014]
	0.65	0.59	0.47		
	0.69	0.79	0.33		
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.61	0.71	0.24	0.067	[0.063, 0.071]
	0.62	0.55	0.37		
	0.64	0.72	0.28		
500	0.65	0.78	0.28	0.011	[0.01, 0.012]
	0.66	0.59	0.47		
	0.69	0.80	0.33		
1500	0.66	0.78	0.30	0.002	[0.002, 0.003]
	0.66	0.60	0.51		
	0.70	0.80	0.36		

Table A.3. Cross-sectional Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for second-order model with constrained second-order factor loadings and variance, and constrained first-order factor variances, for group = 1.

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.58	0.77	0.72	0.04	[0.039, 0.042]
	0.51	0.50	0.43		
	0.68	0.92	0.73		
500	0.51	0.63	0.61	0.01	[0.01, 0.01]
	0.49	0.36	0.32		
	0.63	0.85	0.59		
1500	0.50	0.61	0.60	0.006	[0.006, 0.006]
	0.49	0.34	0.31		
	0.62	0.83	0.58		
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.55	0.69	0.65	0.034	[0.033, 0.035]
	0.51	0.42	0.38		
	0.68	0.87	0.66		
500	0.52	0.63	0.60	0.009	[0.009, 0.009]
	0.49	0.37	0.32		
	0.62	0.82	0.60		
1500	0.51	0.62	0.59	0.006	[0.006, 0.006]
	0.49	0.35	0.32		
	0.62	0.81	0.58		
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.56	0.82	0.54	0.042	[0.04, 0.044]
	0.52	0.51	0.63		
	0.67	0.90	0.89		
500	0.53	0.69	0.40	0.011	[0.01, 0.011]
	0.49	0.36	0.32		
	0.63	0.80	0.69		
1500	0.53	0.68	0.39	0.007	[0.006, 0.007]
	0.49	0.35	0.31		
	0.61	0.79	0.67		
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.56	0.75	0.45	0.036	[0.034, 0.037]
	0.50	0.44	0.53		
	0.66	0.86	0.81		
500	0.53	0.69	0.39	0.009	[0.009, 0.01]
	0.49	0.37	0.33		
	0.63	0.80	0.65		
1500	0.53	0.68	0.39	0.006	[0.006, 0.006]
	0.49	0.36	0.31		
	0.62	0.79	0.64		

Table A.4. Cross-sectional Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for second-order model with constrained second-order factor loadings and variance, and constrained first-order factor variances, for group = 2.

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.61	0.78	0.75	0.026	[0.025, 0.027]
	0.86	0.77	0.90		
	0.34	0.70	0.47		
500	0.55	0.64	0.65	0.008	[0.007, 0.008]
	0.87	0.64	0.84		
	0.28	0.54	0.36		
1500	0.54	0.62	0.64	0.005	[0.005, 0.005]
	0.86	0.61	0.83		
	0.28	0.52	0.36		
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.60	0.71	0.69	0.022	[0.021, 0.023]
	0.85	0.69	0.88		
	0.33	0.60	0.42		
500	0.56	0.65	0.64	0.006	[0.006, 0.006]
	0.84	0.62	0.83		
	0.29	0.53	0.36		
1500	0.55	0.64	0.64	0.003	[0.003, 0.003]
	0.83	0.60	0.82		
	0.29	0.51	0.35		
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.60	0.85	0.53	0.03	[0.029, 0.031]
	0.71	0.76	0.78		
	0.75	0.91	0.67		
500	0.58	0.75	0.39	0.007	[0.007, 0.007]
	0.67	0.61	0.49		
	0.71	0.83	0.37		
1500	0.57	0.73	0.38	0.004	[0.004, 0.004]
	0.66	0.60	0.47		
	0.71	0.82	0.35		
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.60	0.80	0.44	0.025	[0.024, 0.026]
	0.70	0.68	0.72		
	0.74	0.86	0.58		
500	0.57	0.75	0.38	0.006	[0.005, 0.006]
	0.66	0.60	0.52		
	0.71	0.82	0.37		
1500	0.58	0.74	0.38	0.003	[0.003, 0.003]
	0.66	0.60	0.50		
	0.70	0.81	0.36		

Table A.5. Cross-sectional Simulations: Mean estimated second-order factor loadings, error, and associated 95% confidence interval for second-order model with constrained second-order factor loadings and variance, and constrained first-order factor variances.

Sample Size	Λ_{F2}	EASE	CI	EASE (group 2)	CI (group2)
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.47 0.78 0.60	0.056	[0.052, 0.06]	0.055	[0.052, 0.059]
500	0.41 0.83 0.57	0.022	[0.021, 0.023]	0.016	[0.016, 0.017]
1500	0.41 0.83 0.56	0.017	[0.016, 0.017]	0.01	[0.01, 0.011]
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.56 0.86 0.70	0.028	[0.026, 0.029]	NA	NA
500	0.53 0.89 0.67	0.006	[0.006, 0.006]	NA	NA
1500	0.53 0.89 0.67	0.002	[0.002, 0.002]	NA	NA
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.49 0.74 0.67	0.073	[0.068, 0.077]	0.076	[0.072, 0.08]
500	0.43 0.79 0.64	0.021	[0.02, 0.022]	0.02	[0.02, 0.021]
1500	0.42 0.79 0.64	0.014	[0.013, 0.014]	0.012	[0.012, 0.013]
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.58 0.84 0.71	0.037	[0.035, 0.04]	NA	NA
500	0.55 0.88 0.69	0.007	[0.007, 0.008]	NA	NA
1500	0.55 0.89 0.68	0.003	[0.003, 0.003]	NA	NA

Table A.6. Cross-sectional Simulations: Second-order factor loading population values for both groups in the data with invariant second-order structure, or for group 1 in the data with non-invariant second-order structure.

	Λ_{F2}
$F1_a$	0.55
$F1_b$	0.88
$F1_c$	0.69

Table A.7. Cross-sectional Simulations: Second-order factor loading population values for group 2 in the data with non-invariant second-order structure.

	Λ_{F2}
$F1_a$	0.34
$F1_b$	0.71
$F1_c$	0.54

Table A.8. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for first-order, unconstrained model at time = 1

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.41	0.47	0.52	0.009	[0.009, 0.01]
	0.58	0.36	0.80		
	0.70	0.70	0.52		
500	0.41	0.47	0.52	0.001	[0.001, 0.001]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
1500	0.41	0.47	0.52	0	[0, 0]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.41	0.47	0.53	0.009	[0.008, 0.009]
	0.59	0.38	0.80		
	0.70	0.70	0.53		
500	0.41	0.47	0.52	0.002	[0.002, 0.002]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
1500	0.41	0.47	0.52	0	[0, 0.001]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.41	0.48	0.51	0.011	[0.01, 0.011]
	0.58	0.37	0.79		
	0.70	0.69	0.52		
500	0.41	0.47	0.52	0.001	[0.001, 0.001]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
1500	0.41	0.47	0.52	0	[0, 0]
	0.58	0.37	0.79		
	0.70	0.69	0.53		

Table A.9. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for first-order, unconstrained model at time = 1

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.41	0.48	0.52	0.034	[0.033, 0.036]
	0.59	0.38	0.80		
	0.71	0.70	0.53		
500	0.41	0.47	0.52	0.004	[0.004, 0.005]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
1500	0.41	0.47	0.52	0.002	[0.001, 0.002]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.41	0.48	0.52	0.028	[0.027, 0.029]
	0.58	0.37	0.80		
	0.72	0.70	0.53		
500	0.41	0.47	0.52	0.004	[0.004, 0.004]
	0.58	0.37	0.78		
	0.70	0.69	0.54		
1500	0.41	0.47	0.52	0.001	[0.001, 0.001]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.41	0.49	0.52	0.029	[0.028, 0.03]
	0.58	0.36	0.80		
	0.72	0.71	0.53		
500	0.41	0.47	0.52	0.005	[0.004, 0.005]
	0.58	0.37	0.80		
	0.70	0.70	0.53		
1500	0.41	0.47	0.52	0.001	[0.001, 0.001]
	0.58	0.37	0.79		
	0.70	0.69	0.53		

Table A.10. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for first-order, unconstrained model at time = 2

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.63	0.69	0.41	0.009	[0.009, 0.01]
	0.58	0.88	0.63		
	0.34	0.80	0.68		
500	0.62	0.69	0.41	0.002	[0.002, 0.002]
	0.55	0.87	0.63		
	0.34	0.80	0.67		
1500	0.62	0.69	0.41	0.001	[0.001, 0.001]
	0.56	0.88	0.62		
	0.34	0.80	0.67		
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.63	0.68	0.40	0.01	[0.01, 0.011]
	0.56	0.88	0.64		
	0.34	0.80	0.66		
500	0.62	0.69	0.41	0.002	[0.002, 0.002]
	0.55	0.88	0.62		
	0.34	0.80	0.67		
1500	0.62	0.69	0.41	0.001	[0.001, 0.001]
	0.56	0.88	0.63		
	0.34	0.80	0.67		
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.63	0.69	0.40	0.009	[0.009, 0.009]
	0.57	0.87	0.62		
	0.34	0.81	0.68		
500	0.61	0.69	0.41	0.002	[0.002, 0.002]
	0.56	0.88	0.63		
	0.35	0.80	0.67		
1500	0.62	0.69	0.41	0.001	[0.001, 0.001]
	0.55	0.88	0.63		
	0.35	0.80	0.67		

Table A.11. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for first-order, unconstrained model at time = 2

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.63	0.55	0.34	0.037	[0.036, 0.038]
	0.69	0.89	0.80		
	0.40	0.63	0.68		
500	0.61	0.55	0.34	0.017	[0.016, 0.017]
	0.69	0.88	0.81		
	0.40	0.63	0.67		
1500	0.62	0.56	0.34	0.014	[0.014, 0.014]
	0.69	0.88	0.81		
	0.41	0.63	0.67		
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.62	0.54	0.34	0.042	[0.041, 0.043]
	0.70	0.87	0.82		
	0.40	0.63	0.68		
500	0.61	0.56	0.34	0.016	[0.016, 0.017]
	0.69	0.88	0.80		
	0.40	0.63	0.67		
1500	0.62	0.55	0.34	0.014	[0.014, 0.014]
	0.69	0.88	0.80		
	0.41	0.62	0.67		
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.63	0.55	0.34	0.041	[0.039, 0.042]
	0.69	0.88	0.82		
	0.41	0.63	0.68		
500	0.62	0.55	0.34	0.017	[0.016, 0.017]
	0.69	0.88	0.80		
	0.41	0.63	0.67		
1500	0.62	0.55	0.34	0.014	[0.013, 0.014]
	0.69	0.88	0.80		
	0.41	0.63	0.67		

Table A.12. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for first-order, unconstrained model at time = 3

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.81	0.40	0.32	0.01	[0.01, 0.01]
	0.31	0.82	0.84		
	0.33	0.72	0.79		
500	0.82	0.40	0.32	0.002	[0.002, 0.002]
	0.30	0.81	0.83		
	0.32	0.72	0.78		
1500	0.82	0.40	0.32	0.001	[0.001, 0.001]
	0.30	0.81	0.83		
	0.32	0.72	0.78		
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.86	0.41	0.34	0.011	[0.01, 0.011]
	0.32	0.85	0.86		
	0.33	0.75	0.82		
500	0.86	0.42	0.33	0.002	[0.002, 0.002]
	0.31	0.84	0.86		
	0.32	0.74	0.81		
1500	0.85	0.42	0.33	0.001	[0.001, 0.001]
	0.31	0.84	0.86		
	0.33	0.74	0.81		
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.86	0.42	0.33	0.01	[0.01, 0.011]
	0.32	0.84	0.88		
	0.34	0.75	0.80		
500	0.87	0.42	0.33	0.002	[0.002, 0.002]
	0.31	0.84	0.86		
	0.32	0.74	0.81		
1500	0.85	0.42	0.33	0.001	[0.001, 0.001]
	0.32	0.84	0.86		
	0.33	0.74	0.81		

Table A.13. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for first-order, unconstrained model at time = 3

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.81	0.32	0.32	0.12	[0.118, 0.122]
	0.81	0.74	0.42		
	0.78	0.33	0.83		
500	0.82	0.30	0.31	0.098	[0.097, 0.099]
	0.81	0.73	0.40		
	0.78	0.32	0.84		
1500	0.82	0.30	0.31	0.093	[0.093, 0.094]
	0.81	0.72	0.41		
	0.78	0.32	0.83		
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.85	0.32	0.33	0.124	[0.122, 0.126]
	0.84	0.76	0.43		
	0.81	0.34	0.86		
500	0.85	0.32	0.32	0.101	[0.1, 0.102]
	0.84	0.76	0.42		
	0.80	0.34	0.87		
1500	0.85	0.31	0.33	0.098	[0.097, 0.098]
	0.84	0.75	0.42		
	0.81	0.33	0.86		
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.84	0.33	0.32	0.124	[0.122, 0.126]
	0.83	0.74	0.43		
	0.81	0.34	0.87		
500	0.85	0.32	0.32	0.101	[0.1, 0.102]
	0.84	0.75	0.42		
	0.81	0.34	0.88		
1500	0.85	0.31	0.32	0.097	[0.097, 0.098]
	0.84	0.75	0.42		
	0.81	0.34	0.86		

Table A.14. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for autoregressive model with across-time constraints on factor variances, covariances, and autoregressions at time = 1

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.41	0.47	0.52	0.009	[0.008, 0.009]
	0.58	0.36	0.80		
	0.71	0.70	0.52		
500	0.41	0.47	0.52	0.001	[0.001, 0.001]
	0.58	0.37	0.79		
	0.71	0.69	0.53		
1500	0.41	0.47	0.52	0	[0, 0]
	0.58	0.37	0.79		
	0.70	0.70	0.53		
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.41	0.47	0.52	0.008	[0.007, 0.008]
	0.59	0.38	0.79		
	0.70	0.70	0.53		
500	0.41	0.47	0.52	0.001	[0.001, 0.001]
	0.58	0.38	0.79		
	0.70	0.69	0.53		
1500	0.41	0.47	0.52	0	[0, 0]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.40	0.48	0.52	0.009	[0.009, 0.009]
	0.58	0.37	0.79		
	0.71	0.68	0.53		
500	0.41	0.47	0.52	0.001	[0.001, 0.001]
	0.58	0.37	0.79		
	0.71	0.69	0.53		
1500	0.41	0.47	0.52	0	[0, 0]
	0.58	0.37	0.79		
	0.70	0.69	0.53		

Table A.15. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for autoregressive model with across-time constraints on factor variances, covariances, and autoregressions at time = 1

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.41	0.48	0.52	0.032	[0.03, 0.033]
	0.59	0.38	0.80		
	0.72	0.70	0.53		
500	0.41	0.47	0.52	0.004	[0.004, 0.004]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
1500	0.41	0.47	0.52	0.001	[0.001, 0.002]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.41	0.47	0.52	0.023	[0.023, 0.024]
	0.58	0.37	0.81		
	0.71	0.71	0.52		
500	0.41	0.47	0.52	0.004	[0.004, 0.004]
	0.58	0.37	0.78		
	0.70	0.69	0.53		
1500	0.41	0.47	0.52	0.001	[0.001, 0.001]
	0.58	0.37	0.79		
	0.70	0.69	0.53		
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.40	0.48	0.52	0.025	[0.024, 0.026]
	0.59	0.36	0.79		
	0.72	0.71	0.53		
500	0.41	0.47	0.52	0.004	[0.004, 0.004]
	0.58	0.37	0.79		
	0.71	0.69	0.53		
1500	0.41	0.47	0.52	0.001	[0.001, 0.001]
	0.58	0.37	0.79		
	0.70	0.69	0.53		

Table A.16. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for autoregressive model with across-time constraints on factor variances, covariances, and autoregressions at time = 2

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.59	0.67	0.40	0.026	[0.026, 0.027]
	0.56	0.86	0.62		
	0.34	0.77	0.65		
500	0.59	0.67	0.40	0.02	[0.02, 0.02]
	0.54	0.85	0.61		
	0.34	0.78	0.66		
1500	0.60	0.67	0.40	0.02	[0.02, 0.02]
	0.54	0.86	0.61		
	0.34	0.79	0.65		
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.60	0.66	0.38	0.026	[0.025, 0.026]
	0.53	0.83	0.60		
	0.32	0.76	0.63		
500	0.59	0.66	0.39	0.02	[0.02, 0.02]
	0.53	0.84	0.60		
	0.33	0.77	0.64		
1500	0.59	0.66	0.39	0.019	[0.019, 0.019]
	0.53	0.84	0.60		
	0.33	0.77	0.64		
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.58	0.66	0.38	0.025	[0.024, 0.025]
	0.54	0.83	0.59		
	0.33	0.78	0.65		
500	0.58	0.66	0.39	0.019	[0.019, 0.02]
	0.53	0.84	0.60		
	0.33	0.77	0.64		
1500	0.58	0.66	0.39	0.019	[0.019, 0.019]
	0.53	0.84	0.60		
	0.33	0.77	0.64		

Table A.17. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for autoregressive model with across-time constraints on factor variances, covariances, and autoregressions at time = 2

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.61	0.53	0.33	0.071	[0.069, 0.072]
	0.67	0.86	0.79		
	0.39	0.61	0.66		
500	0.60	0.54	0.33	0.052	[0.051, 0.052]
	0.67	0.86	0.79		
	0.40	0.61	0.65		
1500	0.60	0.54	0.34	0.049	[0.049, 0.05]
	0.67	0.86	0.79		
	0.40	0.61	0.65		
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.59	0.52	0.33	0.068	[0.067, 0.07]
	0.67	0.84	0.78		
	0.38	0.59	0.64		
500	0.58	0.53	0.33	0.05	[0.049, 0.05]
	0.66	0.84	0.77		
	0.39	0.60	0.64		
1500	0.59	0.53	0.33	0.047	[0.047, 0.048]
	0.66	0.84	0.77		
	0.39	0.60	0.64		
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.59	0.53	0.33	0.067	[0.066, 0.069]
	0.66	0.84	0.78		
	0.39	0.60	0.64		
500	0.59	0.53	0.33	0.05	[0.049, 0.05]
	0.65	0.84	0.77		
	0.39	0.60	0.64		
1500	0.59	0.53	0.33	0.047	[0.047, 0.048]
	0.66	0.84	0.77		
	0.39	0.60	0.64		

Table A.18. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for autoregressive model with across-time constraints on factor variances, covariances, and autoregressions at time = 3

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.82	0.39	0.31	0.034	[0.034, 0.035]
	0.30	0.79	0.82		
	0.31	0.70	0.77		
500	0.82	0.40	0.32	0.027	[0.027, 0.028]
	0.29	0.79	0.81		
	0.30	0.70	0.76		
1500	0.82	0.39	0.32	0.026	[0.026, 0.026]
	0.29	0.79	0.81		
	0.30	0.70	0.76		
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.81	0.39	0.32	0.033	[0.032, 0.034]
	0.30	0.81	0.81		
	0.31	0.71	0.78		
500	0.81	0.40	0.32	0.027	[0.026, 0.027]
	0.30	0.80	0.82		
	0.31	0.71	0.77		
1500	0.81	0.40	0.32	0.026	[0.026, 0.026]
	0.30	0.80	0.82		
	0.31	0.71	0.77		
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.84	0.40	0.31	0.034	[0.033, 0.035]
	0.30	0.80	0.83		
	0.31	0.71	0.76		
500	0.86	0.40	0.32	0.028	[0.028, 0.029]
	0.29	0.80	0.82		
	0.30	0.71	0.77		
1500	0.84	0.40	0.32	0.027	[0.027, 0.027]
	0.29	0.80	0.82		
	0.30	0.71	0.77		

Table A.19. Longitudinal Simulations: Mean estimated first-order factor loadings, EASE, and associated 95% confidence interval for autoregressive model with across-time constraints on factor variances, covariances, and autoregressions at time = 3

Sample Size	Λ_{F1a}	Λ_{F1b}	Λ_{F1c}	EASE	CI
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.79	0.30	0.31	0.108	[0.106, 0.11]
	0.79	0.69	0.40		
	0.75	0.33	0.81		
500	0.80	0.30	0.31	0.089	[0.088, 0.09]
	0.79	0.69	0.39		
	0.76	0.32	0.81		
1500	0.80	0.30	0.30	0.085	[0.084, 0.085]
	0.79	0.69	0.40		
	0.76	0.32	0.81		
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.81	0.30	0.31	0.111	[0.108, 0.113]
	0.80	0.72	0.40		
	0.77	0.32	0.83		
500	0.81	0.30	0.31	0.091	[0.09, 0.092]
	0.80	0.72	0.40		
	0.77	0.32	0.82		
1500	0.81	0.30	0.31	0.088	[0.087, 0.088]
	0.80	0.71	0.40		
	0.77	0.32	0.82		
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.80	0.30	0.30	0.11	[0.108, 0.113]
	0.79	0.72	0.39		
	0.77	0.32	0.85		
500	0.81	0.30	0.31	0.092	[0.091, 0.093]
	0.79	0.71	0.40		
	0.77	0.32	0.84		
1500	0.81	0.30	0.31	0.088	[0.088, 0.089]
	0.80	0.71	0.40		
	0.77	0.32	0.83		

Table A.20. Longitudinal Simulations: Mean estimated "second-order" regressions (i.e. autoregressions), error, and associated 95% confidence interval for autoregressive model with across-time constraints on factor variances, covariances, and autoregressions. The data with invariant autoregressions have EASE and 95% CI relative to only the single set of population values, while the data with non-invariant autoregressions have EASE and 95% CI to both population values.

Sample Size	Λ_{ARs}	EASE	CI	EASE (time 2-3)	CI (time 2-3)
Measurement Model: metrically non-invariant, S-O Model: non-invariant					
100	0.22 0.22 0.22	0.019	[0.018, 0.02]	0.019	[0.018, 0.021]
500	0.23 0.22 0.22	0.009	[0.008, 0.009]	0.008	[0.008, 0.008]
1500	0.22 0.21 0.22	0.008	[0.007, 0.008]	0.006	[0.005, 0.006]
Measurement Model: metrically non-invariant, S-O Model: invariant					
100	0.30 0.30 0.30	0.016	[0.015, 0.017]	NA	NA
500	0.30 0.30 0.30	0.003	[0.003, 0.003]	NA	NA
1500	0.30 0.30 0.30	0.001	[0.001, 0.001]	NA	NA
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.30 0.31 0.30	0.017	[0.015, 0.018]	NA	NA
500	0.30 0.30 0.30	0.003	[0.003, 0.003]	NA	NA
1500	0.30 0.30 0.30	0.001	[0.001, 0.001]	NA	NA

Table A.21. Longitudinal Simulations: Mean estimated "second-order" regressions (i.e. autoregressions), error, and associated 95% confidence interval for autoregressive model with across-time constraints on factor variances, covariances, and autoregressions. The data with invariant autoregressions have EASE and 95% CI relative to only the single set of population values, while the data with non-invariant autoregressions have EASE and 95% CI to both population values.

Sample Size	Λ_{AR}	EASE	CI	EASE (time 2-3)	CI (time 2-3)
Measurement Model: configurally non-invariant, S-O Model: non-invariant					
100	0.21 0.23 0.22	0.021	[0.02, 0.022]	0.02	[0.019, 0.021]
500	0.21 0.22 0.22	0.009	[0.009, 0.009]	0.008	[0.007, 0.008]
1500	0.22 0.22 0.22	0.007	[0.007, 0.007]	0.006	[0.006, 0.006]
Measurement Model: configurally non-invariant, S-O Model: invariant					
100	0.29 0.30 0.29	0.016	[0.015, 0.017]	NA	NA
500	0.30 0.30 0.30	0.003	[0.003, 0.003]	NA	NA
1500	0.30 0.30 0.30	0.001	[0.001, 0.001]	NA	NA
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg					
100	0.30 0.29 0.30	0.016	[0.015, 0.017]	NA	NA
500	0.30 0.30 0.30	0.003	[0.003, 0.003]	NA	NA
1500	0.30 0.30 0.30	0.001	[0.001, 0.001]	NA	NA

B. Appendix: Model fit statistic tables

Table B.1. Cross-sectional Simulations, model fit information: Df difference, mean likelihood difference of the null variance/covariance model with constrained second-order model, mean likelihood difference of previous model (constrained loadings and errors) with constrained second-order model, AIC difference, mean RMSEA (SD), RMSEA skew

Sample Size	Diff df	Diff_{null} -2LL	Diff_{null} AIC	Diff_{prev} -2LL	Diff_{prev} AIC	RMSEA	RMSEA skew
Measurement Model: metrically non-invariant, S-O Model: non-invariant							
100	54	64.30	25372.45	1.40	-0.61	0.002 (0.002)	0.10
500	54	72.63	18180.76	1.54	-0.30	0.001 (0)	0.23
1500	54	108.47	216.80	2.91	1.07	0.001 (0)	0.14
Measurement Model: metrically non-invariant, S-O Model: invariant							
100	54	62.17	25370.34	1.75	-0.14	0.002 (0.002)	0.19
500	54	65.75	18173.97	2.81	1.06	0 (0)	0.16
1500	54	87.62	195.88	6.31	4.59	0 (0)	0.07
Measurement Model: configurally non-invariant, S-O Model: non-invariant							
100	54	62.99	25371.24	1.61	-0.40	0.002 (0.002)	0.22
500	54	74.75	18183.01	1.90	-0.15	0.001 (0)	0.18
1500	54	112.42	220.69	2.38	0.65	0.001 (0)	0.00
Measurement Model: configurally non-invariant, S-O Model: invariant							
100	54	62.56	25370.70	1.65	-0.25	0.002 (0.002)	0.05
500	54	67.70	18175.83	3.35	1.69	0.001 (0)	0.16
1500	54	92.40	200.60	8.52	6.71	0 (0)	0.19

Table B.2. Cross-sectional Simulations, additional model fit information: Percent of models with non-significant likelihood ratio test comparisons ($p > .05$). SO Unconstrained are second-order models featuring no constraints across groups, SO Constrained 1 models are the former models with between-group constraints on the second-order factor loadings, SO Constrained 2 models are the SO Con1 models with constraints on the second-order "manifest errors" (i.e. first-order factor variances) as well as the second-order loadings, and SO Constrained 3 models are SO Con2 models featuring additional across-group constraints on the second-order factor variance.

Size	1. SOUnconstr$_{null}$	2. SOConstr1$_{prev}$	3. SOConstr1$_{null}$	4. SOConstr2$_{prev}$	5. SOConstr2$_{null}$
Measurement Model: metrically non-invariant, S-O Model: non-invariant [$\% \to 0$]					
100	0.847	0.977	0.849	0.679	0.768
500	0.976	0.969	0.973	0.029	0.528
1500	0.978	0.931	0.967	0.000	0.000
Measurement Model: metrically non-invariant, S-O Model: invariant [$\% \to 1$]					
100	0.863	0.958	0.865	0.838	0.822
500	0.969	0.881	0.956	0.305	0.773
1500	0.963	0.712	0.923	0.000	0.236
Measurement Model: configurally non-invariant, S-O Model: non-invariant [$\% \to 0$]					
100	0.876	0.931	0.864	0.725	0.797
500	0.951	0.857	0.944	0.031	0.458
1500	0.976	0.585	0.909	0.000	0.000
Measurement Model: configurally non-invariant, S-O Model: invariant [$\% \to 1$]					
100	0.857	0.902	0.857	0.873	0.832
500	0.957	0.612	0.896	0.461	0.738
1500	0.977	0.109	0.691	0.014	0.188

Table B.3. Longitudinal Simulations, model fit information: Percent of models with non-significant likelihood ratio test ($p > .05$). Here, SO Unconstrained are first-order models featuring factor autoregression across time points, SO Constrained 1 models are the former models with between-group constraints on the first-order factor covariances

Size	1. SOUnconstr_{null}	2. SOConstr1_{null}	3. SOConstr1_{prev}
Measurement Model: metrically non-invariant, S-O Model: non-invariant [$\% \to 0$]			
100	0.567	0.505	0.714
500	0.916	0.794	0.159
1500	0.955	0.482	0.000
Measurement Model: metrically non-invariant, S-O Model: invariant [$\% \to 1$]			
100	0.497	0.492	0.811
500	0.848	0.848	0.868
1500	0.967	0.967	0.969
Measurement Model: metrically non-invariant, S-O Model: non-inv cov, invar autoreg [$\% \to 0$]			
100	0.503	0.434	0.639
500	0.843	0.733	0.145
1500	0.963	0.472	0.000
Measurement Model: configurally non-invariant, S-O Model: non-invariant [$\% \to 0$]			
100	0.517	0.468	0.721
500	0.931	0.762	0.120
1500	0.960	0.496	0.000
Measurement Model: configurally non-invariant, S-O Model: invariant [$\% \to 1$]			
100	0.536	0.524	0.905
500	0.919	0.920	0.962
1500	0.967	0.971	0.978
Measurement Model: configurally non-invariant, S-O Model: non-inv cov, invar autoreg [$\% \to 0$]			
100	0.522	0.475	0.735
500	0.940	0.805	0.158
1500	0.961	0.453	0.000

Figure B.1. Second-order model structure with final model estimates for ages 2 and 3. Underlined values were fixed to the displayed value; non-underlined items were estimated. Values in boxes were constrained to invariance between groups. Because of the large sample size, all estimates were significant.

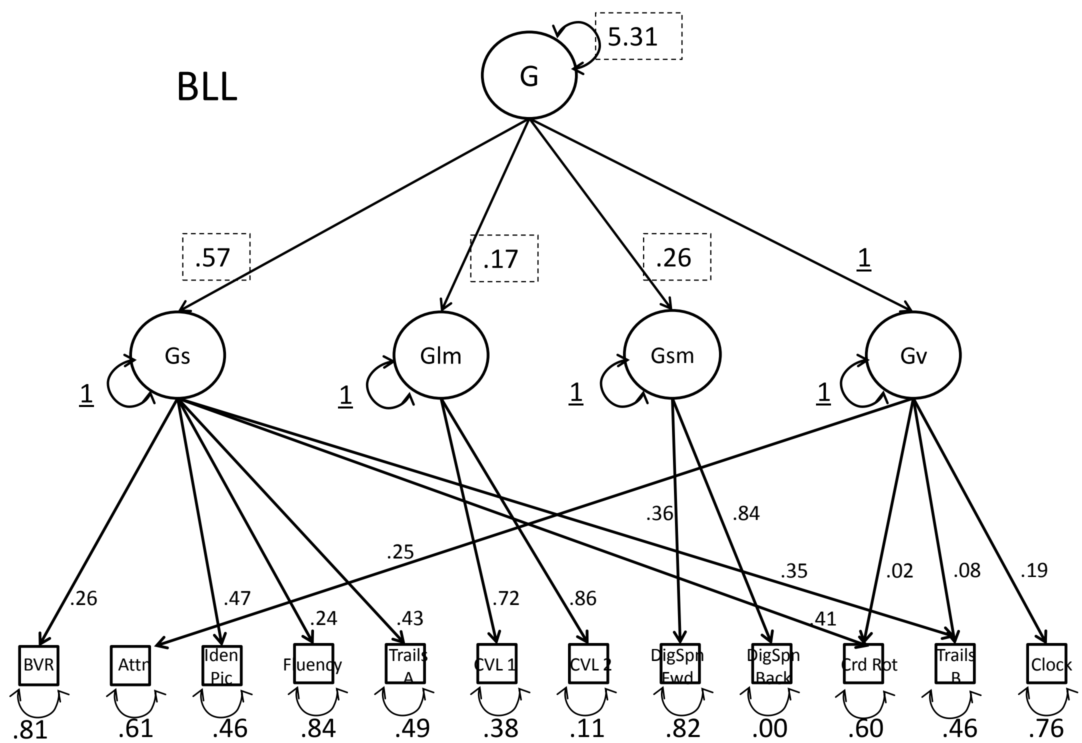


Figure B.2. Second-order model structure with final model estimates for ages 2 and 3. Underlined values were fixed to the displayed value; non-underlined items were estimated. Values in boxes were constrained to invariance between groups. Because of the large sample size, all estimates were significant.

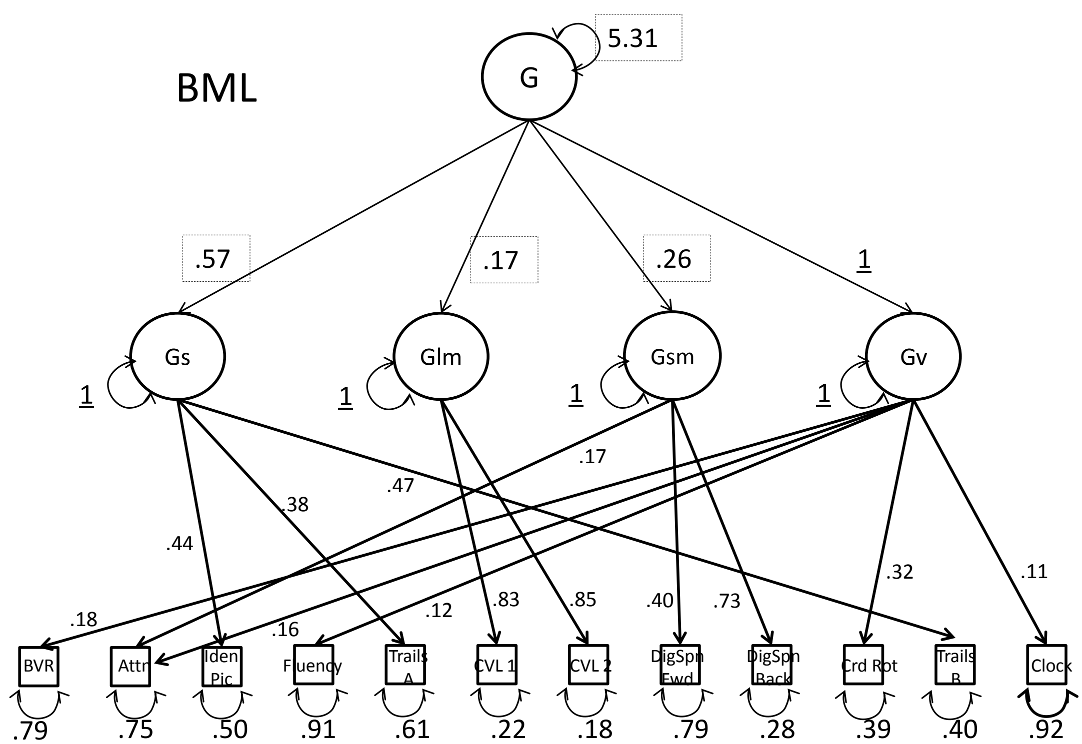


Figure B.3. Second-order model structure with final model estimates for ages 2 and 3. Underlined values were fixed to the displayed value; non-underlined items were estimated. Values in boxes were constrained to invariance between groups. Because of the large sample size, all estimates were significant.

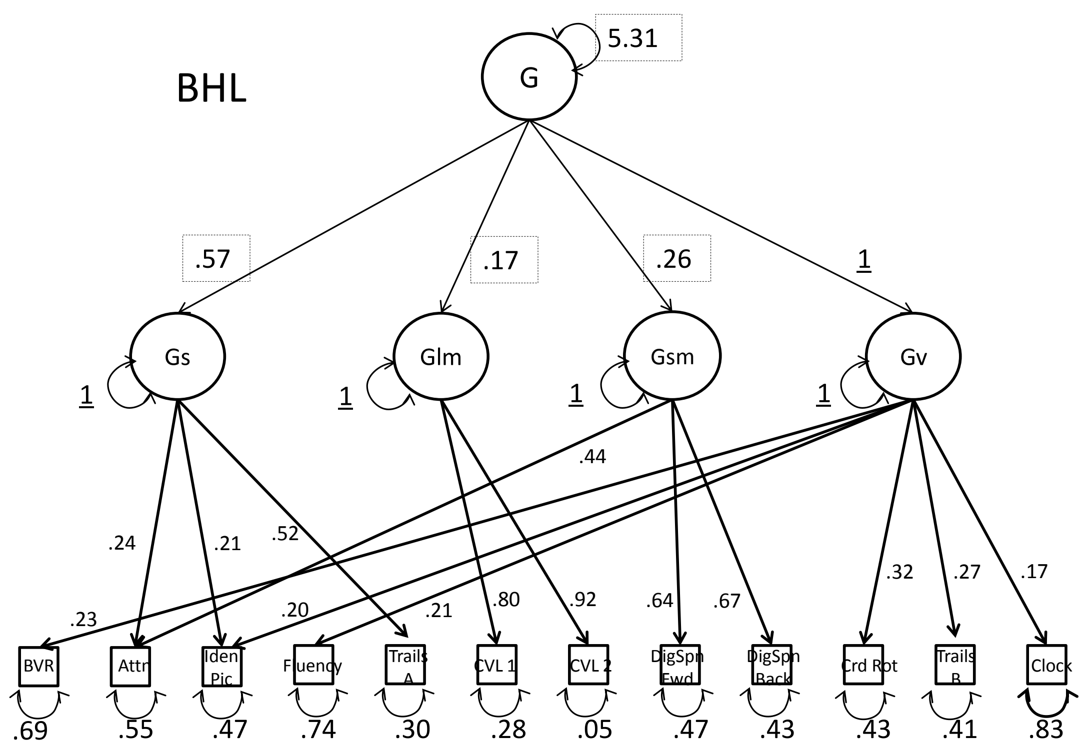


Figure B.4. Second-order model structure with final model estimates for ages 2 and 3. Underlined values were fixed to the displayed value; non-underlined items were estimated. Values in boxes were constrained to invariance between groups. Because of the large sample size, all estimates were significant.

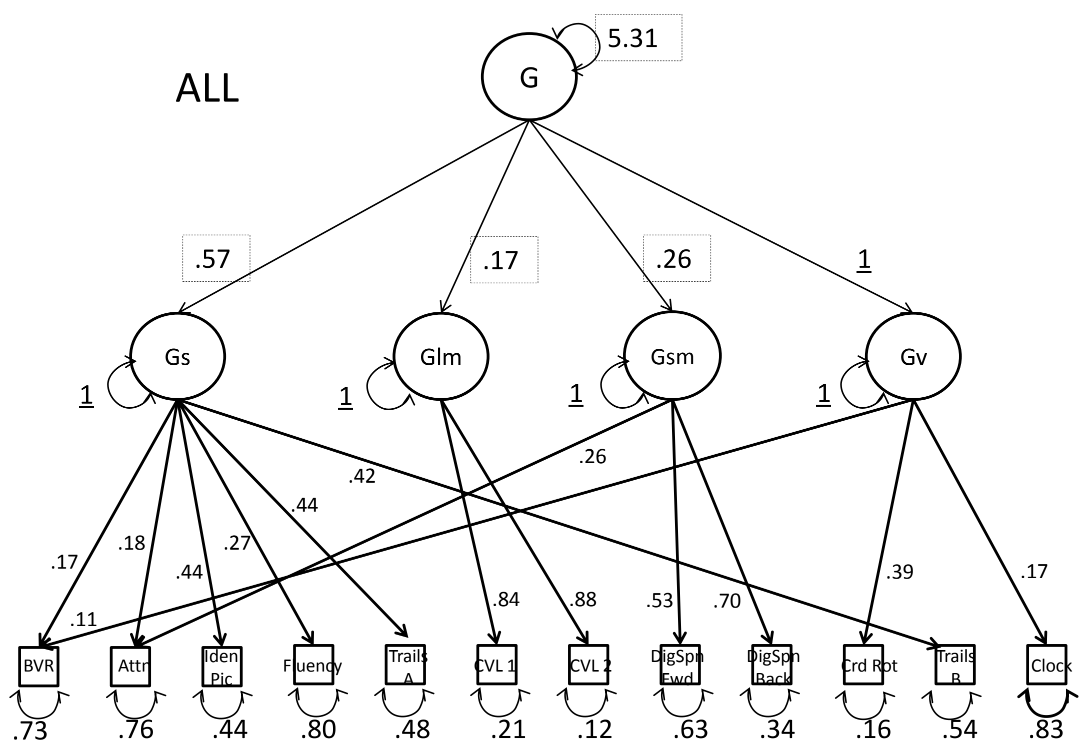


Figure B.5. Second-order model structure with final model estimates for ages 2 and 3. Underlined values were fixed to the displayed value; non-underlined items were estimated. Values in boxes were constrained to invariance between groups. Because of the large sample size, all estimates were significant.

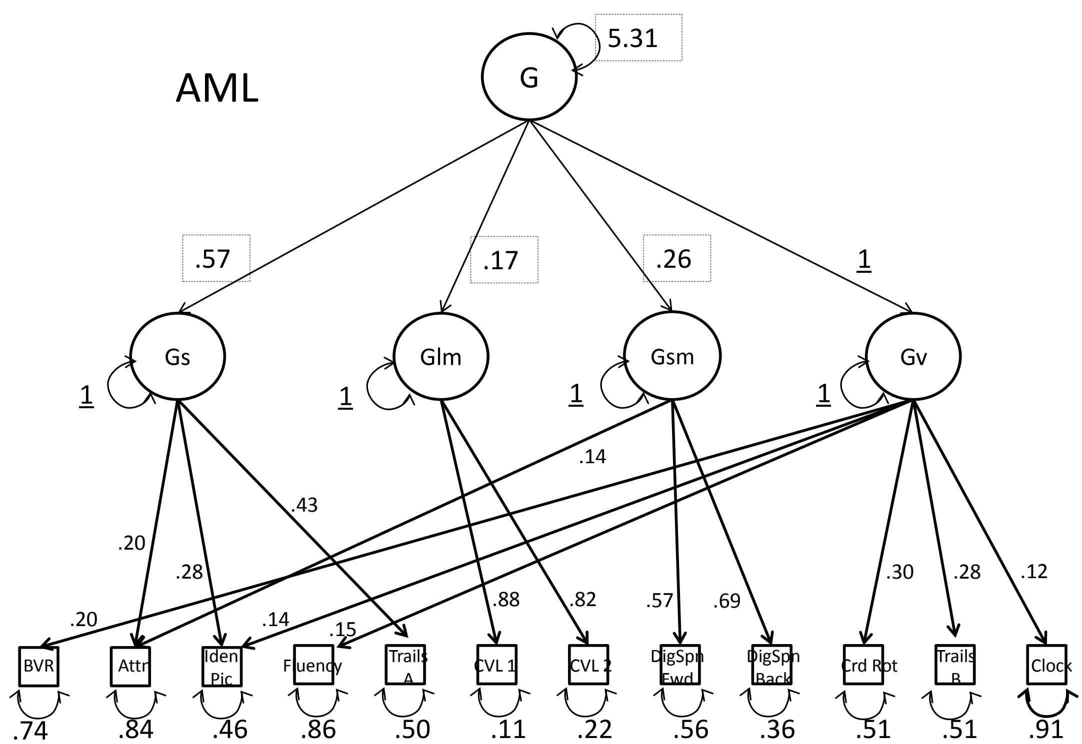


Figure B.6. Second-order model structure with final model estimates for ages 2 and 3. Underlined values were fixed to the displayed value; non-underlined items were estimated. Values in boxes were constrained to invariance between groups. Because of the large sample size, all estimates were significant.

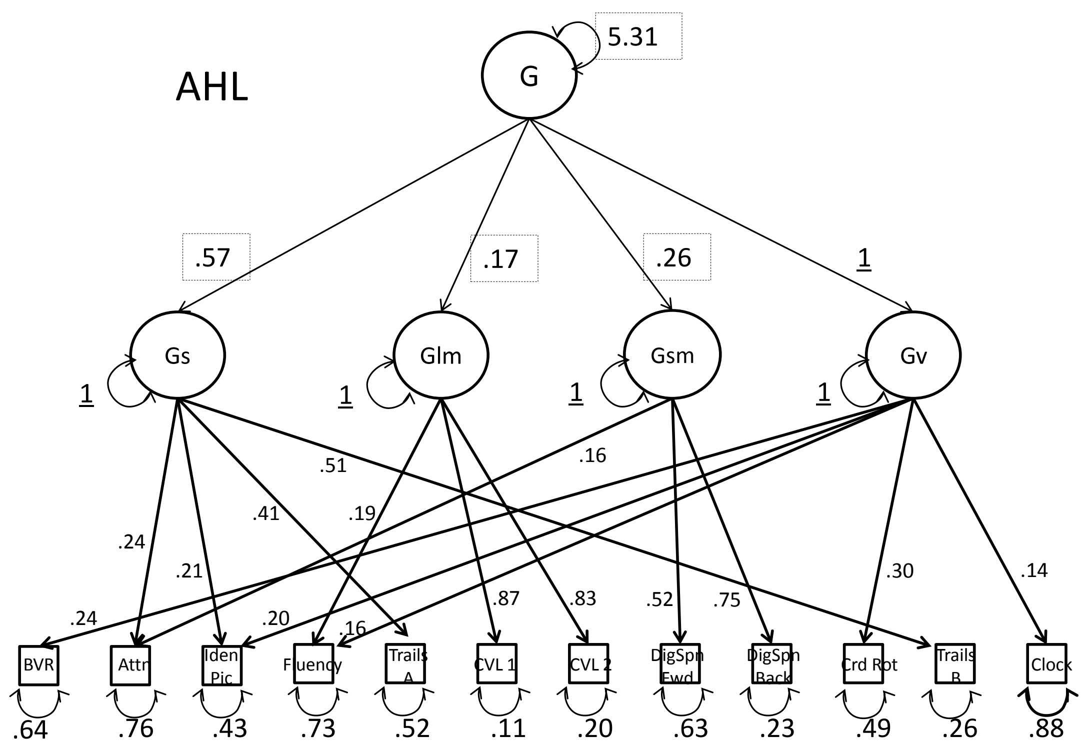


Table B.4. HANDLS model fit information: model comparisons of $-2LL, AIC, df$ and the χ^2 likelihood ratio p value

Comparison Model	Diff_{-2LL}	Diff_{AIC}	Diff_{df}	p
Base Model				
HND Null Model				
FO Free	398.27	-151.73	275	$< .001$
SO Free	439.13	-134.87	287	$< .001$
SO Constrained Ldngs & Var	478.03	-135.97	307	$< .001$
HND FO Free Model				
FO Constrained	47.74	-12.26	30	.02
HND SO Free Model				
SO Constrained Loadings	10.76	-1.86	15	.02
HND SO Constrained Loadings				
SO Constrained Ldngs & Var	10.76	0.76	5	.06

Table B.5. Early Steps model fit information: model comparisons of $-2LL, AIC, df$ and the χ^2 likelihood ratio p value

Base Model				
Comparison Model	Diff$_{-2LL}$	Diff$_{AIC}$	Diff$_{df}$	p
ES Null Model (multiple groups)				
FO Free	-111995.7		10442	≈ 1
ES FO Free Model				
FO Constrained	-6751.9	-14617.2	60	≈ 1
ES SO Dynamics Free				
SO Free	331.3	307.4	12	p ¡ .001
SO Constrained Dynamics	-4379.7	-4396	8	≈ 1
ES SO Constrained Dynamics				
SO Constrained Dynamics & Ldngs	374.7	-339	18	p ¡ .001

Bibliography

(2006). 2006 poverty guidelines for the 48 contiguous states and the District of Columbia.

Achenbach, T. M. and Rescorla, L. (2000). *ASEBA Preschool Forms & Profiles: An Integrated System of Multi-informant Assessment*. Aseba.

Achenbach, T. M. and Rescorla, L. (2001). *ASEBA School-Age Forms & Profiles*. Aseba, Burlington, VT.

Ahmavaara, Y. (1954). The mathematical theory of factorial invariance under selection. *Psychometrika*, 19(1):27–38.

Aitken, A. C. (1935). Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, 4(2):106–110.

Alwin, D. F. and Jackson, D. J. (1981). Applications of simultaneous factor analysis to issues of factorial invariance. In Jackson, D. J. and Borgatta, E. F., editors, *Factor analysis and measurement in sociological research: A multidimensional perspective*, pages 249–280. Sage, Beverly Hills, CA.

Bentler, P. M. and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3):588–606.

Berger, C. R. (1968). Sex differences related to self-esteem factor structure. *Journal of consulting and clinical psychology*, 32(4):442.

Beydoun, M. A. (2010). Dietary antioxidant intake and its association with cognitive function: socio-demographic differentials and the mediating role of depressive symptoms. Unpublished presentation.

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P., and Fox, J. (2012). OpenMx: Multipurpose software for statistical modeling, version 1.2. http://openmx.psyc.virginia.edu.

Bollen, K. and Lennox, R. (1991). Conventional wisdom on measurement: a structural equation perspective. *Psychological Bulletin*, 110(2):305–314.

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

Byrne, R. W. (2003). Imitation as behavior parsing. *Philosophical Transactions of the Royal Society of London, B*, 358(1431):529–536.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2):245–276.

Chen, F. F., Sousa, K. H., and West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(3):471–492.

Cheung, G. W. and Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of management*, 25(1):1–27.

Cronbach, L. J. and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4):281–302.

Darton, R. A. (1980). Rotation in factor analysis. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 29(3):pp. 167–194.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43(2):121–149.

Dishion, T. J., Shaw, D., Connell, A., Gardner, F., Weaver, C., and Wilson, M. (2008). The family check-up with high-risk indigent families: Preventing problem behavior by increasing parents positive behavior support in early childhood. *Child development*, 79(5):1395–1414.

Dodson, G. T., Boker, S. M., Evans, M. K., and Zonderman, A. B. (2014). An investigation of construct validity using higher order invariance.

Drapeau, A., Beaulieu-Prévost, D., Marchand, A., Boyer, R., Préville, M., and Kairouz, S. (2010). A life-course and time perspective on the construct validity of psychological distress in women and men. measurement invariance of the k6 across gender. *BMC medical research methodology*, 10(1):68.

Estabrook, R. (2012). Factorial invariance: Tools and concepts for strengthening research. *Measurement in sport and exercise psychology*, pages 53–63.

Evans, M. K., Lepkowski, J. M., Powe, N. R., LaVeist, T., Kuczmarski, M. F., and Zonderman, A. B. (2010). Healthy aging in neighborhoods of diversity across the life span (handls): A longitudinal, epidemiologic, urban study of health, race, and socioeconomic status. *Ethnicity & Disease*, 20:267–275.

Eysenck, H. J. and Eysenck, S. B. (1969). *Personality structure and measurement*. Routledge.

Gerbing, D. W. and Anderson, J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. *Multivariate Behavioral Research*, 20(3):255–271.

Golay, P. and Lecerf, T. (2011). Orthogonal higher order structure and confirmatory factor analysis of the french wechsler adult intelligence scale (wais-iii). *Psychological assessment*, 23(1):143.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? testing measurement invariance using the confirmatory factor analysis framework. *Medical care*, 44(11 Suppl 3):S78.

Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30:179–185. 10.1007/BF02289447.

Horn, J. L. and Blankson, N. (2005). Foundations for better understanding of cognitive abilities. *Contemporary intellectual assessment: Theories, tests, and issues*, 2:41–67.

Horn, J. L. and McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental aging research*, 18(3):117–144.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2):183–202.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, 57(2):239–251.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 57:409–426.

Lawley, D. (1944). Iv.a note on karl pearson's selection formulæ. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences*, 62(01):28–30.

Marsh, H. W. (1985). The structure of masculinity/femininity: An application of confirmatory factor analysis to higher-order factor structures and factorial invariance. *Multivariate Behavioral Research*, 20(4):427–449.

Marsh, H. W. and Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. *Psychological bulletin*, 97(3):562.

McDonald, R. P. (2011). Measuring latent quantities. *Psychometrika*, 76(4):511–536.

Meredith, W. (1964a). Notes on factorial invariance. *Psychometrika*, 29(2):177–185.

Meredith, W. (1964b). Rotation to achieve factorial invariance. *Psychometrika*, 29(2):187–206.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58:525–543.

Meredith, W. and Horn, J. (2001). The role of factorial invariance in modeling growth and change. In Collins, L. M. and Sayer, A. G., editors, *New methods for the analysis of change. Decade of behavior*, pages 203–240. American Psychological Association, Washington, DC.

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

Merkle, E. C. and Zeileis, A. (2013). Tests of measurement invariance without subgroups: a generalization of classical methods. *Psychometrika*, 78(1):59–82.

Millsap, R. E. and Kwok, O.-M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological methods*, 9(1):93.

Millsap, R. E. and Meredith, W. (2007). Factorial invariance: Historical perspectives and new problems. *Factor analysis at 100: historical developments and future directions*, 131.

Molenaar, P. C. M. and Nesselroade, J. R. (2012). Merging the idiographic filter with dynamic factor analysis to model process. *Applied Developmental Science*, 16(4):210–219.

Nesselroade, J. R. (2001). Intraindividual variability in development within and between individuals. *European Psychologist*, 6(3):187–193.

Nesselroade, J. R., Gerstorf, D., Hardy, S. A., and Ram, N. (2007). Idiographic filters for psychological constructs. *Measurement*, 5(4):217–235.

Nesselroade, J. R. and Molenaar, P. C. M. (2003). Quantitative models for developmental processes. In Valsiner, J. and Connolly, K. J., editors, *Handbook of Developmental Psychology*. Sage, London.

Norton, S., Cosco, T., Doyle, F., Done, J., and Sacker, A. (2013). The hospital anxiety and depression scale: A meta confirmatory factor analysis. *Journal of Psychosomatic Research*, 74(1):74 – 81.

Pearson, K. (1903). Mathematical contributions to the theory of evolution. xi.–on the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London*, A:1–66.

Pedraza, O. and Mungas, D. (2008). Measurement in cross-cultural neuropsychology. *Neuropsychology review*, 18(3):184–193.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Schlotz, W., Yim, I. S., Zoccola, P. M., Jansen, L., and Schulz, P. (2011). The perceived stress reactivity scale: Measurement invariance, stability, and validity in three countries. *Psychological Assessment*, 23(1):80–94.

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

Spearman, C. (1904a). ”general intelligence,” objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292.

Spearman, C. (1904b). The proof and measurement of association between two things. *American Journal of Psychology*, 15:72–101.

Steenkamp, J.-B. E. M. and Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1):pp. 78–107.

Thompson, B. and Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, 56(2):197–208.

Thomson, G. H. and Lederunn, W. (1939). The influence of multivariate selection on the factorial analysis of ability. *British Journal of Psychology. General Section*, 29(3):288–306.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, 38(5):406.

Thurstone, L. L. (1947). *Multiple Factor Analysis*. University of Chicago Press, Chicago.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the lifespan. *Developmental Psychology*, 45(4):1097–1118.

Vandenberg, R. J. and 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.

Widaman, K. F. and Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The science of pre-*

vention: Methodological advances from alcohol and substance abuse research, pages 281–324.

Zhang, Z., Browne, M. W., and Nesselroade, J. R. (2011). Higher-order factor invariance and idiographic mapping of constructs to observables. *Applied Developmental Science*, 15(4):186–200.