SMGroup

SM Journal of Biometrics & Biostatistics

Research Article

A Didactic Investigation of Perfect Fit in Second-Order Confirmatory Factor Analysis: Exploratory Structural Equation Modeling and Bayesian Approaches

Larry R Price

Department of Mathematics, Texas State University, USA

Article Information

Received date: Jul 05, 2017
Accepted date: Jul 24, 2017
Published date: Jul 27, 2017

***Corresponding author**

Larry R Price, Department of Mathematics, Texas State University, USA, Email: lprice@txstate.edu

Distributed under Creative Commons CC-BY 4.0

Keywords Confirmatory factor analysis; Bayesian structural equation modeling, exploratory structural equation modeling, second-order factor models, Wechsler Preschool and Primary Scale of Intelligence-IV

Abstract

Confirmatory Factor Analysis (CFA) plays an integral role in establishing evidence for the validity of test scores. This study provides a didactic strategy on the systematic investigation of perfect model-data fit in CFA. Specific steps presented include (a) investigating the impact of sample size on model fit indices, power and Type II error, (b) demonstration of how Muthén and Asparouhov's (2012) [1] ESEM approach is used to aid in evaluating the second-order factor model for simple structure, and (c) illustrating the tenability of Bayesian Structural Equation Modeling (BSEM) in resolving a non-positive definite matrix solution and in capturing the relationships between the measurement and latent variable parts of the second-order model in a way that provides an optimal tradeoff between simple structure and perfect model-data fit. The ESEM hierarchical approach identified loadings contradicting the original factor analytic results. Bayesian second-order CFA revealed that latent regressions were inflated in the original second-order CFA resulting in an in admissible solution due to a non-positive definite latent variable matrix. Respecification of the factor model using BSEM informed by the ESEM analysis eliminated the inadmissible solution and provided unbiased parameter estimates across sample sizes of N=100, 300, 600 and 1000.

Introduction

Confirmatory Factor Analysis (CFA) plays an integral role in establishing evidence for the validity of test scores. For example, CFA is used in test development is to verify that a factor structure based on empirical data aligns with theory. Consider the case where the goal of test development is to develop normative scores based on a factor analytic structure posited by theory. Provided the model fit is acceptable, scale scores developed based on a normative sample are subsequently used to make important decisions for diagnostic classification or ranking persons based on ability or achievement. However, CFA is highly restrictive due to the requirement of zero loadings on non-hypothesized factors for a particular model under investigation. The restrictive nature of CFA often leads to competing questions about alternative model structures that may yield better model-data fit and increased precision in parameter estimates. However, it is important to ensure that no statistical anomalies are present in the measurement and latent variable parts of the model in any factor analysis model. In empirical research, although perfect model fit in factor analysis is a target goal, the need for rigorous evaluation beyond a single analysis is critical. To this end, the need for close inspection and follow-up analyses of perfect model fit is clearly highlighted in this investigation.

The impetus for this investigation arose from the observation of perfect model fit in the hierarchical CFA conducted on the Wechsler Preschool and Primary Test of Intelligence-Fourth Edition(WPPSI-IV; Wechsler, 2012) [2] during the test development phase. In the original hierarchical CFA conducted on the WPPSI-IV, although perfect fit was obtained, a non-positive definite covariance matrix was observed in the latent variable (factor) matrix yielding the solution in admissible. As illustrated in this didactic, the restrictive nature of the hierarchical CFA generated inflated latent variable (factor) parameter estimates in the WPPSI-IV factor structure, which in turn, generated a non positive-definite latent variable (factor) matrix. Additionally, in the original hierarchical CFA, no supplemental analyses were conducted. For example, related questions include (a) How might the hierarchal CFA parameter estimates (and their standard errors) have differed based on a larger or smaller sample size as compared to the original normative sample? (b) Would the appearance of the non-positive definite latent variable matrix have appeared with different size conditions? (c) Would inflated latent variable correlations appeared at larger and smaller sample sizes other than the sample used in the original hierarchical CFA? This didactic investigation provides a systematic way to address these questions. Ultimately, the goal here is to provide guidance on how to identify the best hierarchical factor structure in a way that is supported by rigorous psychometric

OPEN ACCESS
ISSN: 2573-5470

and statistical procedures rather than rely on a single sample-based hierarchical factor analysis.

The goal of CFA is to obtain estimates for each parameter of the measurement model that produces predicted variance-covariance matrix ($\Sigma\hat{}$) that resembles the sample variance-covariance matrix (S) as closely as possible [3].The discrepancy between the two matrices is expressed as a fitting function F(S,$\Sigma\hat{}$). In the just-identified case the factor model holds exactly in the population (i.e., S=$\Sigma\hat{}$) and a perfect model-data fit is exhibited by (a) a $\chi 2$ of zero (0 df); p=1.0. In the over-identified case, F(S,$\Sigma\hat{}$)=0 and the criteria for a perfect model-data fit includes: (a) $\chi 2$ goodness-of-fitindex is not statistically significant with a range between p=.50-1.0 (Jöreskog & Sörbom,1996a) [4], (b)a value=1.0 for a goodness-of-fit index such as the comparative fit index (CFI), (c) the root mean square error of approximation (RMSEA)=0.0 and (d)a residual matrix of the discrepancy between S and $\Sigma\hat{}$containing 0.0 for all parameter estimates. In CFA, perfect model-data fit is rarely obtained and as Bollen (1989) [5] notes, even when a perfect model-data fit is obtained, there are other plausible models where perfect fit is possible. As previously noted, CFA is restrictive requiring that subtests load exclusively on specific theoretical factors. Under this assumption, CFA is consistent with the restrictive independent cluster model and it has the advantage of motivating researchers to develop parsimonious models. Although the idea of having pure items or sub tests that load on a single factor is a central tenet of CFA, this is not are quirement of a well-defined, useful factor structure, nor even a requirement of traditional definitions of simple structure (Brown, 2006; McDonald, 1999; Thurstone, 1947) [6-8]. The concept of simple structure was introduced by Thurstone (1947) [8] where he argued that a well-defined, useful factor structure should meet certain criteria. The principles of simple structure are more comprehensive than the exclusive requirement that only specific items or subtests load on specific factors (with no cross-loadings). The principles of simple structure include (a) each row of the factor matrix having at least one loading close to zero, (b) for each column of the factor matrix there should be as many variables with zero loadings or near-zero loadings as there are factors, (c) for every pair of factors there should be several variables with loadings on one factor but not in the other, (d) when there are four or more factors, a large portion of the variable should have negligible (close to zero) loadings on any pair of factors, and (e) for every pair of factors in the factor matrix there should be only a small number of variables with non-zero loadings (Kerlinger and Lee, 2000) [9].

The aim of this didactic investigation is to present a systematic approach to ensure the criteria for simple structure are met while simultaneously evaluating the psychometric integrity of the factor solution in the presence of perfect model fit. Specially, the evaluation of psychometric integrity in this investigation includes examining the existence of bias in factor loadings over variations in sample size, presence or absence of singularity in the variance-covariance latent variable matrix and adherence to the guidelines for simple structure. To facilitate presentation, a three-step procedure is provided using the standardization sample from the young sample (ages 2years, 6months to 3years, 11 months, N=600) Wechsler Preschool and Primary Test of Intelligence-Fourth Edition (WPPSI-IV; Wechsler, 2012) [2]. The original CFA conducted on the WPPSI-IV produced a perfect model fit; however, singularity existed in the solution (i.e. a non-positive definite covariance matrix was observed in the factor solution).

Study Goals

A search of published research examining methodological strategies to evaluate the condition of perfect model-data fit in second-order CFA yielded zero citations. To address this gap in the literature, the goal of this study is to present a didactic on how to proceed in the systematic investigation of perfect model data fit. Specifically, detailed steps are provided related to how researchers can proceed with the goal of identifying the best factor structure supported by rigorous psychometric and statistical procedures. The presentation is based on empirical data used in the original hierarchical CFA conducted during the standardization phase of the WPPSI-IV. A factor analytic study such as the original CFA is routinely conducted by the test publisher prior to the development and publication of normative scores. Step one of the procedure involved evaluating the impact of sample size on parameter estimation bias and singularity in the factor solution (i.e. observance of a non-positive covariance matrix among the structural part of the model). Step two includes using (Muthén and Asparouhov's (2012) [1]. ESEM approach to aid in evaluating the second-order factor model for simple structure and diagnosing problems leading to observance of a non-positive definite latent variable matrix (i.e. singularity). Step three uses the information gleaned from the ESEM analysis to inform a Bayesian Structural Equation Modeling (BSEM) in resolving the non-positive definite matrix solution and capturing the relationships between the measurement and latent variable parts of the second-order model in a way that provides an optimal tradeoff between simple structure and perfect model-data fit. The methodological approach provided here offers a didactic, systematic strategy for investigating the occurrence of perfect model-data fit in second-order factor analytic studies. Ideally, the strategy provided will serve as a guide for theoretical and applied researchers in improving the sensitivity and precision of CFA.

In this didactic example, perfect model fit was evaluated relative to the theoretical framework of the Wechsler Preschool and Primary Test of Intelligence-Fourth Edition (WPPSI-IV; Wechsler, 2012) [2]. The second-order (hierarchical) factor structure and subtest composition of the WPPSI-IV illustrated in Figure 1 reflects current theory and practice of intelligence assessment in children (Alfonso, et al., 2005) [10]. The WPPSI-IV is an innovative measure of cognitive ability for preschool age and young children that is grounded in contemporary theory and research. The six subtests included on the WPPSI-IV specific to the young sample are (a) Information, (b) Receptive Vocabulary, (c) Block design, (d) Object Assembly, (e) Picture Memory, and (f) Zoo Locations. The Information and Receptive Vocabulary subtests comprise the first-order Verbal Comprehension factor; Block Design and Object Assembly comprise the first-order Visual Spatial factor, and Picture Memory and Zoo Locations comprise the first-order Working Memory factor (see Figure 1). The Information subtest measures a child's ability to acquire, retain and retrieve general factual knowledge. It involves crystallized intelligence, long-term memory, and the ability to retain and retrieve knowledge from the school environment. The Receptive Vocabulary subtest measures a child's ability to comprehend verbal directions, auditory and visual discrimination, auditory memory, auditory processing and the integration of visual perception and auditory input. The Block Design subtest is designed to measure the ability to analyze and synthesize abstract visual stimuli. The test also includes nonverbal concept formation visual perception and

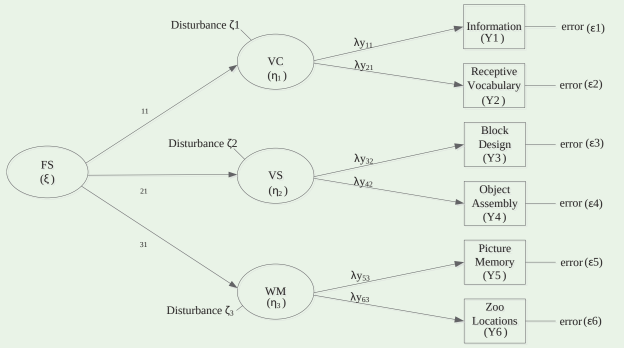

Figure 1: WPPSI-IV Hierarchal Confirmatory Factor Analysis Model.

organization, simultaneous processing, visual-motor coordination, earning and the ability to discern figure in visual stimuli. The Object Assembly subtest is designed to assess visual-perceptual organization, integration and synthesis of part-whole relationships, nonverbal reasoning, cognitive flexibility, and visual-motor-coordination. The Picture Memory subtest assesses working memory in the form visual perception and organization, concentration, and visual recognition and memory of important details. The Zoo Locations subtest assesses visual-spatial working memory-based skills (Figure1).

Sample

A subsample of the standardization (normative) sample for the Wechsler Preschool and Primary Test of Intelligence-Fourth Edition (WPPSI-IV; Wechsler, 2012) [2] ages 2 years, 6 months to 3 years, 11 months (young sample, N=600) served as the sample for this investigation. Four out of six sub tests displayed mild or moderate skewness; multivariate kurtosis was acceptable. Table 1 provides descriptive statistics and subtest score reliability estimates for the sample.

The second-order confirmatory factor analysis (a.k.a. HCFA) is provided in figure1. The second-order CFA model is expressed in notation LISREL (Jöreskog & Sörbom, 1996a) [4] in Equation1. Table 2 provides the maximum likelihood parameter estimates for the original second-order CFA. The second-order CFA is over-identified and yielded a perfect model-data fit (i.e.$F(S,\hat{\Sigma}) = 0$, $\chi 2=2.0(6)$, $p=.91$; CFI =1.0; RMSEA=0.0; SRMR=0.0).

Table 1: Descriptive statistics for WPPSI-IV subtests.

Subtest	*M*	*SD*	Skewness (Z-score)	Kurtosis (Z-score)	Reliability
Information	9.90	3.0	-2.58*	1.54	.91
Receptive Vocabulary	9.90	3.0	-0.06	0.32	.91
Block Design	9.91	3.0	-4.03*	5.37	.85
Object Assembly	9.70	3.1	2.19*	1.26	.85
Picture Memory	9.90	3.2	-4.45*	2.10	.91
Zoo Locations	9.93	3.0	-1.19	1.87	.90
Multivariate				3.29	

Note: Subtests scores are based on total scores for item responses then normalized into scale scores.
*skewness significant at p < .05

Equation1.Second-Order Factor Analysis Model

$\mathbf{y} = \Lambda_y(\Gamma\xi + \varsigma) + \varepsilon$

Λ_y = matrix of first-order factor loadings (p x m).

Γ = matrix of second-order factor loadings (m x n).

ξ = vector of second-order factors (n x 1).

ς =vector of second-order unique components.

ε = vector of first-order unique components.

Φ =covariance matrix of second-order factors (n x n).

Ψ =covariance matrix of second-order uniqueness(m x m).

Φ_ε =diagonal covariance matrix of first-order uniqueness (p x p).

Methods

Number of Factors per Items/Sub tests and Sample Size

The SEM literature on the use of maximum likelihood estimation on model parameters, regarding the percentage of proper solutions, accuracy of parameter estimates and the appropriateness of the overall chi-square reveals that large sample sizes (N) are required for unbiased parameter estimates. For example, Anderson and Gerbing (1988) [11] recommend an N of between 100 to 150 subjects,

Table 2: WPPSI-IV second-order (hierarchical) Confirmatory Factor Analysis.

Factor by Subtest	Loading	S.E.	Confidence Interval	
Verbal Comprehension			Lower95%	Upper95%
Information	0.76	0.03	0.71	0.82
Receptive				
Vocabulary	0.75	0.03	0.69	0.79
Visual Spatial				
Block Design	0.66	0.04	0.59	0.73
ObjectAssembly	0.61	0.04	0.53	0.68
Working Memory				
PictureMemory	0.73	0.05	0.64	0.82
Zoo Locations	0.5	0.04	0.42	0.58

Note: N=600. Loadings are standardized. Confidence intervals based on 1000 bootstrap replications.

Composite		Full Scale Factor		Confidence	Interval
				Lower 95%	Upper 95%
Verbal Comprehension	0.92(.05) ((.0(.03)			0.85	0.909
Visual Spatial	-	0.92 (.05)		0.84	1
Working Memory	-	-	0.90 (.05)	0.83	1

Note: N=600. Loadings are standardized. Confidence intervals based on 1000 bootstrap replications. Standard errors are in parentheses in lower composite table.

Boomsma (1983) [12] recommends N=400. Bentler and Chou (1987) [13] recommend a ratio also was five subjects per variable as being sufficient in normal and elliptical distributions when the latent variables have multiple indicators. MacCallum, Browne and Sugawara (1996) [14] provide support for large N noting that the power and precision of parameter estimates increase monotonically with sample size andthedegreesoffreedom. However, they also noted that adequate power can be achieved with relatively modest levels of N when degrees of freedom are small. Ding, Velicer and Harlow (1995) [15] reported that the likelihood of fully proper solutions increased with increasing the number of indicators per factor, sample size and magnitude of factor loadings. Regarding the optimal number of indicators (*p*) per factor(*f*; i.e.#*p/f*), research suggests thatat least three indicators per factor are desirable, but under certain circumstance two may be sufficient (Brown, 2006, p.72; Bollen, 1989) [5,5]. Circumstances where two indicators per factor are acceptable include (a) when every latent variable is correlated with at least one other latent variable and (b) the errors between indicators are uncorrelated. The WPPSI-IV second-order factor structure includes two subtests per factor with factors being correlated with one another and the assumption that errors of measurement are uncorrelated. In step one of this didactic, we investigate the role of sample size in (a) model-data fit, (b) power, (c) Type II error and (d) observing a non-positive covariance matrix.

Sample Size

A statistically significant χ2of model-fit indicates a discrepancy between the sample variance-covariance matrix (**S**) and the predicted variance-covariance matrix (**Σ**ˆ.) However, the χ2 is problematic due to its sensitivity to sample size, yielding solutions that are rejected when using large sample size seven when differences between **Σ**ˆ and **S** are negligible (Yuan, 2010) [16]. Additionally, the role of sample size was of interest in relation to observing the presence of a non-positive definite latent variable covariance matrix. In step one of our approach, Markov chain Monte Carlo (MCMC, Brooks et.al, 2011) [17] simulation was used to examine the sampling distribution of the parameter estimates and error structure of the WPPSI-IV second-order factor model across sample sizes of N=100, 300, 600 and 1,000. In the simulation study, parameter estimates for the N=600 (standardization sample for the WPPSI-IV) served as the starting values for all simulation conditions. Additionally, a power analysis at each sample size condition was conducted. M*plus*version7.3 (Muthén & Muthén, 2012) [18] was used to conduct the simulation study. Percent bias was defined as the population parameter value minus the average parameter value divided population parameter value times one-hundred (100). Ninety-five percent (95%) coverage was defined as the proportion of replications for which the 95% confidence interval contained the true population parameter value. Details describing the steps in conducting a simulation study such as the one in this study are provided in Muthén & Muthén, 2002; Price, 2012) [19,20].

Results from the simulation study are reported in Table3and reveal that perfect model fit for the over-identified second-order model resolved (i.e. was no longer observed) at sample sizes of 100, 300 and 600. Particularly problematic was the observation that at a sample size of N=100,59% of the solutions were inadmissible due to a non-positive definite latent variable matrix. Across all sample sizes, bias in factor loadings remained below 5% (Table 3). However, at sample size N=300, 35% of the solutions were in admissible due to a non-positive definite latent variable matrix. At sample size N=600, 7% of the solutions were in admissible due to a non-positive definite latent variable matrix. In summary, the results of the simulation study revealed that while parameter estimates (factor loadings) were unbiased across all sample sizes, singularity in the second-order latent variable matrix remained problematic. In summary, sample size played an important role in diagnosing perfect model fit however it did not provide a diagnostic solution specific to singularity in the latent variable covariance matrix (Table 3).

Exploratory Structural Equation Modeling and Factor Analysis

One goal in this investigation was to illustrate the use of ESEM as a viable confirmatory alternative to CFA on the basis of strong theoretical assumptions regarding the expected factor structure. Therefore, in step two of this didactic, Asparouhov and Muthén's (2009) [21] exploratory structural equation modeling (ESEM) framework was employed to address perfect fit and resulting improper solutions manifested as singularity in the second-order latent variable covariance matrix. The ESEM strategy provides a bridge between traditional EFA and CFA in that it (a) provide a statistical test of loadings in the measurement model and (b) allows for identification of potentially important cross-loadings and (c) captures measurement error in the respective sub tests. In traditional EFAthe decision to retain or exclude items or subtests involves some degree of subjectivity (Fabrigar & Wegener, 2012; Mulaik, 1972) [22,23]. In comparison to CFA, ESEM provides a way to apply EFA in a rigorous manner yielding a more accurate picture of the underlying factor structure relative to the constructs posited by theory. For example, allowing small cross-loadings to be included in a model provides a mechanism to identify items that are imperfect indicators of a construct. This step is beneficial because some degree of irrelevant association with the other constructs is able to be included in the model. Such irrelevant association in a model is identified as a form of systematic measurement error. Critical to the present investigation, is that when cross-loadings, even small ones, are not estimated, the only way to represent relationships between specific indicators and other constructs is through the latent factor correlations (Asparouhov & Muthén, 2009) [21]. Under these circumstances, latent factor correlations end up being overestimated in many applications of CFA - an artifact observed in the present study. In addition to inflated latent factor correlations, in hierarchical factor analysis, the issue of singularity (a.k.a. a non positive-definite latent variable variance-covariance matrix) is possible to exist. In the present study, singularity in the hierarchical latent variable variance-covariance matrix was particularly problematic because of the high correlation among Verbal Comprehension, Visual Spatial, and Working Memory factors at level-1 of the hierarchy. Importantly, ESEM offers the same advantages as CFA analysis in terms of fit indexes, standard errors, and tests of significance. Additionally, the ESEM framework is highly flexible and allows one to model correlated residuals and conduct tests of measurement invariance. In this way, ESEM provides a unified approach between CFA, EFA and SEM.

The main difference between CFA and ESEM/EFA is that all cross-loadings are estimated in EFA/ESEM and not in CFA. Model fitting with ESEM proceeded by allowing all six subtests (Figure 2) to load on each of the three factors (a) Verbal Comprehension, (b) Visual Spatial, and (c)Working Memory. Importantly, if there are at least moderate cross-loadings (e.g.,≥.30) in the true population and

Table 3: Monte Carlo Results for Hierarchical Confirmatory Factor Analysis of WPPSI-IV.

	N = 100					N = 300				
Factor by Subtest	**Population**	**Average**	**% Bias**	**95% Coverage**	**Power**	**Population**	**Average**	**% Bias**	**95% Coverage**	**Power**
Verbal Compression										
Information	1.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00
Receptive Vocabulary	0.97	0.99	2.06	0.93*	1.00	0.98	0.98	0.00	0.95	1.00
Visual Spatial										
Block Design	1.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00
Object Assembly	0.94	0.96	2.12	0.94*	0.99	0.94	0.95	1.06	0.96	1.00
Working Memory										
Picture Memory	1.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00
Zoo Locations	0.66	0.67	1.51	0.96	0.97	0.66	0.66	0.00	0.95	1.00
Full Scale										
Verbal Compression	1.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00
Visual Spatial	0.85	0.87	2.35	0.95	1.00	0.85	0.86	1.17	0.94*	1.00
Working Memory	0.97	0.99	2.06	0.96	1.00	0.97	0.98	1.03	0.96	1.00

Note. Unstandardized estimates. MCMC iterations = 1000; Chi-Square = 6.08/SD = 3.5; RMSEA = 0.03/SD = .04; SRMR = .03/SD =.009. *Coverage < 95%. 59% solutions inadmissible due to non-positive-definite latent variable (factor) matrix. This may occur because of (a) a correlation greater or equal to 1 between 2 latent variables, (b) or a linear dependency among more than 2 latent variables, or (c) a negative variance for a latent variable. % bias = population parameter value - the average parameter value/population parameter value * 100. 95% coverage = the proportion of replications for which the 95% confidence interval contains the true population parameter value.

	N = 600					N = 1000				
Factor by Subtest	**Population**	**Average**	**% Bias**	**95% Coverage**	**Power**	**Population**	**Average**	**% Bias**	**95% Coverage**	**Power**
Verbal Comprehension										
Information	1.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00
Receptive Vocabulary	0.97	0.98	1.03	0.95	1.00	0.97	0.97	0.00	0.95	1.00
Visual Spatial										
Block Design	1.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00
Object Assembly	0.93	0.94	1.08	0.95	1.00	0.94	0.94	0.00	0.96	1.00
Working Memory										
Picture Memory	1.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00
Zoo Locations	0.65	0.66	1.53	0.94*	1.00	0.66	0.66	0.00	0.95	1.00
Full Scale										
Verbal Comprehension	1.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00
Visual Spatial	0.84	0.85	1.20	0.95	1.00	0.85	0.86	1.17	0.94*	1.00
Working Memory	0.97	0.97	0.00	0.96	1.00	0.97	0.98	1.03	0.96	1.00

Note. Unstandardized estimates. MCMC iterations = 1000; Chi-Square = 6.2/SD = 3.5; RMSEA = 0.01/SD = .01; SRMR = .01/SD =.003. *Coverage < 95%. 7% solutions inadmissible due to non-positive-definite latent variable (factor) matrix. This may occur because of (a) a correlation greater or equal to 1 between 2 latent variables, (b) or a linear dependency among more than 2 latent variables, or (c) a negative variance for a latent variable. % bias = population parameter value - the average parameter value/population parameter value * 100. 95% coverage = the proportion of replications for which the 95% confidence interval contains the true population parameter value.

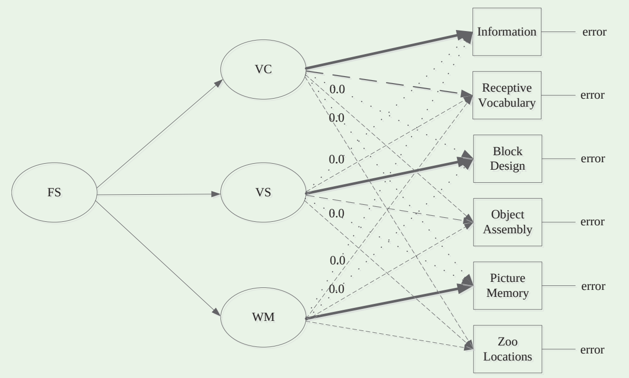

Figure 2: Hierarchical Exploratory SEM Analysis (HESEM).
Note: Heavy bold paths=anchor test; Dotted paths are set to zero because they serve as anchors in the simultaneous solution of other composities; Dashed paths are required for the exploratory solution. VC= verbal compherension, VS= visual spatial, WM=working memory.

these loadings are constrained to be zero, then estimated regression weights in the second-order portion of the model are likely to be inflated (Asparouhov & Muthén, 2009; Marsh, 2010) [21,24]. The occurrence of inflated factor regressions with factor models having moderate cross-loadings holds true for orthogonal and oblique factor models alike. In table 4, the results of the exploratory/confirmatory approach using ESEM approach are provided for the WPPSI-IV.

The results of the ESEM revealed additional loadings on non-hypothesized factors in comparison to the original CFA model. Although the observed non-hypothesized loadings were not statistically significant, they verified that (a) the assumptions of the CFA model were untenable for the N = 600 standardization sample and (b) violations of the tenets of simple structure were revealed. Additionally, statistical significance was not observed on all subtests hypothesized to load on their respective factors. For example, for factor one (Verbal Comprehension), the Receptive Vocabulary subtest displayed an adequate loading, but was not statistically significant. For factor two (Visual Spatial), the Object Assembly subtest displayed an adequate loading, but was not statistically significant. Latent factor variances for VC, VS, WM were set to 1.0 in the ESEM to overcome zero degrees of freedom and non-positive definite covariance matrix problems in the second-order latent variable matrix. Although the ESEM analysis provided greater diagnostic utility by revealing problematic portions of the WPPSI-IV factor structure, the final solution provided was not optimal and did not meet our target criteria for simple structure. In an attempt to further improve on model refinement, the second-order factor model was investigated using Bayesian SEM in step three of the procedure.

Bayesian structural equation modeling

In step three, Bayesian SEM was employed to further examine the identified problems in the second-order or hierarchical factor model. The goal is step three was to leverage Bayesian SEM to specify a model that (a) resolved the singularity problems in the latent factor covariance matrix and (b) met the criteria for simple structure and (c) exhibited unbiased parameter estimates across samples sizes of 100, 300, 600 and 1,000. Bayesian statistical thinking can be viewed as an extension of the traditional (i.e., frequentist) approach, in that it formalizes aspects of the statistical analysis that are left to uninformed judgment by researchers in classical statistical analyses (Hoff, 2009; Lee, 2007) [25]. A difference between Bayesian and Classical probability is that in the Bayesian framework, the data are fixed and the parameters are random. Conversely, in classical probability, the data are random and the estimated parameters are fixed. Because the parameters estimated in a Bayesian analysis are random, researchers are able to make direct probabilistic statements about (a) point estimates such as the mean, median, mode, or regression weights and (b) random parameters lying within specific upper and lower limits (i.e. credibility intervals). Also, Bayesian credible intervals often contain frequent is to coverage probabilities, often being close to Bayesian coverage levels (Hoff, 2009). A particular strength of the Bayesian approach to factor analysis is that one can refine the model at critical parts (i.e. through use of assigning priors to parameters) in order to gain a clearer picture of the underlying structure and use MCMC re sampling to examine the performance of the model.

Table 4: Exploratory Second-Order SEM of WPPSI-IV young sample.

Subtest	**Factor**	***p*-value**	**Factor**	***p*-value**	**Factor**	***p*-value**	**Reliability**
	1		2		3		
Block Design	0.00 (.00)	-	0.68 (.09)	0.00	-0.03 (.08)	0.69	.85
Object Assembly	0.03 (.28)	0.92	0.67 (.62)	0.28	-0.10 (.43)	0.80	.85
Information	0.83 (.12)	0.00	0.00 (.00)	-	0.00 (.00)	-	.91
Receptive Vocabulary	0.49 (.35)	0.16	0.26 (.34)	0.47	0.00 (.00)	-	.91
Picture Memory	0.19 (.18)	0.29	-0.15 (.18)	0.40	0.70 (.06)	0.00	.91
Zoo Locations	0.00 (.00)	-	0.00 (.00)	-	0.52 (.04)	0.00	.90
Composite	Full Scale Factor					*p*-value	
Verbal Comprehension	0.80					0.000	.94
Visual Spatial	-	-	0.97			0.000	.90
Working Memory	-	-	-	-	0.90	0.000	.92

Note. Estimates are standardized; Sample size is N=600; Ages 2:6 to 3:11. Values in parentheses are standard errors. Reliability estimates are based on split-half technique with Spearman-Brown correction. "-" results from a zero factor loading. Verbal latent variable r-square = .63; Visual Spatial latent variable r-square = .94; Working Memory latent variable r-square is undefined because of a .82 loading. *Latent factor variances VC, VS, WM set to 1.0; this step taken to overcome zero degrees of freedom and non-positive definite covariance matrix problems in the PSI or latent variable matrix.*

Assigning Priors to Model Parameters: Bayesian statistical inference provides a frame work for incorporating information from on level via priors to inform the analysis at other levels (Congdon, 2010). [26]. Assigning priors increases precision of the parameter estimates and is advantageous in complex models. Here, informative priors (~Normal 0.0, .04) were used on factor cross-loadings reflecting a high probability with in a plausible range for the parameter values. Primary factor loadings were set at (~N 0,1); informative inverse gamma priors (~IG0.0, .001) were placed on the intercept and variances of the latent variables (factors).

Bayesian Model Fit and Results: All analyses were performed using *M p*lus, version 7.3. The posterior distributions for all parameters were derived through the Gibbs sampling method. In this method, the conditional distribution of one set of parameters given other sets can be used to maker and omdraws of parameter values, resulting in the approximation of the joint distribution of all parameters (Muthén and Asparouhov, 2012) [1]. In the N=600sample, the PPP was observed as .65 indicating an excellent model-data fit. The Gel man-Rubin convergence diagnostic takes into account the Potential Scale Reduction statistic (PSR; Gelman, Carlin, Sternand Rubin, 2004) [27] and monitors the between chain variation to the within-chain variation. Proper convergence of the MCMC chains over the 50,000 draws was achieved (i.e. observed as 1.02), ideal values are between1 and 1.1. Thus convergence was verified by PSR value of <1.1. Results of the Bayesian second-order factor model demonstrate (a) that the WPPSI-IV is properly specified and (b) improvements in accuracy and precision of the parameter estimates were achieved while simultaneously accounting for challenges in the second-order latent variable (i.e. factor) matrix. Table 5 provides the results of the Bayesian Second-Order CFA.

The results of the Bayesian second-order factor analysis provided a solution to the issues observed in the original second-order CFA and those appearing in the ESEM approach. For example, the second-order factor structure of the WPPSI-IV displayed (a) significant loadings on the hypothesized factors, (b) small, non-significant loadings on non-hypothesized factors (<.06; Table 5), consistency among the second-order factors (e.g., loadings of .89 or .90) relative to their loadings on the subtests, and (d) elimination of the non-positive definite latent variable issue (see footnote in Table 5).

Conclusion and Discussion

The findings from this study provide important did actic information for researchers under circumstances where perfect model-data fit is observed in second-order or hierarchical CFA. The results of the simulation study revealed that the parameter estimates reported in the original second-order CFA were unbiased with power being adequate at sample sizes as small as N=100 (although for some loadings 95% coverage was not observed). However, inadmissible solutions due to a non-positive definite latent variable matrix were observed >10% of the time until the sample size reached N>600. Based on the results here and the accessibility of conducting simulation protocols, researchers in applied psychometrics are encouraged to conduct simulation studies as a regular part of their factor analytic work to provide more rigorous evaluation of the role of sample size, power and parameter bias.

The ESEM approach confirmed the initial problems observed in the WPPSI-IV second-order CFA. For example, small to moderate cross-loadings were identified violating the tenets of simple structure and non-significant loadings were observed on some of the hypothesized subtest-factor relationships. Also, the non-positive definite latent variable matrix remained problematic in the ESEM analysis. Taken together, these results provide evidence that while the ESEM is more sensitive than traditional CFA at revealing the second-order factor structure, it did not provide information enough useful for resolving the problem of a non-positive latent variable matrix and achieving simple structure.

In the final step of this didactic, analytic results of the Bayesian SEM analysis revealed that latent (factor) regressions were inflated in the original second-order CFA. This result is highly important since, from a theoretical perspective, it is important that the size of the relationships between the hierarchical factors in the model are accurately captured. The ability to include prior information into

Table 5: Bayesian Second-Order SEM of WPPSI-IV Young Standardization Sample N=600. (Bayes estimation with small variance priors on cross-loadings of .04).

Subtest	Factor	*p*-value	Factor	*p*-value	Factor	*p*-value	
	1		2		3		
Block Design	0.03	0.40	0.61	0.00	0.04	0.37	
Object Assembly	0.05	0.32	0.55	0.00	0.02	0.43	
Information	0.71	0.00	0.02	0.43	0.05	0.33	
Receptive Vocabulary	0.71	0.00	0.04	0.34	0.01	0.45	
Picture Memory	0.05	0.33	0.01	0.47	0.67	0.00	
Zoo Locations	-0.01	0.45	0.04	0.35	0.49	0.00	
Composite	Full Scale Factor					*p*-value	
Verbal Comprehension	0.89					0.000	
Visual Spatial	-	-	0.89			0.000	
Working Memory	-	-	-	-	0.90	0.000	

Note. MCMC based on 200,000 iterations with 2 chains using Gibbs sampling. PPP=.64; PSR <1.05 Verbal latent variable r-square = .79; Visual Spatial latent variable r-square = .79; Working Memory latent variable r-square = .81. *No non-positive-definite matrix problems even though the latent variable variance were not set to 1.0* the hierarchical factor model also proved beneficial to improving the accuracy of the factor structure. For example, assigning informative priors provided away to capture the impact of the cross loadings with while maintaining the theoretical tenets of the WPPSI-IV structure. In this way, the guidelines of simple structure were met. The Bayesian approach also provided an efficient way to model the distributional characteristics of the subtest scores. For example, instead of transforming subtest scores or using elliptical estimators to adjust for skewness, subtest score distributions were modeled according to their true distributional characteristics. Perhaps most importantly, inadmissible solutions due to an on-positive latent variable matrix were eliminated. Overall, the didactic presented here provides a systematic approach to critically examine the structure of a second-order CFA model. Ideally, application of this strategy will contribute to improving the validity of scores based on a second-order CFA in test development.

References

1. Muthén B O and Asparouhov T. Bayesian structural equation modeling: A more flexible representation of substantive theory. Psychological Methods.2012; 17: 313-335.
2. Wechsler D. Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition. San Antonio, TX: NCS Pearson.
3. Kaplan D. Structural equation modeling: Foundations and extensions, 2nd edition.
4. Thousand Oaks: Sage Publications.2009.
5. Jöreskog K, and Sörbom D. LISREL 8: User's reference guide. Chicago, IL: Scientific Software International. (1996a).
6. Dunson D B, Palomo J and Bollen K (2005). Bayesian structural equation modeling.
7. Technical Report 2005-5.Statistics and Applied Mathematical Sciences Institute (SAMSI), Research TrianglePark, N.C.
8. Brown,T. Confirmatory factor analysis for applied research. NewYork: Guilford Press.2006.
9. McDonald R P. Test theory: A unified treatment. Mahwah NJ: Lawrence Erlbaum. 1999.
10. Thurstone L L. Multiple factor analysis. Chicago: University of Chicago Press. 1947
11. Kerlinger F N, Lee, H B. Foundations of behavioral research, 4th edition. Belmont CA: Cengage Learning. 2000.
12. Alfonso V C, Flanagan D P, & Radwan S. The impact of the Cattell-Horn-Carroll Theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P.L. Harrison (Eds.) Contemporary Intellectual Assessment: Theories, Tests & Issues. New York, NY: Guilford Press. 2005; 185-202.
13. Anderson J C, Gerbing D W. Structural equation modeling in practice: A review and recommended two step approach. Psychological Bulletin. 1988; 103: 411-423.
14. Boomsma A. On the robustness of LISREL against small sample size and non-normality. Amsterdam: Sociometric Research Foundation.1983.
15. Bentler P M, Chou C. Practical issues in structural equation modeling. Sociological Methods and Research. 1987; 16: 78-117.
16. MacCallum R C, Browne M W , Sugawara H. M. Power analysis is and determination of sample size for covariance structure modeling. Psychological Methods. 1996; 1: 130-149.
17. Ding C, Velicer, W F Harlow L L. The effects of estimation methods, number of indicators per factor and improper solutions on structural equation model fit indices. Structural Equation Modeling. 1995; 2: 119-144.
18. Yuan KH. Fit indices versus test statistics. Multivariate Behavioral Research, 2010; 40: 115-148.
19. Brooks S, Gelman A, Jones G, Meng X. Handbook of Markov Chain Monte Carlo. BocaRaton FL: Chapman-HallCRC. 2011.
20. Muthén B O. Multilevel covariance structure analysis. Sociological methods & research. 1994; 22: 376-398.
21. Muthén B O & Muthén L K. How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling. 2002; 4: 599-620.
22. Price LR. Small sample properties of Bayesian multivariate autoregressive time series models. Structural Equation Modeling. 2012; 19: 51-64.
23. Asparouhov,T.and Muthén B O. Exploratory structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal. 2009; 16: 397-438.
24. Fabrigar L R, Wegener D T. Exploratory factor analysis. New York, NY: Oxford University Press. 2012.
25. Mulaik S A. The foundations of factor analysis. NewYork, NY: McGraw-Hill.1972.
26. Marsh H W, Lüdtke O ,Muthén B O, Asparouhov T, Morin A J S, Trautwein U. A new look at the Big Five structure through exploratory structural equation modeling. Psychological Assessment. 2010; 22: 471-491.
27. Lee S Y. Structural equation modeling: A Bayesian approach. NewYork: Wiley. 2007.
28. Congdon, P.D. Applied Bayesian hierarchical methods. BocaRaton, FL:CRC. 2010.
29. Gelman A, Stern H, Carlin B, Rubin D. Bayesian data analysis. Boca Raton, FL: CRC Press. 2004.