

Higher–Order Factor Invariance and Idiographic Mapping of Constructs to Observables

Zhiyong Zhang

The University of Notre Dame

Michael W. Browne

The Ohio State University

John R. Nesselroade

The University of Virginia

Abstract

Building tailored construct representations that filter out irrelevant idiosyncratic information hampering the articulation of general lawful (nomothetic) relations by means of individual level factor analysis (P–technique) has been explored with initially promising results (Nesselroade, Gerstorf, Hardy, & Ram, 2007). We present a specification of an alternative to traditional factor invariance (higher–order invariance) that further formalizes these ideas and examine the results of some simulation experiments designed to evaluate the appropriateness of the implementation. The results also have implications for the use of between–persons variation and subgroup comparisons to establish relations among variables.

Keywords: Factor invariance, measurement invariance, higher–order invariance, likelihood ratio test

Unobserved or latent variables (constructs, factors) are a fixture of contemporary behavioral research and theorizing. Interpreting the relations among manifest variables by tying them to latent variables (factors) in a structural model has been common practice in behavioral research for several decades. When the key variables of psychological theory are not directly observable, in order to continue elaborating theory and building a knowledge base ways are sought to ensure that the same abstractions are being considered in different circumstances (e.g., different samples, different ages).

The concept of factorial invariance (e.g., Horn & McArdle, 1992; Meredith, 1964, 1993; Millsap & Meredith, 2007; Thurstone, 1947) has been a linchpin in dealing with the concerns associated with the use of latent variables. An offspring of factorial invariance, measurement invariance (MI), underlies evaluation of the measurement model component in structural equation modeling. Some form of factorial invariance has been generally accepted as the evidence that, indeed, the same abstract attribute is being measured in different circumstances.

Factor invariance has also served as the basis for distinguishing between quantitative and qualitative changes in developmental science (Baltes & Nesselroade, 1970; Nesselroade, 1970). Indeed, factor invariance was the “holy grail” of classical psychological measurement for most of the 20th century and bears

implicitly, if not explicitly, on other approaches to measurement. Still other influences of the perceived importance of factorial invariance were the promotion of the simple structure criterion for factor rotation which was believed to enhance the chances of demonstrating factor invariance from one study to another and the use of marker variables (e.g., Cattell, 1957; Thurstone, 1957) to identify major factors across studies involving different test batteries.

Factorial Invariance and Research Practice

The generally highly-regarded significance of factor invariance bears on research practice in various, sometimes conflicting ways. One of the key ones has been the sustained effort to use identical batteries of manifest variables to measure different comparison groups (e.g., different samples–same time, same sample–different times) in anticipation of conducting empirical tests of invariance. But problems arise when, for a variety of logical reasons, it makes sense for measurement batteries to differ from one instance to another. For example, in longitudinal studies covering substantial intervals of the lifespan, measurement batteries are periodically reconstituted to be more age-appropriate for the longitudinal participants. Or, in cross-sectional comparisons of vastly different age groups, items may deliberately be varied in content to better match them to known age/cohort differences. When the batteries of manifest variables characterizing the data sets being compared are different, factor invariance concepts have been difficult to utilize even though Thurstone (1947, 1957) defined invariance in terms of the factorial description of a test remaining the same when it is moved from one battery to another that involves the same common factors. This was clearly a reference to individual variables rather than entire sets of variables and forms part of the rationale for the use of marker variables in testing the hypothesis that the factor intercorrelation matrices do not differ statistically from one case to another.

Thurstone (1957) also explicitly discussed invariance in the case of two different test batteries (but the same underlying factors) being given to two different populations and argued that it should still be possible to identify the same factors in both data sets. He pointed out, however, that the factors would have to be identified independently in the two sets. Just how this is to be done is neither obvious nor a matter of common knowledge. For example, authoritarian behavior of fathers and submissiveness of their sons might be the target concepts in a study of their relations, but in one study authoritarian behavior of fathers might be indexed by measures of physical coercion while in another study it might be indexed only via verbal behavior measures. Similarly, sons' submissiveness in one study might be exhibited only in measures of docility while in the other it might be measured with ratings of passive-aggressive behaviors. Thus, for the case of different subsets of manifest variables, even though one cannot reach the invariance conclusion on the relations between the first-order factors and the observed variables, as in the usual factor invariance analysis, some way of establishing and assessing that the same factors or constructs are involved remains desirable.

In the basic factor analysis model each observed variable is constituted as a linear combination of the abstract factors plus a unique part. In this sense a given factor can be a cause of not only the variables explicitly included in a study, but of many other variables that are not included in the measurement battery. According to factor invariance theory deriving from the basic model, should one somehow also observe these latter variables and pool them with the original ones, the loadings of the original variables on the factors should remain unchanged. Similarly, if a small test battery is selected from a larger one, the factor loadings for all the tests should not change (Thurstone, 1947).

At the operational level, the conclusion is inescapable that there are many instances where it is necessary or desirable to be able to rest the conclusion that the "same" factors are being measured on somewhat different batteries of manifest variables. Arguments for higher-order invariance (Nesselroade et al., 2007) represent an extension of the reasons for this conclusion and an attempt to deal with the matter.

Higher-Order Factor Invariance

The invariance proposals by Nesselroade et al. (2007) involved focusing on invariance at the level of factor intercorrelations rather than the primary factor loadings, thereby defining invariance at a more abstract level. This is why it was termed higher-order invariance. In contrast to reasons given above for why measurement batteries might differ across circumstances, Nesselroade et al. argued that factor loading patterns (linkages between observable measures and factors) can reflect the idiosyncracies of individual histories of conditioning, learning, language usage, etc. that, to some extent, could interfere with demonstrating invariance and should be filtered from the measurements prior to fitting more general models.¹ Their approach to demonstrating higher-order invariance and the positive measurement features that it implies consisted of four main steps. First, use individual-level analyses initially, even though one's goal is to reach general conclusions.² So, for example, P-technique factor analyses on multiple individuals' data sets is a beginning point for seeking relations between latent and manifest variables that apply more generally (see e.g., Zevon & Tellegen, 1982). Second, as parsimoniously as possible allow first-order factor loading patterns to reflect idiosyncracies, thus ameliorating a strict view of the traditional conception of factor invariance. Third, try to fit a factor model that, while allowing for idiosyncrasy in the factor loadings, constrains the interrelations among the first-order factors (e.g., factor intercorrelations) to be invariant from one individual to another. Fourth, if the model fits, the factor intercorrelation matrices do not differ for the individuals, and one can try to fit a second-order factor solution (loadings of first-order factors on second-order factors) that will be invariant, in the traditional sense, across individuals.

Targeting the factor intercorrelations instead of the primary factor loading patterns obviously involves some major shifts in thinking about invariance and measurement properties. The key one is the focus on relations among constructs rather than the relations between constructs and manifest variables. The commentaries accompanying Nesselroade et al. (2007) identify many of the "red flags" that spring from the contrast between traditional notions of invariance and the higher-order invariance concepts. But with higher-order invariance, for example, there is no need to assume that what is objectively the same stimulus set (e.g., a set of test items) is functionally the same stimulus for different individuals or subgroups. Nor need one assume that particular response categories are the same from one individual to another. For instance, in addition to genetic variation, one person's learning/conditioning history may have helped to shape their respiration rate differently than another's. Although respiration rate in every case reflects sympathetic nervous system activity to some extent, it also reflects idiosyncrasy that is irrelevant to respiration rate as an indicator of sympathetic arousal. From this perspective, the purposes of a measurement model include both linking observed variables to target constructs and filtering out the idiosyncracies of manifest variables.

An implication of recognizing idiosyncrasy is that the primary factor loading patterns are not necessarily expected to be invariant—the filtering they reflect is different for different people or different subgroups. But, from the standpoint of building a nomothetic science, one still hopes for (and can try to verify) invariance at some level. In the higher-order invariance case, the intercorrelations among the latent variables or constructs was chosen as the locus of invariant relations. This is tantamount to arguing that theoretically-relevant mechanisms and processes are the same for different individuals or groups (invariant / nomothetic) but the "raw materials" on which they operate might not be (idiosyncratic / idiographic). For example, schedules of reinforcement in an operant conditioning paradigm have invariant relations to extinction patterns whether the manifest behavior is bar pressing by a rat or pecking by a pigeon. Bar pressing

¹Although sets of within-individual variation (P-technique) were used to illustrate higher-order invariance, it should be noted that the arguments apply to sets of between-persons variation (subgroup comparisons) as well (see e.g., Nesselroade & Estabrook, 2008).

²This is in keeping with the arguments presented so forcefully by Molenaar (2004) for the importance of studying behavior initially at the individual level. The valuable extensions of P-technique, collectively referred to as dynamic factor models (e.g., Browne & Nesselroade, 2005; McArdle, 1982; Molenaar, 1985; Nesselroade, McArdle, Aggen, & Meyers, 2002) have greatly enriched the possibilities for making the more intensive study of the individual a precursor to developing general lawful relations.

and pecking are very different manifest variables but reinforcement schedules and extinction patterns are the same abstractions whether the subject is a rat or a pigeon or the behavior involves a paw or a beak.

Thus, following the four steps given above indicates a factor solution that meets the traditional definition of factor invariance albeit with the invariant factors being at least one level removed in abstraction from the typical level at which invariance is sought i.e., at the second-order level instead of the first-order level. With traditional factor invariance manifested at a higher-order level and some idiosyncrasy characterizing the relations between the observed variables and the first-order factors, it can be argued that the first-order factors are providing a “filtering” action that somewhat idiosyncratically links behavior patterns to invariant higher-order constructs, hence the use of the term “idiographic filter” in Nesselrode et al. (2007). Because it is possible to calculate the loadings of the observed variables directly on the higher-order factors via transformations such as Cattell-White and Schmidt-Leiman (e.g., Cattell, 1966; Loehlin, 1998; Schmidt & Leiman, 1957), one can identify idiosyncratic patterns of observed variables defining invariant higher-order constructs.

We have argued that there are sound reasons for approaching the identification of invariance at the level of factor intercorrelations. However, there is no guarantee that it can be demonstrated in a given instance. Whether or not higher-order invariance is found to be tenable is an empirical question to be answered from one’s data. If higher-order invariance does obtain, then it has important implications for measurement approaches as well as the further development and testing of theory, as we have suggested.

In the remainder of this article, we build on Thurstone’s explorations of invariance on different sets of manifest variables (Thurstone, 1957) and Cattell’s arguments for the use of marker variables (Cattell, 1957) to clarify and extend in important ways the arguments by Nesselrode et al. (2007) regarding higher-order invariance. First, we formalize higher-order invariance based on multiple preexisting populations with equal factor correlation matrices but different manifest variable correlation matrices, factor loading matrices, and unique factor covariance matrices. Second, we discuss conditions for a testable hypothesis of higher-order invariance. Third, we show how to evaluate higher-order invariance through multiple group analysis. Fourth, we conduct several random sampling experiments to explore the factors that influence the test of higher-order invariance and how well the likelihood ratio test can detect higher-order invariance or the lack of it.

Testing for Higher-Order Invariance

Assume there are g pre-existing infinite populations with manifest variable covariance matrices

$$\Sigma_i = \Lambda_i \Phi_i \Lambda_i' + \Psi_i, \quad i = 1, \dots, g, \quad (1)$$

with

$$\Phi_1 = \Phi_2 = \dots = \Phi_g = \Phi, \quad (2)$$

where Σ_i is the population covariance matrix, Λ_i is the factor loading matrix, Φ_i is the common factor correlation matrix, and Ψ_i is a diagonal covariance matrix of the unique factors for the i -th population. Note that the manifest variable covariance matrices, the factor loading matrices, and the unique factor covariance matrices may vary from one population to another but all factor correlation matrices are equal.

The equality of factor correlation matrices in (2) is not testable without further constraints because whenever (1) is true it is possible to carry out independent oblique rotations in each population so that (2) is satisfied. In fact, for any desired factor correlation matrix Φ with symmetric square root $\Phi^{1/2}$, one can construct the transformation matrices $\mathbf{T}_i = \Phi_i^{1/2} \Phi^{-1/2}$. It is easy to see that

$$\mathbf{T}_i^{-1} \Phi_i \mathbf{T}_i^{-1'} = \Phi^{1/2} \Phi_i^{-1/2} \Phi_i \Phi_i^{-1/2} \Phi^{1/2} = \Phi.$$

Let $\Lambda_i^* = \Lambda_i \mathbf{T}_i$, then one has

$$\Lambda_i \Phi_i \Lambda_i' = \Lambda_i^* \Phi \Lambda_i^{*'}.$$

Thus, before (2) can represent a testable hypothesis, it is necessary to identify each Λ_i under oblique rotation. The rotation could be done through any method such as direct quartimin (Browne, 2001; Jennrich, 2007). After rotation, the hypothesis in (2) could then be tested. However, failing to reject the null hypothesis in (2) is not sufficient to conclude that higher-order factor invariance holds.³ For example, it is conceivable that one might find equality of two factor correlation matrices representing two totally different domains, a point made by several of the commentators who discussed Nesselroade et al. (2007).

Thus, additional conditions are required to make sure that a factor from one population matches the corresponding factor from another population. One way is to match factors by “theory”, “inspection”, “psychological intuition of meaning”, or “insightful judgments of psychological similarity” (Cattell, 1957; Reyburn & Raath, 1950). Based on these criteria, the identification conditions for $\Lambda_i, i = 1, \dots, g$ can be decided and the factors from the g groups can be matched. Even if the g groups have no variables in common, the test of factor invariance can still proceed. However, this method may easily fail due to the subjective decisions that are entailed. Thus, some rigorous decision criteria should be used.

Cattell (1957) suggested the use of marker variables in different studies to match factors. Marker variables are variables introduced into a factor analytic experiment to identify, unambiguously, factors on which they are known to be highly loaded. He further suggested that at least two marker variables for each factor should be used. The marker variable approach can also be used to investigate the invariance of factors with different indicating variables in our current application. However, one marker variable for each factor in each group is sufficient to identify the factors given a well-defined factor structure.

We will first illustrate the use of marker variables in a two population example and then discuss their advantages. Assume there are three factors – f_1, f_2 , and f_3 – in each group. Thus, we need three marker variables, M_1, M_2 , and M_3 to match the factors. Further assume there are p_1 and p_2 manifest variables in each population, respectively. Besides the common marker variables, the other manifest variables may or may not be the same in the two populations. The factor loading matrices for the two populations are specified in Figure 1. “+” in the figure represents a large positive loading to be estimated for the marker variables, “?” represents loadings of indeterminate magnitude for the non-marker variables and “0” indicates factor loadings that are equal to 0. The marker loading is free to vary without constraint but an attempt should be made to choose marker variables with substantial expected loadings. All other free loadings are for non-marker variables that can assume any value and for which there is no particular expectation as to their magnitude.

	Group 1				Group 2		
	f_1	f_2	f_3		f_1	f_2	f_3
M_1	+	0	0	M_1	+	0	0
M_2	0	+	0	M_2	0	+	0
M_3	0	0	+	M_3	0	0	+
A_1	?	?	?	B_1	?	?	?
A_2	?	?	?	B_2	?	?	?
A_3	?	?	?	B_3	?	?	?
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$A_{p_1 - 3}$?	?	?	$B_{p_2 - 3}$?	?	?

Figure 1: A demonstration of the use of marker variables. + indicates freely-estimated positive marker variable loadings. ? represents all other free loadings which can assume any values, depending on the data.

³Molenaar (2008) in a written commentary on Nesselroade et al. (2007) noted this as a problem needing to be solved if higher-order invariance is to be formalized.

There are several advantages of using marker variables besides its automatic role for model identification. First, they provide researchers a substantive basis on which to match factors so matching factors across studies can retain a theory-driven, substantive flavor. Second, given a firm substrate for the matching provided by markers, putatively unique features of the populations or groups can be indulged by including different non-marker variables in the respective choices of manifest variables. For example, test batteries can be customized for different groups to better “flesh out” what is conceptualized to be the nature of the factors in the different groups. Third, sometimes, different groups may react differently to what is ostensibly the same set of variables, rendering the test batteries functionally different. With the assurance of continuity provided by marker variables, that different reactions have occurred can be detected in the estimated factor loadings, if they are allowed to vary from group to group. Thus, marker variables can play a valuable role in the context of building measurement systems that afford a way to make comparisons among individuals or groups while also allowing for idiosyncrasy in the overall measurement schemes. This general line of reasoning regarding marker variables has much in common with Thurstone’s notion of “factor pure” tests.

Procedure for testing higher-order factor invariance

After matching the factors on the basis of the marker variables, we can assess higher-order factor invariance through testing $\Phi_1 = \Phi_2 = \dots = \Phi_g$. Now we illustrate the procedure for finite samples. Assume that there are g finite samples that come from the corresponding populations. The sample size of the i th sample is n_i and there are p_i manifest variables (including q marker variables) that indicate q factors in each sample. Let $\mathbf{S}_i, i = 1, \dots, g$ denote the observed covariance matrices and $\Sigma_i = \Lambda_i \Phi_i \Lambda_i' + \Psi_i, i = 1, \dots, g$ denote the model indicated covariance matrix. To identify the model, the factor loadings for the marker variables and non-marker variables are specified as in Figure 1. To estimate the parameters, we can minimize the discrepancy function (e.g., Jöreskog, 1971),

$$F(\Sigma_{\dots}; \mathbf{S}_{\dots}) = \sum_{i=1}^g \frac{n_i}{n} (\log |\Sigma_i| + \text{tr}(\Sigma_i^{-1} \mathbf{S}_i) - \log |\mathbf{S}_i| - p_i) \quad (3)$$

where $n = \sum_{i=1}^g n_i$ and \mathbf{M}_{\dots} represents short hand denotation of g matrices $\mathbf{M}_1, \dots, \mathbf{M}_g$. This discrepancy function is a nonnegative scalar valued function that is equal to zero if and only if $\Sigma_i = \mathbf{S}_i, i = 1, \dots, g$. The discrepancy function value is a useful measure of goodness of fit of a model where a model with better fit should have a minimum discrepancy function value closer to zero.

To test higher-order factor invariance, consider three models, a saturated model, a non-invariant model, and an invariant model. The saturated model, denoted by **Model S**, is

$$\mathbf{Model S: } \Sigma_i^{(s)}, \quad i = 1, \dots, g \text{ are any positive definite matrices.} \quad (4)$$

Note that the minimum discrepancy function is zero for the saturated model at $\widehat{\Sigma}_i^{(s)} = \mathbf{S}_i, i = 1, \dots, g$. The non-invariant model, denoted by **Model N**, is a g -group factor model with different factor correlation matrices in each group,

$$\mathbf{Model N: } \Sigma_i^{(n)} = \Lambda_i \Phi_i \Lambda_i' + \Psi_i, \quad i = 1, \dots, g. \quad (5)$$

The higher-order invariant model, denoted by **Model H**, is also a g -group factor model but with the same factor correlation matrices specified for each group,

$$\mathbf{Model H: } \Sigma_i^{(h)} = \Lambda_i \Phi \Lambda_i' + \Psi_i, \quad i = 1, \dots, g. \quad (6)$$

The invariant model can be constructed from the non-invariant model by setting $\Phi_i = \Phi$ for $i = 1, \dots, g$.

Let $\Sigma_{\dots}^{(0)}$ represent the g true population covariance matrices. These are known in a random sampling experiment but are not known in practice. Let $\widehat{\Sigma}_{\dots}^{(h)}$ denote the g matrices satisfying Model H that minimize

$F(\boldsymbol{\Sigma}_{\dots}^{(h)}; \mathbf{S}_{\dots})$ and $F(\widehat{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \mathbf{S}_{\dots})$ is the corresponding minimum discrepancy function value. Furthermore, let $\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(h)}$ denote the g matrices satisfying Model H that minimize $F(\boldsymbol{\Sigma}_{\dots}^{(h)}; \boldsymbol{\Sigma}_{\dots}^{(0)})$ and $F(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \boldsymbol{\Sigma}_{\dots}^{(0)})$ is the corresponding minimum discrepancy function value. The statistic $nF(\widehat{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \mathbf{S}_{\dots})$ is the likelihood ratio test statistic for testing Model H as null hypothesis against the saturated model - Model S - as alternative. Its distribution may be approximated by a noncentral chi-square distribution with

$$g \frac{(p-q)^2 - (p+q)}{2} + (g-1) \frac{q(q-1)}{2} \quad (7)$$

degrees of freedom and noncentrality parameter $nF(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \boldsymbol{\Sigma}_{\dots}^{(0)})$. Similarly, when testing Model N as null hypothesis against Model S as alternative, the statistic $nF(\widehat{\boldsymbol{\Sigma}}_{\dots}^{(n)}; \mathbf{S}_{\dots})$ approximately follows a noncentral chi-square distribution with

$$g \frac{(p-q)^2 - (p+q)}{2} \quad (8)$$

degrees of freedom and noncentrality parameter $nF(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(n)}; \boldsymbol{\Sigma}_{\dots}^{(0)})$. Here, $\widehat{\boldsymbol{\Sigma}}_{\dots}^{(n)}$ and $\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(n)}$ are matrices satisfying Model N and Model S that minimize $F(\boldsymbol{\Sigma}_{\dots}^{(n)}; \mathbf{S}_{\dots})$ and $F(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(n)}; \boldsymbol{\Sigma}_{\dots}^{(0)})$ with $F(\widehat{\boldsymbol{\Sigma}}_{\dots}^{(n)}; \mathbf{S}_{\dots})$ and $F(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(n)}; \boldsymbol{\Sigma}_{\dots}^{(0)})$ denoting the corresponding minimum discrepancy function values, respectively.

For the higher-order factor invariance test, we come to the test of Model H as null hypothesis against Model N as alternative hypothesis. The likelihood ratio test statistic is the difference

$$\Delta\chi^2 = nF(\widehat{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \mathbf{S}_{\dots}) - nF(\widehat{\boldsymbol{\Sigma}}_{\dots}^{(n)}; \mathbf{S}_{\dots}) \quad (9)$$

whose distribution may be approximated by a noncentral chi-square distribution with

$$(g-1) \frac{q(q-1)}{2}$$

degrees of freedom and noncentrality parameter

$$n \left[F(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \boldsymbol{\Sigma}_{\dots}^{(0)}) - F(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(n)}; \boldsymbol{\Sigma}_{\dots}^{(0)}) \right] = nF(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \boldsymbol{\Sigma}_{\dots}^{(0)}) \quad (10)$$

because $\boldsymbol{\Sigma}_{\dots}^{(0)}$ is assumed to satisfy the alternative model so that $\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(n)} = \boldsymbol{\Sigma}_{\dots}^{(0)}$.

With higher-order invariance so that $\boldsymbol{\Phi}_i = \boldsymbol{\Phi}$ for $i = 1, \dots, g$, $\boldsymbol{\Sigma}_{\dots}^{(0)}$ also satisfies the null hypothesis and $F(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \boldsymbol{\Sigma}_{\dots}^{(0)}) = 0$. For the situation of lack of invariance, $F(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \boldsymbol{\Sigma}_{\dots}^{(0)})$ is larger than zero and is consequently a measure of equality (or inequality) of $\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_g$. With the discrepancy function value $F(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \boldsymbol{\Sigma}_{\dots}^{(0)})$, we can calculate the RMSEA, ε , (Browne & Cudeck, 1993) as an effect size measure of inequality of factor correlation matrices $\boldsymbol{\Phi}_i, i = 1, \dots, g$,

$$\varepsilon = \sqrt{\frac{F(\widetilde{\boldsymbol{\Sigma}}_{\dots}^{(h)}; \boldsymbol{\Sigma}_{\dots}^{(0)})}{(g-1)p(p-1)/2}}. \quad (11)$$

Note that the RMSEA is independent of sample size.

Practically, two conditions need to be considered in testing higher-order invariance. First, before conducting the test of Model H as null hypothesis against Model N as alternative hypothesis, we need to check the factor loadings for the marker variables. Because the factors are solely indicated by the marker

variables, the factor loadings for the marker variables should be sufficiently large. We suggest visually verifying that the loadings are substantial, for example, a loading (after standardization) that is larger than 0.5 can be considered as substantial based on substantive theory. Unfortunately, if the loadings for marker variables are too small, such as less than 0.1, we may be forced to conclude that there is no factor invariance. Note that although the marker variables are the common variables across groups, we should not constrain their non-zero loadings to be the same to avoid unnecessary restrictions on factor loadings.

After verifying that the loadings for marker variables are substantial, we can continue to test the invariance of factor correlation matrices. We first estimate Φ_i , $i = 1, \dots, g$ from Model N and obtain the chi-square value $\chi_{df1}^2 = nF(\hat{\Sigma}^{(n)}; \mathbf{S} \dots)$ with $df1$ representing the degrees of freedom defined by Eq (8).

Then we set $\Phi_1 = \Phi_2 = \dots = \Phi_g$ and obtain the $\chi_{df2}^2 = nF(\hat{\Sigma}^{(h)}; \mathbf{S} \dots)$ with $df2$ defined by Eq (8). If the difference $\Delta\chi^2 = \chi_{df2}^2 - \chi_{df1}^2$ as the likelihood ratio test statistic is smaller than the critical value from the chi-square distribution with degrees of freedom $(g-1)q(q-1)/2$, we have failed to reject the null hypothesis and thus conclude that higher-order factor invariance prevails. Otherwise, the factor correlation matrices are non-invariant.

Differences Between Common Factor Invariance and Higher-Order Invariance

An essential difference between pursuing traditional factor invariance and higher-order invariance is that the former requires that the same set of variables be used across comparison groups. As shown in Figure 2, for common factor invariance, the same manifest variables $A1, A2, A3, A4$, and $A5$ are used for both groups. However, for higher-order invariance, only the maker variables $M1$ and $M2$ need be common variables. The other variables can be totally different as is the case with $A3, A4$, and $A5$ with $B3, B4$, and $B5$.

This difference can be made still clearer by further investigating the definition of factor invariance. If a set of variables X can be decomposed into common and unique factors, X is factorable. If the factor matrix remains the same under selection, X is said to be factorially invariant (Meredith, 1993). Thus, traditional factor invariance focuses on testing the invariance of factor loadings of the same set of manifest variables across groups (e.g., Horn & McArdle, 1992; Meredith, 1964, 1993). If invariance obtains at that level, the observed variables are argued to have the same meaning or to measure the same construct across groups (e.g., Bollen, 1989; French & Finch, 2006). If this form of invariance does not hold, the observed variables are construed as showing different meanings in the different subgroups, i.e., the observed variables are essentially different variables for different groups in which case the latent constructs indicated by the observed variables are construed as different across groups.⁴

A prime question for higher-order invariance to answer is whether or not the factors indicated by different sets of variables, the selection of which has been fine-tuned to conform to the features of each measured population, are the same factors. To ensure the factors are from the same domain, prior knowledge can be used and the marker variable approach discussed earlier is one way to use such knowledge. Demonstrating that the factor loadings of the markers are substantial offers some concrete evidence that the factors are as hypothesized. Additionally testing the invariance of factor correlation matrices supports an inference that the factors are invariant. Identifying the patterns by which different manifest variables load the same factors further enriches substantive understanding of the factor in different individuals or subgroups.

Random Sampling Experiments

We have discussed reasons for investigating higher-order invariance and how to test for it. In this section, we explore possible influences on the test of higher-order invariance through random sampling

⁴Although measurement invariance may exist when factor invariance does not hold (Meredith, 1993), without the latter researchers generally conclude that the variables do not measure the same construct.

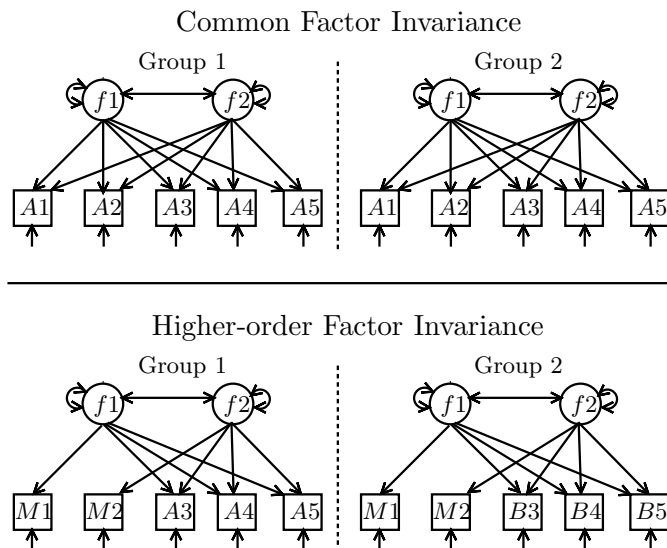


Figure 2: The comparisons of the usual factor invariance and higher-order factor invariance. In the top figure, the correlations between f_1 and f_2 are freely estimated and the factor loadings are often constrained to be the same for the two groups. In the bottom figure, the correlations are set to be equal but all factor loadings are freely estimated.

experiments. In the first three experiments, we investigate how the number of non-marker variables, the loadings of marker variables, and the communalities affect the factor correlation estimates. In the fourth experiment, we look at the performance of the likelihood ratio test in detecting higher-order invariance.

Experiment 1: Using only marker variables

We have indicated how marker variables can be used to identify the same factors under higher order invariance conditions. Using only one marker variable for each factor provides a way to test higher-order invariance and idiographic features of the non-marker variables simultaneously. However, it is not suitable to use only marker variables for higher-order invariance analysis. If the marker variables alone are used, the number of observed variables (p) is equal to the number of marker variables (q), which is also the number of factors. The estimates of the factor correlations are the manifest variable correlations. It can be shown that the correlations between manifest variables underestimate the factor correlations.

When $p = q$, the factor loading matrix $\mathbf{\Lambda} = \mathbf{D}_\lambda$ is a diagonal matrix and thus the population manifest variable covariance matrix

$$\mathbf{\Sigma} = \mathbf{D}_\lambda \mathbf{\Phi} \mathbf{D}_\lambda + \mathbf{D}_\psi,$$

where $\mathbf{\Phi}$ is factor correlation matrix, and \mathbf{D}_ψ is unique factor covariance matrix and is also a diagonal matrix. Let \mathbf{D}_σ denote the diagonal matrix formed by the diagonal elements of $\mathbf{\Sigma}$. Then $\mathbf{P} = \mathbf{D}_\sigma^{-1/2} \mathbf{\Sigma} \mathbf{D}_\sigma^{-1/2}$ becomes the manifest variable correlation matrix and

$$\mathbf{P} = \mathbf{D}_\zeta \mathbf{\Phi} \mathbf{D}_\zeta + \mathbf{D}_\psi^*$$

where $\mathbf{D}_\zeta = \mathbf{D}_\sigma^{-1/2} \mathbf{D}_\lambda$ and $\mathbf{D}_\psi^* = \mathbf{D}_\sigma^{-1/2} \mathbf{D}_\psi \mathbf{D}_\sigma^{-1/2}$. The population factor correlation matrix $\mathbf{\Phi}$ are thus related to \mathbf{P} by the attenuation formulae

$$\text{Off}[\mathbf{P}] = \text{Off}[\mathbf{D}_\zeta \mathbf{\Phi} \mathbf{D}_\zeta]$$

where $\text{Off}[\mathbf{A}]$ represents the matrix formed from the off-diagonal elements of \mathbf{A} with zeros in the diagonal and \mathbf{D}_ζ is also a diagonal matrix of square roots of communalities. Because communalities lie between zero and one, the correlation between the manifest variables are in general smaller than correlations between factors. Consequently, sample correlations between manifest variables will be both biased and inconsistent estimators of correlations between factors.

For the purpose of demonstration, let $p = q = 3$ and set

$$\mathbf{\Lambda} = \begin{bmatrix} .707 & 0 & 0 \\ 0 & .707 & 0 \\ 0 & 0 & .707 \end{bmatrix} \quad \mathbf{\Phi} = \begin{bmatrix} 1 & .5 & .4 \\ .5 & 1 & .3 \\ .4 & .3 & 1 \end{bmatrix},$$

then the manifest variable correlation matrix is (the communalities are set at .5)

$$\mathbf{P} = \begin{bmatrix} 1 & .25 & .2 \\ .25 & 1 & .15 \\ .2 & .15 & 1 \end{bmatrix}.$$

Clearly, in this simple example the correlations of manifest variables are only half the size of the corresponding correlations between factors.

Experiment 2: The effect of non-marker variables

We have shown that using only the marker variables is not sufficient to obtain accurate factor correlation estimates. Now, we investigate how the number of non-marker variables influences the estimation of the factor correlation matrix. Increasing the number of non-marker variables increases the model degrees of freedom which are determined by

$$\text{df} = \frac{(p - q)^2 - (p + q)}{2}.$$

For the purpose of demonstration, we set $q = 3$ and change the value of p . We compare the factor correlation estimates with $p = 6, 9, 12$, and 15 , corresponding to factor models with 2, 3, 4 and 5 indicating variables for each of the three factors. The factor correlation matrix is set as

$$\mathbf{\Phi} = \begin{bmatrix} 1 & .5 & .4 \\ .5 & 1 & .3 \\ .4 & .3 & 1 \end{bmatrix}, \quad (12)$$

and the uniqueness covariance matrix $\mathbf{\Psi} = (\psi_{ij})$ is set as $\psi_{ii} = 0.5, i = 1, \dots, p$ and $\psi_{ij} = 0, i \neq j, i, j = 1, \dots, p$. With this setting, all communalities have been chosen to be 0.5 so that effects due to an arbitrary choice of different communalities are avoided.

The specification of the factor loading matrices for models with the increasing number of non-marker variables (different p) is given in Table 1. The degrees of freedom for the four models are 0, 12, 33, and 63, respectively. Note that when adding non-marker variables we specify the factor loadings to be the same as the other non-marker variables to control the effects of the magnitude of the factor loadings.

To compare the factor correlation estimates with different numbers of non-marker variables, we simulate R sets of data using the factor loading matrix specifications in Table 1 and estimate the factor correlations. Let $\hat{\phi}_{ij}^{(r)}$ denote the estimated factor correlation between the i -th and j -th factors obtained from the r -th replication and let $\phi_{ij}^{(o)}$ be the corresponding population value. Furthermore, let z_r stand for the Average Squared Error (ASE) measure of accuracy of $\hat{\mathbf{\Phi}}$ obtained from the r -th replication where

$$z_r = \frac{2}{q(q-1)} \sum_{j=2}^q \sum_{i=1}^{j-1} (\hat{\phi}_{ij}^{(r)} - \phi_{ij}^{(o)})^2.$$

Table 1: The influences of non-marker variables on factor correlation estimates

model	Λ	df	EASE	CI
$p = 6, q = 3$	$\begin{bmatrix} .707 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & .707 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & .707 \\ .410 & .410 & .000 \\ .000 & .439 & .439 \\ .422 & .000 & .422 \end{bmatrix}$	0	.01437	[.01429, .01446]
$p = 9, q = 3$	$\begin{bmatrix} .707 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & .707 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & .707 \\ .410 & .410 & .000 \\ .000 & .439 & .439 \\ .422 & .000 & .422 \\ .410 & .410 & .000 \\ .000 & .439 & .439 \\ .422 & .000 & .422 \end{bmatrix}$	12	.00808	[.00804, .00812]
$p = 12, q = 3$	$\begin{bmatrix} .707 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & .707 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & .707 \\ .410 & .410 & .000 \\ .000 & .439 & .439 \\ .422 & .000 & .422 \\ .410 & .410 & .000 \\ .000 & .439 & .439 \\ .422 & .000 & .422 \\ .410 & .410 & .000 \\ .000 & .439 & .439 \\ .422 & .000 & .422 \end{bmatrix}$	33	.00722	[.00717, .00727]
$p = 15, q = 3$	—*	63	.00674	[.00670, .00678]

Note. df: degrees of freedom. *: The factor loading matrix here has the same structure as the one above with three more observed non-marker variables. The factor loading matrix is omitted to save space.

The Expected Average Squared Error (EASE, denoted by μ_z) as a measure of accuracy of the estimates of Φ for a model can be estimated by

$$\bar{z} = \frac{1}{R} \sum_{r=1}^R z_r.$$

Furthermore, a $100(1 - \alpha)\%$ confidence interval for EASE could be constructed based on the Central Limit Theorem using

$$\bar{z} - n_{\alpha/2} \frac{s_z}{\sqrt{R}} \leq \mu_z \leq \bar{z} + n_{\alpha/2} \frac{s_z}{\sqrt{R}},$$

where $n_{\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the standard normal distribution and

$$s_z^2 = \frac{1}{R-1} \sum_{r=1}^R (z_r - \bar{z})^2.$$

Table 1 presents the EASE and its 95% confidence interval for the factor correlation estimates from different models with the sample size of 500 based on $R = 100,000$ replications. The results clearly show that with more non-marker variables, the EASE becomes smaller. The confidence intervals for the EASE do not overlap for the four models. Thus, the use of non-marker variables can help estimate the factor correlations more accurately. Furthermore, based on the magnitude of the EASE, it seems that increasing the number of non-marker variables makes a big difference initially but an asymptote appears to be approached with further increases.

Table 2: The influences of factor loadings and communalities on factor correlation estimates

Model	Λ	ζ^2	EASE	CI
M1	$\begin{bmatrix} .707 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & .707 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & .707 \\ .410 & .410 & .000 \\ .000 & .439 & .439 \\ .422 & .000 & .422 \\ .410 & .410 & .000 \\ .000 & .439 & .439 \\ .422 & .000 & .422 \end{bmatrix}$	$\begin{bmatrix} .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \end{bmatrix}$.00808	[.00804, .00812]
M2	$\begin{bmatrix} .707 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & .707 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & .707 \\ .707 & .000 & .000 \\ .000 & .707 & .000 \\ .000 & .000 & .707 \\ .707 & .000 & .000 \\ .000 & .707 & .000 \\ .000 & .000 & .707 \end{bmatrix}$	$\begin{bmatrix} .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \end{bmatrix}$.00638	[.00635, .00641]
M3	$\begin{bmatrix} .316 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & .316 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & .316 \\ .707 & .000 & .000 \\ .000 & .707 & .000 \\ .000 & .000 & .707 \\ .707 & .000 & .000 \\ .000 & .707 & .000 \\ .000 & .000 & .707 \end{bmatrix}$	$\begin{bmatrix} .1 \\ .1 \\ .1 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \\ .5 \end{bmatrix}$.04771	[.04747, .04794]
M4	$\begin{bmatrix} .707 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & .707 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & .707 \\ .316 & .000 & .000 \\ .000 & .316 & .000 \\ .000 & .000 & .316 \\ .316 & .000 & .000 \\ .000 & .316 & .000 \\ .000 & .000 & .316 \end{bmatrix}$	$\begin{bmatrix} .5 \\ .5 \\ .5 \\ .1 \\ .1 \\ .1 \\ .1 \\ .1 \\ .1 \end{bmatrix}$.02311	[.02299, .02323]

Note. ζ^2 : Communalities.

Experiment 3: The effects of factor loadings and communalities

To investigate whether the distribution of factor loadings and the magnitude of the communalities influence the factor correlation estimates, we designed an experiment in the following way. First, the factor correlation matrix is the same as in Experiment 2 and the number of manifest variables is fixed at 9. Four models with different factor loading matrices and communalities are summarized in Table 2

To compare the factor correlation estimates, we again obtain the EASE for each model based on $R = 100,000$ replications of simulation with the sample size 500. The EASE and its 95% confidence interval for the four models are given in Table 2. Models M1 and M2 have the same communalities (0.5) but Model M2 has a simpler structure than Model M1. The EASEs from the two models are very close although overall M2 seems to have better factor correlation estimates. The difference between M2 and M3 lies in the communalities for marker variables in M3 being much smaller. The much larger EASE for M3 compared to M1 indicates that low communalities in marker variables can severely reduce the accuracy of the factor correlation estimates. M3 and M4 have the same factor structure. However, M3 has low communalities on the marker variables while M4 has low communalities on the non-marker variables. The much lower ASE for M4 compared to M3 suggests that the communalities of marker variables play a more important role than those of non-marker variables in estimating the factor correlations.

Experiment 4: Finite sample properties of likelihood ratio test for higher-order invariance

We have suggested the use of the likelihood ratio test for higher-order invariance. Asymptotically, the likelihood ratio test statistic $\Delta\chi^2$ has a non-central chi-square distribution. However, the performance of the test in detecting higher-order invariance and lack of invariance in finite samples is yet to be evaluated. Thus, in this simulation experiment, we investigate how well the likelihood ratio test can detect higher-order invariance. In other words, we evaluate how well the non-central chi-square distribution can approximate the empirical distribution of the likelihood ratio test statistics.

Simulation design. The simulation focuses on the higher-order invariance of two populations (groups). For the first group, the population parameters are set as in M1 in Table 2. For the second group, the population parameters are determined according to the effect size of the inequality of the factor correlations.

The RMSEA (ε) in Eq (11) can be used as a measure of effect size in factor correlation inequality in this experiment. To determine the population RMSEA, we need to know $\Sigma_{\dots}^{(0)}$ and it is known only in random sample experiments. For the two group in the experiment, the factor correlation matrices are set, respectively, as

$$\Phi_1 = \begin{bmatrix} \mathbf{1} & .5 & .4 \\ .5 & \mathbf{1} & .3 \\ .4 & .3 & \mathbf{1} \end{bmatrix} \text{ and } \Phi_2 = \begin{bmatrix} \mathbf{1} & .5 + d & .4 + d \\ .5 + d & \mathbf{1} & .3 + d \\ .4 + d & .3 + d & \mathbf{1} \end{bmatrix}, \quad (13)$$

where d is the difference between corresponding factor correlations of the two groups in the experiment. To facilitate comparisons, the communalities are always fixed at 0.5. Thus, the factor loadings are different for different effect sizes. Furthermore, the variances for the unique factors are always fixed at 0.5.

Under these specifications, for each value of d , the population covariance matrices $\Sigma_{\dots}^{(0)}$ are determined and the RMSEA can be calculated. In this experiment, we implemented four levels of RMSEA: 0, 0.05, 0.08, and 0.12 corresponding to zero, small, medium, and large effect sizes, respectively. For the four levels of RMSEA, the corresponding d values are 0, $-.227$, $-.35$, and $-.499$, respectively.

The likelihood ratio test statistic has an asymptotic non-central chi-square distribution with $d = (g - 1)q(q - 1)/2$ degrees of freedom and noncentrality parameter $\delta = d(n_1 + n_2)\varepsilon^2$ where n_1 and n_2 are sample sizes of the two simulated groups, respectively. Note that when the RMSEA is $\varepsilon = 0$, there exists higher-order invariance and the noncentrality parameter is 0.

In this experiment, we obtained the empirical distribution for the likelihood ratio test statistic based on $R = 2000$ simulation replications for sample sizes at $n_1 = n_2 = 100$ to 500 with an interval of 100 for each of the two groups. For each simulation replication, the likelihood ratio test statistic was retained as the chi-square difference ($\Delta\chi_i^2, i = 1, \dots, R$) between the invariance model and the non-invariance model.

Simulation results. Let $\chi_{1-\alpha}^2(3)$ be the critical value of a chi-square distribution with $d = (g - 1)q(q - 1)/2 = 3$ degrees of freedom at the level α . With the likelihood ratio test statistic from Monte Carlo simulation, we calculated $\hat{\alpha}$ by

$$\hat{\alpha} = \frac{\#[\Delta\chi_i^2 > \chi_{1-\alpha}^2(3)]}{R},$$

where $\#[\Delta\chi_i^2 > \chi_{1-\alpha}^2(3)]$ is the number of replications of simulation where $\Delta\chi_i^2$ exceeded $\chi_{1-\alpha}^2(3)$. When RMSEA = 0, the $\hat{\alpha}$ is the estimated Type I error and when the RMSEA > 0, the $\hat{\alpha}$ is estimated power.

For the simulation, the asymptotic non-central chi-square distribution for the likelihood ratio statistic is known. Thus, we can calculate

$$\tilde{\alpha} = \Pr[\chi^2(3, \delta) > \chi_{1-\alpha}^2(3)]$$

where $\delta = d(n_1 + n_2)\varepsilon^2$ and $\chi^2(d, \delta)$ denotes a non-central chi-square distribution with d degrees of freedom and non-centrality parameter δ . Note when RMSEA = 0, $\delta = 0$ and thus $\tilde{\alpha} = \alpha$. If the non-central chi-square distribution can approximate the distribution of the likelihood test statistics well, we should expect that $\tilde{\alpha}$ is close to $\hat{\alpha}$.

Table 3: Comparison of the empirical distribution and the non-central chi-square approximation distribution of the likelihood ratio test statistic

	RMSEA	d	Sample size ($n_1 = n_2$)				
			100	200	300	400	500
$\hat{\alpha}$	0	0	0.087	0.068	0.065	0.059	0.052
	0.05	-0.227	0.185	0.267	0.411	0.515	0.618
	0.08	-0.35	0.370	0.636	0.817	0.917	0.963
	0.12	-0.499	0.675	0.947	0.990	1.000	1.000
$\tilde{\alpha}$	0	0	0.05	0.05	0.05	0.05	0.05
	0.05	-0.227	0.153	0.275	0.400	0.518	0.623
	0.08	-0.35	0.345	0.634	0.824	0.924	0.970
	0.12	-0.499	0.692	0.951	0.995	1.000	1.000

Table 3 presents $\hat{\alpha}$ and $\tilde{\alpha}$ with different sample sizes and effect sizes when $\alpha = 0.05$. When $\text{RMSEA} = d = 0$, $\hat{\alpha}$ is the proportion of simulations that reject the null hypothesis that there is higher-order invariance. With small sample sizes, $\hat{\alpha}$ is greater than the nominal level 0.05. With increased sample size, it comes closer to 0.05. When $\text{RMSEA} > 0$, $\hat{\alpha}$ and $\tilde{\alpha}$ are measures of power in detecting non-invariance. Overall, with increased sample size, the power increases. Across the rows of the table, $\hat{\alpha}$ is close to $\tilde{\alpha}$ especially with a greater RMSEA.

To further investigate possible deviations of the non-central chi-square distribution from the empirical distribution of the likelihood ratio test statistic, we compared the estimated density of the likelihood ratio test statistic directly with the non-central chi-square density. The density functions for the likelihood ratio test statistic from a sample of 2,000 Monte Carlo replications were estimated by applying the filtered polynomial method, originally proposed by Elphinstone (1983, 1985) and resuscitated by Heinzmann (2008).⁵

Figure 3 portrays the histograms and estimated density functions for the likelihood ratio test statistics in this experiment. First, it clearly shows that the estimated density functions using the filtered polynomial method capture the characteristics of the corresponding histograms. Second, it demonstrates that the distribution of the likelihood ratio test statistics does not seem to change with increased sample size when $\text{RMSEA} = 0$. Third, it also demonstrates that the distribution of the likelihood ratio test statistic approaches symmetry and the variation of the test statistic increases as the sample size and RMSEA increase, provided that $\text{RMSEA} > 0$.

To enable a further comparison of the estimated actual distribution of the likelihood ratio test statistic with the asymptotic non-central chi-square distribution, Q-Q plots are shown in Figure 4. The ordinate of each graph represents values of the quantile function of the noncentral chi-square distribution, $\chi^2(d, \delta)$. Degrees of freedom d , are

$$d = g(g - 1)q(q - 1)/2 = 1 \cdot 3 \cdot 2/2 = 3$$

throughout. The noncentrality parameter, δ is a function of total sample size, $n_1 + n_2$, and of the RMSEA, ε , given by

$$\delta = d(n_1 + n_2)\varepsilon^2.$$

For example, if the RMSEA is $\varepsilon = 0.05$ and $n_1 = n_2 = 200$, then the noncentrality parameter is $\delta = 3$. The abscissa of the graph represents corresponding values of the quantile function of the actual distribution of the likelihood ratio test statistic. This function was estimated when applying the filtered polynomial method to the sample of 2000 Monte Carlo replications. To avoid displaying inaccurate regions of the Q-Q plot occurring in the tails, it is shown only between the 1st and 99th percentiles of the noncentral chi-square distribution.

Figure 4 suggests that this non-central chi-square approximation to the likelihood ratio test statistic is fully adequate over the ranges of values of the RMSEA and sample sizes considered except possibly when

⁵We thank Dr. Guangjian Zhang for his help with the density plots.

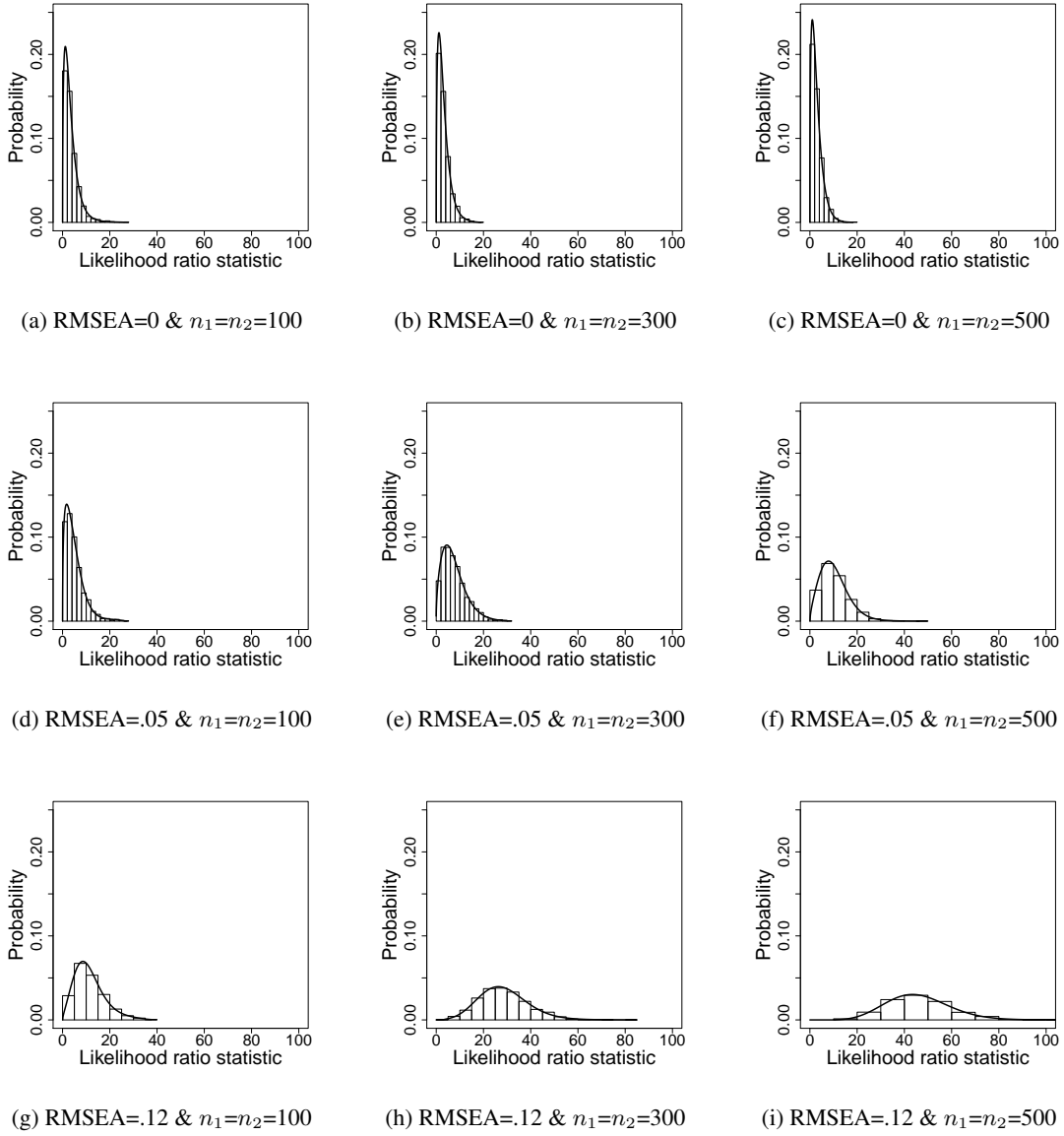


Figure 3: Histograms and estimated density plots for the likelihood ratio test statistics.

$n_1 = n_2 = 100$. The increasing length of the quantile-quantile lines in Figure 4 reflects the increasing variation of the distribution of the likelihood ratio statistic as sample sizes and RMSEA increase, as seen in Figure 3. We conclude from this experiment that the likelihood ratio test can detect both invariance and non-invariance in the higher-order invariance context.

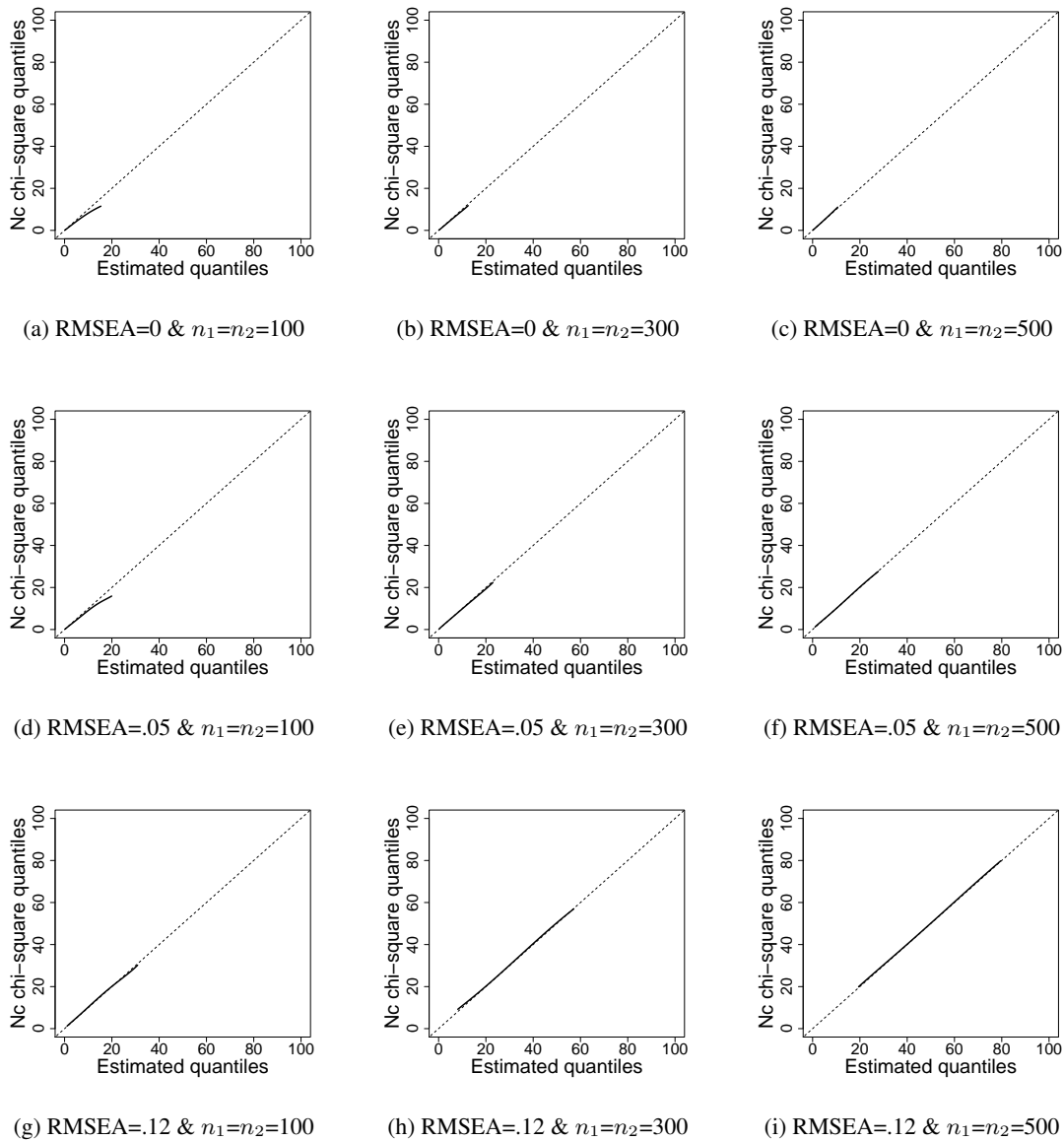


Figure 4: QQ plot. The abscissa and ordinate represent the quantiles from estimation based on the filtered polynomial density function and the non-central chis-square distribution, respectively. The dotted line is the reference line with slope equal to 1.

Discussion and Conclusions

Measurement remains one of the most vital concerns of the behavioral and social sciences. A lack of valid measurement approaches foregoes any hope of empirically testing theoretical constructions that might lead to explanation and thus limits scientific progress to continued descriptive work. Researchers in

behavioral and social science measurement have not been idle and over the past four or five decades we have benefited from extensions of classical test theory (Lord & Novick, 1968), innovations deriving from mathematical psychology (Krantz & Tversky, 1971), the development of item response theory (IRT) (Rasch, 1960), the introduction and subsequent development of generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson, Webb, & Rowley, 1989), and so on. For the past three decades a major point of concentration has been the development and elaboration of the measurement model approach within the framework of structural equation modeling (SEM). This last springs from the common factor analysis model and relies heavily on the rich traditions of factorial invariance to build a measurement scheme that explicitly confronts various reliability, validity, and scaling issues.

Despite the promise and popularity of SEM-based measurement approaches, we have found them wanting when they are evaluated in light of what we believe to be the role of the individual as the unit of analysis in a science of behavior. The essential arguments, which have been presented elsewhere (Nesselroade et al., 2007), and were summarized above, identify a “filtering” role for factors to eliminate idiosyncrasy that is irrelevant, yet present and detrimental to the purpose of establishing lawful relations. This role for factors was referred to as “the idiographic filter” by Nesselroade et al. (2007). A proposed solution, and the one featured herein, is to invoke the logic and methods of factorial invariance at the level of the interrelations among the primary factors while allowing the associated primary factor loadings to reflect idiosyncratic features of individuals or subgroups. This then provides an invariant reference frame for the higher-order constructs while allowing them to project into the measurement space in non-invariant patterns. For example, transformations such as the Schmidt-Leiman and Cattell-White transformations (Loehlin, 1998) can be used to obtain the loadings of observed variables on the higher order factors.

Although the strategy just described seems to be a promising way to nullify unwanted idiosyncrasy while enhancing measurement properties, one obviously cannot be completely “laissez-faire” in mapping factor loading patterns to individuals or subgroups. Even though the procedures constrain the primary factors to intercorrelate identically across individuals or subgroups and thus dictate invariant second order factor patterns (what we have labeled higher-order invariance), for the procedure to yield a convincing solution one wants to have confidence that the somewhat different factor loading patterns still represent the “same” underlying factors. This needs to be more substantial than just a “feeling” and, as was discussed earlier, there are minimum loading pattern conditions to be observed in establishing higher-order invariance. At the same time, however, it was made clear that there is absolutely no requirement that identical variable sets are necessary for establishing factorial invariance, even in the traditional sense of that term.

That individuals and subgroups differ in important ways in how they construe verbal items, in their learning/conditioning histories, etc., seems ample reason for entertaining the possibility that a rigidly standardized measurement framework at the *observable* level may not be the most appropriate and compelling way to proceed with the assessment of abstract constructs. Admittedly, this assertion “flies in the face” of measurement tradition in behavioral and social science but, rather than being rejected outright, the potential value of such an approach needs to be assessed. The simulation studies we conducted lean support to the higher-order invariance conception as a viable and promising option for constructing and evaluating idiosyncratically tailored schemes for measuring abstract constructs. What are some of the more general implications of these proposals and the supporting results that we have reported here? Here are three worth noting.

1. Recognizing and actively nullifying idiosyncrasy that is irrelevant to one’s measurement purpose but which can intrude into the process of empirical inquiry, rather than ignoring it, leaving it to “average out,” is possible and methods and techniques for so doing deserve to be further elaborated and applied, even though they may challenge long-held beliefs regarding “proper” measurement in the social and behavioral sciences.
2. To the extent that idiosyncrasy in standardized measurement frameworks tends to dilute the strength

of relations among important constructs, getting rid of it through filtering operations such as those discussed herein can be expected to result in stronger relations.

3. In the higher-order invariance context, the design of studies can be innovative in building measurement batteries that are better matched to one's experimental aims. For example, longitudinal studies can deliberately plan to mold the nature of the test battery to the age-level of the participants at each occasion of measurement while maintaining their research focus on the same constructs. In cross-sectional studies involving subgroup comparisons based on age, ethnicity, gender, etc., test batteries can be tailored to the subgroup while still measuring the same constructs.

Finally, along with various colleagues we reiterate the importance of re-focusing on the individual as the proper unit of analysis in psychological research (see e.g., Carlson, 1971; Lamiell, 1997; Molenaar, 2004; Nesselroade & Molenaar, 2010). The study of individual differences, despite its popularity and productivity, still needs to work at "reconstituting" a meaningful individual from all those differences that are studied. The higher order invariance conception that we have examined herein is aimed at helping to promote an individual emphasis by offering a more informed basis for aggregating information about individuals in the effort to articulate general relations. Clearly, however, the emphases and the approach extend to comparisons at the subgroup level as well as the individual level.

References

- Baltes, P. B., & Nesselroade, J. R. (1970). Multivariate longitudinal and cross-sectional sequences for analyzing ontogenetic and generational change: A methodological note. *Developmental Psychology*, 2, 163–168.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York : Wiley.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36(1), 111–150.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of ARMA time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Psychometrics: A festschrift to Roderick P. McDonald* (pp. 415–452). Mahwah, NJ: Lawrence Erlbaum Associates.
- Carlson, R. (1971). Where is the person in personality research? *Psychological Bulletin*, 75(3), 203–219.
- Cattell, R. B. (1957). *Personality and motivation structure and measurement*. New York: World Book Co.
- Cattell, R. B. (Ed.). (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurement: Theory of generalizability for scores and profiles*. New York: Wiley.
- Elphinstone, C. D. (1983). A target distribution model for nonparametric density estimation. *Communications in Statistics, A, Theory and Methods*, 12(2), 161–198.
- Elphinstone, C. D. (1985). *A method of distribution and density estimation*. Unpublished ph. d. thesis, University of South Africa, Pretoria, South Africa.
- French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(3), 378–402.
- Heinzmann, D. (2008). A filtered polynomial approach to density estimation. *Computational Statistics*, 23, 343–360.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117–144.
- Jennrich, R. I. (2007). Rotation methods, algorithms, and standard errors. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 315–336). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426.
- Krantz, D. H., & Tversky, A. (1971). Conjoint-measurement analysis of composition rules in psychology. *Psychological Review*, 151–169.
- Lamiell, J. T. (1997). Individuals and the differences between them. In R. Hogan, J. A. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 117–141). New York: Academic Press.
- Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.
- McArdle, J. J. (1982). *Structural equation modeling of an individual system: Preliminary results from "A case study in episodic alcoholism"*. (Unpublished manuscript, Department of Psychology, University of Denver)
- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29, 177–185.
- Meredith, W. (1993). Measurement invariance, factor analysis and factor invariance. *Psychometrika*, 58, 525–543.
- Millsap, R. E., & Meredith, W. (2007). Factorial invariance: Historical perspectives and new problems. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 131–152). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50, 181–202.
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology – this time forever. *Measurement: Interdisciplinary Research and Perspectives*, 2, 201–218.
- Molenaar, P. C. M. (2008, January). Personal communication.
- Nesselroade, J. R. (1970). Application of multivariate strategies to problems of measuring and structuring long-term change. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory* (pp. 193–207). New York: Academic Press.

- Nesselroade, J. R., & Estabrook, C. R. (2008). Factor invariance, measurement, and studying development over the lifespan. In C. K. Hertzog & H. Bosworth (Eds.), *Festschrift for K. W. Schaie*. Washington, D. C.: American Psychological Association.
- Nesselroade, J. R., Gerstorf, D., Hardy, S. A., & Ram, N. (2007). Idiographic filters for psychological constructs. *Measurement: Interdisciplinary Research and Perspectives*, 5, 217–235.
- Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. M. (2002). Alternative dynamic factor models for multivariate time-series analyses. In D. M. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Advances and techniques* (pp. 235–265). Mahwah, NJ: Lawrence Erlbaum Associates.
- Nesselroade, J. R., & Molenaar, P. C. M. (2010). Emphasizing intraindividual variability in the study of development over the lifespan. In W. F. Overton (Ed.), *Handbook of lifespan development*. New York: John Wiley & Sons.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut.
- Reyburn, H., & Raath, M. (1950). Primary factors of personality. *British journal of mathematical and statistical psychology*, 3, 150–158.
- Schmidt, J., & Leiman, J. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 53–61.
- Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist*, 44(2), 922–932.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Thurstone, L. L. (1957). *Multiple factor analysis (fifth impression)*. Chicago: University of Chicago Press.
- Zevon, M., & Tellegen, A. (1982). The structure of mood change: Idiographic/nomothetic analysis. *Journal of Personality and Social Psychology*, 43, 111–122.