Bifactor and Hierarchical Models: Specification, Inference, and Interpretation

Kristian E. Markon

University of Iowa

Preprint of Markon, K. E. (In press). Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annual Review of Clinical Psychology*.

Contact Information: Kristian E. Markon, Department of Psychological and Brain Sciences, University of Iowa, 312 Stuit Hall, Iowa City, IA 52242, kristian-markon@uiowa.edu

ABSTRACT

Bifactor and other hierarchical models have become central to representing and explaining observations in psychopathology, health, and other areas of clinical science, as well as in the behavioral sciences more broadly. This prominence comes after a relatively rapid period of rediscovery, however, and certain of their features remain poorly understood. Here, hierarchical models are compared and contrasted with other models of superordinate structure, with a focus on implications for model comparisons and interpretation. Issues pertaining to the specification and estimation of bifactor and other hierarchical models are reviewed, in exploratory as well as confirmatory modeling scenarios, as are emerging findings about model fit and selection. Bifactor and other hierarchical models provide a powerful mechanism for parsing shared and unique components of variance, but care is required in specifying and making inferences about them.

Keywords: hierarchical, higher-order, bifactor, model equivalence, model complexity

INTRODUCTION

Bifactor models are now ubiquitous in structural modeling of psychopathology. They have been central to general factor models of psychopathology (e.g., Caspi et al. 2014, Laceulle et al. 2015, Lahey et al. 2012, Stochl et al. 2015), and have become a prominent focus in modeling a range of phenomena, as diverse as internalizing psychopathology (Naragon-Gainey et al. 2016), externaling psychopathology (Krueger et al. 2007), psychosis (Shevlin et al. 2017), somatic-related psychopathology (Witthöft et al. 2016), cognitive functioning (Frisby & Beaujean 2015), and constructs central to prominent therapeutic paradigms (Aguado et al. 2015). They have also become central to modeling method effects, such as informant (Bauer et al. 2013), keying (Gu et al. 2017; Tomas & Oliver 1999), and other effects (DeMars 2006), and have been used to explicate fundamental elements of measurement theory (Eid et al. 2017).

Although bifactor and other hierarchical models are now commonplace, this was not always so. Their current ubiquity follows a long period of relative neglect (Reise 2012), having been derived in the early twentieth century (Holzinger & Harman 1938, Holzinger & Swineford 1937), before being somewhat overlooked for a number of decades and then being rediscovered more recently. Bifactor models were mistakenly dismissed as equivalent to and redundant with other superordinate structural models (e.g., Adcock 1964, Humphreys 1981, Wherry 1959, Reise 2012, Yung et al. 1999); as differences between bifactor models and other types of superordinate structural models became more recognized (Yung et al. 1999), interest in bifactor models reemerged.

HIERARCHICAL AND HIGHER-ORDER MODELS: DIFFERENTIATING SUPERORDINATE STRUCTURAL ACCOUNTS

The origins of bifactor models are closely intertwined with the origins of factor analysis itself and theories of general cognitive ability. Spearman, for example, in an attempt to represent Galton's (1883) theories of general cognitive ability, developed a model which distinguished between superordinate general factors, orthogonal subordinate specific factors, and error factors (Figure 1a; Beaujean 2015, Spearman 1904a, b). Although this model as formulated was not estimable, it comprised basic features of bifactor models. As such it laid foundation for the idea that factors could differ in their explanatory breadth, and sparked development of alternative models of subordinate and superordinate factors and their relationships.

Correlational, Higher-Order, and Hierarchical Models

In particular, two approaches to conceptualizing subordinate-superordinate factor relationships developed, especially in the context of understanding general cognitive ability (Beaujean 2015, Carrol & Schweiker 1951). These two paradigms differ in how levels of abstraction are modeled: in one, superordinate factors are at a greater level of abstraction because they influence subordinate factors; in another, superordinate factors are at a greater level of abstraction because they influence a greater breadth of observed variables.

In one paradigm, exemplified by the work of Thurstone (1944), superordinate factors such as general cognitive ability were seen as explaining subordinate factors in a "bottom-up" manner. In this paradigm, subordinate factors were seen as theoretically salient, and superordinate factors such as g were seen representing the tendency for subordinate factors to be correlated (Figure 2a). This "bottom-up" paradigm is represented by the *higher-order* model (Figure 2b), in which general factors explain specific factors, and the latter are nested in the former. Superordinate factors, in this "bottom-up" paradigm, are superordinate because they influence subordinate factors.

In contrast, in the “top-down” paradigm, subordinate factors explained residual phenomena that were not explained by superordinate factors such as g. In this paradigm, superordinate factors such as g were theoretically salient. This paradigm is represented by bifactor and other *hierarchical* models (Figure 1b), where general and specific factors are orthogonal and uncorrelated with one another, explaining distinct nonnested components of shared variance among indicators. Superordinate factors, in this "top-down" paradigm, are superordinate relative to subordinate factors because they influence a greater breadth of observed variables.

Different terminology has been used for different types of hierarchical models, but for the sake of clarity, here I adopt the terminology of Yung and colleagues (1999), who were following an an even older literature. In this tradition, bifactor models are a particular type of hierarchical model where there is one superordinate factor and multiple subordinate factors. Other, non-bifactor hierarchical models might have multiple superordinate factors, or more than two levels of hierarchy, such that the distinction between general (G) and specific (S) factors (Figure 1) is vague. In all hierarchical models, though, factors directly influence observed variables, and not more subordinate factors.

Relationships between higher-order, hierarchical, and correlational models are central to understanding the statistical characteristics of hierarchical and bifactor models and their interpretation. As bifactor and other hierarchical models have been increasingly adopted, questions have emerged about their statistical characteristics and the characteristics of inferential methods used with them, in terms of estimation and model selection accuracy. Questions about their formulation and interpretation have also emerged, with relationships between superordinate structural models playing a key role.

Constraints and Nesting Relationships.

It is important to recognize that the higher-order model is nested in the hierarchical model, as a constrained version of the latter. In a bifactor or other hierarchical model, superordinate factors can have direct associations with observed variables (Figure 1b); in a higher-order model, in contrast, associations between superordinate factors and observed variables are mediated through subordinate factors (Figure 2b). These mediation paths can be thought of as model constraints on the associations between superordinate factors and observed variables, in that in a higher-order model, the association between a superordinate factor and observed variable is proportional to the product of the path between the superordinate factor and subordinate factor and the path between the subordinate factor and observed variable. In a hierarchical model in contrast, a direct association between a superordinate variable and observed variable can be freely estimated, and is not dependent on mediating subpaths.

In general, this nested relationship implies that higher-order and hierarchical models are not equivalent to one another, although hierarchical models with appropriate constraints can be equivalent to higher-order models. A variant of higher-order model where superordinate factors have direct effects on observed variables (Figure 1e) is equivalent to a hierarchical model (Young et al. 1999), although this type of higher-order model is less common.

Reise (2012) articulated how different factor models of interest are related to bifactor and other hierarchical models. For example, by fixing superordinate factor loadings to zero and freeing factor correlations, one obtains a correlated-factors model. By imposing appropriate constraints on the superordinate-to-observed variable loadings of a bifactor model, to represent them in terms of subordinate factors, one obtains a higher-order model. Finally, by fixing the subordinate factor model loadings to zero, one obtains a unidimensional factor model.

These constraints are critical to interpreting phenomena involving bifactor models, such as the relative fit of bifactor models compared to to other models. In contrast to many model comparison scenarios, where constraints are placed by fixing parameter estimates to particular values (such as fixing a path to zero), with bifactor model comparisons, the constraints involved often are implicit, through structural differences between the models. These constraints also have implications for interpreting different methods for estimating bifactor models, as those different estimation methods may imply different constraints.

MODEL SPECIFICATION AND INTERPRETATION

Unique challenges arise in specification of bifactor and other hierarchical models due to their scope and overlap with other model classes. In exploratory factor analysis (EFA) bifactor models pose challenges for specification and rotation of loading matrices, because they are somewhat quasi-confirmatory, necessitating *a priori* specified loading patterns in the superordinate loadings, but not in the subordinate loadings. Similarly, with both exploratory and confirmatory factor analysis (CFA), bifactor models raise special issues with regard to model specification and identification due to their equivalence with various classes of models. Finally, interpretation of models and parameter estimates can be challenging or unintuitive due to the relativistic meaning of structural features, and qualitative changes in model interpretation with changes in estimates.

Exploratory Factor Analysis: Transformations and Rotations

Rank-Deficient Transformations: Schmid-Leiman and Related Methods. The Schmid-Leiman family of methods (referred to here as the rank-deficient transformation [RDT] family of methods, following Waller 2017, Schmid & Leiman 1957; also independently discussed by Thompson 1939, Thurstone

1947, and Wherry and Winer [Wherry 1959, Wherry & Winer 1953]) comprise an older exploratory approach to hierarchical modeling. These methods now take different forms, but all re-express an exploratory factor model, from one having fewer factors with higher-order structure (that is, factor correlations or higher-order factors) to one having more factors without higher-order structure (a bifactor model). RDT methods can be thought of as a way to re-express a correlational or higher-order model in a bifactor form, and illustrate the nested relationship between the two types of models.

An early form of RDT, the Schmid-Leiman transformation (SLT), works by re-estimating a correlated-factors model as a higher-order factor model, re-expressing the factor correlations of that model in terms of higher-order factor loadings. The indirect effects of superordinate factors on observed variables, via higher-order and lower-order path combinations, are then re-expressed as direct superordinate-observed variable paths of a bifactor model. In the SLT, an EFA model having p lower-order factors is rotated using an oblique independent clusters rotation, and a higher-order factor model is estimated based on the factor correlations from this oblique rotation. Once the higher-order paths have been estimated, they can be multiplied by appropriate lower-order paths to obtain estimates of the implied direct associations between the superordinate variable and observed variables as might be obtained in a bifactor model. This results in a matrix of loadings on the observed variables, including the loadings of superordinate factors on the observed variables, and of subordinate factors on the observed variables. What is obtained is a set of loadings which structurally resembles a bifactor model, but, because it was obtained through a higher-order framework, is a bifactor model with those constraints, and actually a re-representation of a higher-order model.

Waller (2017) has provided an alternate approach to RDT, one which differs from the original SLT in that it involves a single factor analysis. This approach, referred to as a direct SLT, is actually a form of rotation; rather than transform the parameters of a higher-order factor model as in the original SLT, in a direct SLT one performs a targeted rotation of an augmented loading matrix from an oblique factor analysis model. Like the original SLT, however, direct SLT results in a loading matrix that instantiates a more constrained higher-order or correlational model while resembling a bifactor or other hierarchical model.

Waller's (2017) work on RDT is important because it illustrates directly how nested relationships between higher-order and hierarchical models can change expression. In the direct SLT, one rotates an oblique loading matrix from a p-factor model where a column of zeros have been added—that is, one rotates a (p+1)-column matrix, but where one column consists of zeros. After rotation, the loading matrix takes a hierarchical structural form. In this way, one can think of a higher-order structure as either a constrained version of a hierarchical structure, or one where some parameters have been fixed to zero in a specific way. The form of the constraints can change, even though the underlying structural nesting is the same.

Analytic Bifactor Rotations. Analytic bifactor rotations (Jennrich & Bentler 2011, 2012) provide an alternative exploratory approach to bifactor structure. In this approach, an EFA loading matrix is explicitly rotated to a bifactor criterion. Unlike the SLT, which involves re-representing estimates from an oblique rotation of a p-factor model, where p is the number of lower-order factors, in an analytic bifactor rotation, the loading matrix of a (p+1)-factor EFA model, comprising subordinate factors and one superordinate factor, is rotated directly.

Analytic bifactor rotations work by rotating the entire (p+1)-column loading matrix such that the p-column subordinate loading submatrix is rotated to an independent clusters criterion. Different orthogonal or oblique independent clusters rotations could be used, but commonly studied formulations have focused on quartimin and geomin rotations (through the analytic bifactor biquartimin and bi-geomin rotations; Jennrich and Bentler 2011, 2012). Mansolf and Reise (2016) have noted that rotation of the superordinate factor loadings is indirect and implicit, in contrast to rotation of the subordinate factor loadings, which is direct and explicit in that it forms the basis for the rotation criterion.

Targeted Bifactor Rotations. Another approach to delineating hierarchical structure using exploratory factor analysis is to use a bifactor model as a target in a procrustes (that is, targeted) rotation. In contrast to analytic bifactor rotations, where an observed loading matrix is rotated to directly optimize a mathematical function of the observed loadings themselves, in procrustes rotation, a loading matrix is rotated to be maximally similar to some second target loading matrix. By choosing a target matrix having a bifactor structure, one can rotate the observed loading matrix to a bifactor form.

With bifactor target rotations, a paradox arises in that a bifactor target is needed in order to delineate a bifactor structure, diminishing the advantages of an exploratory approach. One solution, suggested by Reise and colleagues (Reise et al. 2010), is to use a partially specified target loading matrix, where some of the loadings are used as a target but not others. Loadings that are more theoretically central, or known with greater confidence, might be used as targets, for example, such as the superordinate factor loadings or a subset of the subordinate factor loadings.

Another solution, proposed by Abad and colleagues (2017), is to use a SLT loading matrix as a target. Although the SLT loading matrix actually instantiates a higher-order model, by rotating a matrix based on a (p+1)-factor model, rather than the p-factor model of the SLT, one retains the information in the (p+1)-factor model, resulting in a hierarchical model that is as similar as possible to that suggested by the SLT. Abad and colleagues (2017) further suggest using this procrustes-rotated matrix as a target itself in a subsequent step, and iterating this process repeatedly until the rotated loading matrix stabilizes. Note that this differs subtly but critically from the direct SLT proposed by Waller (2017): whereas Abad and colleagues' method rotates a (p+1)-factor loading matrix using a pseudo-(p+1)-factor target (that is, a target that appears to have p+1 factors, but only has the statistical information of p factors) Waller's method rotates a pseudo-(p+1)-factor loading matrix using a pseudo-(p+1)-factor target.

Both analytic bifactor rotations and the SLT are susceptible to bias and other problems in estimation of higher-order factor loadings when there are cross loadings at the lower-order level (Abad et al. 2017, Reise 2012). When there are cross-loadings at the lower-order level, some of the covariance implied by those cross-loadings is accounted for by higher-order factors, leading to overestimated higher-order loadings and underestimated lower-level loadings (Reise 2012). This leads to potential problems with use of exploratory methods in scenarios where independent clusters structure is only approximate. Abad and colleagues (2017) provide some simulation evidence that using a SLT loading matrix as a rotation target may ameliorate some of these problems, but further research is needed to confirm their findings under a broader range of conditions.

Confirmatory Factor Analysis

The complexity of relationships between different superordinate exploratory factor models reflects broader complexities involved in specification and identification of hierarchical models, including those in a confirmatory context. A path parameter in a hierarchical model might change in meaning or structural importance in unintuitive ways, leading to complexities in interpretation, model identification, and estimability of models.

Empirical Underidentification. One challenge in confirmatory bifactor modeling is empirical underidentification, where models might be underidentified and unestimable or not depending on sample. Green and Yang (2017) highlighted the complexity of such issues with confirmatory bifactor models, demonstrating that bifactor models can be underidentified in samples with what they term a homogenous within and homogenous between (HWHB) covariance structure. In this type of structure, covariances are approximately equal between observed variables measuring the same subordinate factor, and are also approximately equal between observed variables measuring different subordinate factors—that is, the block of covariances corresponding to one subordinate factor (σ_{ii}) are approximately equal (to one another, but not necessarily to those in another block), the block of covariances corresponding to another factor are approximately equal (σ_{jj}), and the block of cross-covariances are approximately equal (σ_{ij}). In this case, one has

$$\begin{aligned} \sigma_{ii} &= \lambda_{Gi}^2 + \lambda_{Si}^2 \\ \sigma_{jj} &= \lambda_{Gj}^2 + \lambda_{Sj}^2 \\ \sigma_{ij} &= \lambda_{Gi}\lambda_{Gj} \end{aligned} \quad (1)$$

where λ_{Gi} is the loading of a measure in block *i* from the general factor, λ_{Gj} is the loading of a measure in block *j* from the general factor, λ_{Si} is the loading of a measure in block i from its specific factor, and λ_{Sj} is the loading of a measure in block *j* from its specific factor.

As Green and Yang (2017) discuss, in this HWHB scenario, there are fewer observed covariances than parameters, leading to underspecification of models and difficulties with estimation. Whether or not such scenarios are typical of psychopathology data is unclear, although it should be noted that such scenarios would be predicted under certain measurement circumstances sometimes deemed ideal (for example, with Rasch-conforming measures, where loadings are equal), or when covariances are uniformly low or high. In these cases, additional constraints may be necessary to identify the model, such as enforcing additional assumptions about equality of loadings.

Factor Correlations in Hierarchical Models. One specific recurring question about model identification is whether bifactor and hierarchical models can or should include correlations among the factors. The standard hierarchical model constrains factor correlations to zero, consistent with the factors being interpreted as residual influences on observed variables controlling for other influences. Nevertheless, various papers in the psychopathology literature have reported results of correlations among the factors, especially among subordinate factors, raising questions about what is identified and how to interpret bifactor models with factor intercorrelations.

Caspi et al. (2014), for example, examined three models of psychopathology, including one bifactor model with correlated subordinate internalizing, externalizing, and thought disorder factors in addition to a superordinate general psychopathology factor. This bifactor model produced inadmissible estimates, suggesting it was not specified correctly, so the authors modified it to eliminate the subordinate thought disorder factor, leaving correlated internalizing and externalizing factors and a general psychopathology factor. Although estimates converged in the overall dataset, producing an estimated negative correlation between internalizing and externalizing, they do not converge in all the age subsamples (within the 38 year-old data they report, for example, negative observed variable residual variances are produced).

The difficulties with convergence encountered by Caspi et al. (2014) are common when factors are allowed to correlate in hierarchical models (Eid et al. 2017), pointing to problems with identification in doing so. Variation in convergence across subsamples also suggests the possibility of empirical underidentification problems, where model identification depends on the sample. Extending the rationale of Green and Yang (2017) from Equation 1, with subordinate factor correlations, one would have

$$\begin{aligned}\sigma_{ii} &= \lambda_{Gi}^2 + \lambda_{Si}^2 \\ \sigma_{jj} &= \lambda_{Gj}^2 + \lambda_{Sj}^2 \\ \sigma_{ij} &= \lambda_{Gi}\lambda_{Gj} + r_S\lambda_{Si}\lambda_{Sj}\end{aligned} \tag{2}$$

where r_S is the correlation between the subordinate factors. This again leads to a scenario where there are potentially more parameters than degrees of freedom to estimate them. When factors are allowed to correlate, even if the within-factor covariances are heterogeneous to some extent, difficulties with model identification might arise if the between-factor covariances are relatively homogeneous (or vice versa). This might occur, for example, if the covariances are uniformly small.

Even if a hierarchical model with factor correlations is identified in a particular sample, interpretive questions can be raised. For example, what does it mean for internalizing and externalizing to be negatively correlated, controlling for a general tendency of disorders to be positively correlated, including internalizing and externalizing disorders? Is this ultimately substantively different from saying that internalizing and externalizing disorders are less correlated with one another than with other forms of psychopathology? Parameters in hierarchical models generally pertain to residualized effects, which complicate interpretation of correlations relative to intuitive conclusions.

Eid and colleagues (2017) explored factor correlations in hierarchical models, including correlations between general and specific factors as well as correlations between different specific factors. They generally concluded that correlations between general and specific factors were theoretically inadmissible, as these would violate interpretations of specific factors as residual effects on variables net the effects of general factors. In contrast, they concluded that, all other things being equal, correlations between specific factors could be admissible, as a form of partial correlation between subordinate variables controlling for the effect of a dominant superordinate factor. However, the bifactor models they considered included loadings from superordinate factors that did not also have loadings from subordinate markers, allowing for estimation of general factor covariance independent of specific factor covariance. It is unclear how subordinate factor correlations could be estimated in the absence of loadings unique to superordinate factors.

Returning to the example of Caspi et al. (2014), it is important to note that, by eliminating a subordinate thought disorder factor, the authors essentially created factor loadings unique to the superordinate general psychopathology factor. That is, in one model, with subordinate internalizing, externalizing, and thought disorder factors, all diagnoses had loadings from subordinate as well as superordinate factors, rendering correlations between subordinate factors unestimable. With the thought disorder factor removed from the model, the corresponding diagnoses (obsessive-compulsive disorder, mania, and schizophrenia) became indicators of the superordinate general factor alone. In that case, the correlation between internalizing and externalizing represents residual relationships between the common mental disorders, independent of internalizing and externalizing and the less common mental disorders.

Considered overall, the literature suggests that estimation of factor correlations in hierarchical and bifactor models depends on the nature of the model being estimated as well as the particular sample of data at hand. General-specific factor correlations are likely inadmissible regardless of scenario. Specific-specific factor correlations may be estimable if the model includes superordinate-limited indicators, and if the data permits estimation of the correlations. If indicators all cross-load between superordinate and subordinate factors— either because they are specified that way or are estimated that way—or if covariances among variables are too homogenous with reference to the factor structure being hypothesized, correlations between specific factors may be inadmissible.

Loadings in Hierarchical Models: Misspecification and the Nature of Factors. Possible interpretive ambiguities extend further when loadings are considered. Interpretation of hierarchical models as such depends on a pattern of zero and nonzero loadings where factors differ substantively in their breadth of associations with observed variables. Although the choice of model specification or rotation may imply this pattern, it does does not guarantee loading estimates will conform to it: a model can be specified as a bifactor model but not estimated as one. This complicates interpretation of hierarchical models, in that a model specified as hierarchical or bifactor might be most appropriately interpreted in a different way, such as a simple structure model.

Eid and colleagues (2017) provide a discussion of types of anomalous estimates that are commonly encountered with hierarchical and bifactor models. Many of these types of anomalous estimates involve loadings that are more circumscribed than intended by the model specification. For example, factors intended to be superordinate in nature might have a number of small, even zero loadings, so that factors specified to be general in nature may be estimated as relatively specific (for example, a putative general cognitive ability factor might have estimated loadings near zero except for those associated with verbal working and short-term memory measures, raising questions about whether or not the factor is actually general if it does not also encompass other types of measures). Factors intended to be subordinate, similarly, can be even more specific in the nature of their estimated loadings than intended (for example, in a study of psychosis, a factor intended to reflect positive symptoms might only have prominent loadings on measures of delusions of reference, but not on other types of delusions or hallucinations). Closely related scenarios also occur where a subordinate factor is specified as part of a best-fitting model, but the estimates imply that the subordinate factor accounts for essentially no variance in measures (Eid et al. 2017)—a scenario analogous to a best-fitting model with a factor not having any substantial loadings.

Combinations of anomalous estimates can raise questions about how to interpret models intended to be hierarchical or bifactor in nature. For example, when factors intended to be superordinate in nature have a number of small loadings, in combination with subordinate factors that are more specific than intended, this can suggest a misspecified bifactor model, raising questions about whether a bifactor-specified model is better interpreted as a simple structure model. Patalay and colleagues (2015), for example, presented an estimated bifactor model of psychopathology in adolescence, specified in terms of putative general psychopathology, internalizing, and externalizing factors. Although the putative general factor did load on all of the variables examined, the loadings on the externalizing variables were relatively small, and the larger loadings were on measures reflecting negative emotionality, especially markers of dysregulated negative emotion (e.g., anger, sleep difficulty, tearfulness). The putative internalizing specific factor, moreover, comprised prominent loadings almost exclusively reflecting anxiety and worry. In this way, even though the model was specified as bifactor in structure, questions might be raised about whether the pattern of estimates is best interpreted as such—whether factors of the model should be interpreted as general psychopathology and internalizing factors, or emotional dysregulation and anxiety factors. These types of scenarios create challenges of interpretation even when a confirmatory model is being specified.

Another example of difficulties interpreting level of breadth of factors occurs when loadings from a putative general factor are extremely large, to the point of being almost perfect—for example, close to 1 on a standardized scale. Even if other loadings from a general factor are not small, if one of the loadings is close to 1 (for example, 0.95 or greater; e.g., Caspi et al. 2014, Laceulle et al. 2015, Stochl et al. 2015), it suggests that the measure is interchangeable with the factor or nearly so. Although it is possible to interpret this in terms of a very strong measure of the factor, it is also plausible that the variable being reflected by the high-loading measure in question has high direct associations with other measured variables, and that the whole system has been misspecified in some way.

Explanations for Anomalous Estimates and Alternative Hierarchical Model Specifications. Various explanations can be offered for anomalous factor loading estimates with hierarchical and bifactor models. When models are misspecified, estimates will tend to converge to those that bring the predicted data likelihood closest to the data likelihood under the unknown true model (assuming maximum likelihood estimation has been used; Lv & Liu 2014, White 1982). If the specified model is relatively flexible relative to the true model, but incorrect—such as when the true model is nested in a misspecified model—the misspecified estimates would come to resemble the true model as much as is possible given model constraints. This could explain why estimates sometimes deviate from an intended hierarchical or bifactor structure: if a simple structure true model were misspecified as a bifactor model, for example, the estimated loadings might come to resemble that of a simple structure model. The general, superordinate factor loadings might then resemble those of a subordinate factor, and estimated subordinate factor loadings might be even more specific than expected.

Eid and colleagues (2017) provide an another explication of bifactor models misspecification, based on how subordinate factors are theoretically related to superordinate factors. Arguing from the perspective of stochastic measurement theory, a framework related to classical test theory and generalizability theory, Eid and colleagues assert that traditional bifactor models can only be appropriately applied when subordinate factors are conceptually nested within superordinate factors. This would occur, for example, if subordinate factors represented samples from subordinate factor domains (such when clinician ratings are sampled randomly from the domain of possible clinicians). Because bifactor models are often applied when subordinate and superordinate factors do not have this relationship, Eid and colleagues (2017) assert that bifactor models are often being misspecified or misapplied in the literature.

Eid and colleagues (2017) propose alternate model specifications for situations in which subordinate and superordinate factors are not nested. These alternate specifications, the bifactor-(S-1) model and bifactor-(SI-1) model, establish one factor as a reference factor against which the subordinate factors are compared. The (S-1) and (SI-1) models differ in how reference factors or domains are defined vis-a-vis gold standard measures, as well as in the number of specific factors, but both define one factor as a reference domain that is fixed relative to the other factors. The authors argue that these specifications are more appropriate than traditional bifactor specifications for the structural modeling problems that are typically encountered in psychopathology research.

MODEL EVALUATION AND COMPARISON

As interest in hierarchical and bifactor models of psychopathology has increased, they have become a standard comparator in structural modeling, especially in questions involving complex, multidimensional clinical constructs. This has reinforced the need for accurate inferences about the fit of different hierarchical models—relative to one another, relative to other types of models, and relative to the data—as well as an increased interest in methods of model evaluation and selection with these types of structures.

Repeated observation of anomalous estimates, such as unexpected loading patterns, in hierarchical models that otherwise appear to fit relatively well has led to questions about their selection and evaluation. If a model specified as hierarchical or bifactor produces estimates that are similar to a simple structure model, for example, why does the simple structure model not fit as well, or even better if parsimony is accounted for, especially given that it would be more constrained and therefore have the advantage of using fewer degrees of freedom to attain similar estimates?

Recent work by various groups has illustrated that hierarchical models are prone to positive bias in model selection, in that typical values of model fit statistics in observed datasets overstate the models' fit at the population level (Bonifay & Cai 2017, Murray & Johnson 2013). Hierarchical and bifactor models tend to overfit to data because they are relatively flexible, such that they are able to fit to datasets relatively well (in the sense of producing higher likelihoods, or lower residuals) in general regardless of population structure or theoretical plausibility (Bonifay & Cai 2017; Reise et al. 2016). This flexibility is accounted for by in part by their greater number of parameters (compared to simple structure models or higher-order models), but not completely, so that many traditional model selection criteria fail to adjust for the bias, depending on sample size and the models being compared.

Bonifay and colleagues (Bonifay & Cai 2017) have shown that hierarchical and bifactor confirmatory models are similar to exploratory factor models in their level of flexibility, and more flexible than non-hierarchical confirmatory factor models, even when the number of parameters is held constant. On average, a given bifactor confirmatory model will tend to fit to any given random dataset better than a non-hierarchical confirmatory model, and almost, but not quite as well as, an exploratory factor model. This is independent of the population model, and reflects the ability of hierarchical models to capture chance features of any given sample that might not replicate. In the sense of fit, confirmatory hierarchical and bifactor models are similar to exploratory factor models in their expansiveness.

Figure 3 illustrates the tendency of bifactor models to fit better than other models of superordinate structure. 1000 simulated data samples were randomly generated under one of three conditions: a correlated simple structure population model, a bifactor population model, or an unstructured condition with essentially no model. The two structured population models were taken from estimates reported by Caspi and colleagues (2014); in the unstructured condition, data were uniformly sampled from ranges matching those of the data used in that paper, so as to randomly sample from all possible datasets. Figure 3 provides densities of the difference in log-likelihoods under a correlated simple structure model and a bifactor model fit to each sample.

As is illustrated by the figure, the bifactor model fits better in terms of log-likelihood under all three conditions. The bifactor model fits best on average when the population has a bifactor structure, but it also fits better when the population is a simple structure model. Most importantly perhaps, it also fits better than a correlated factors model when data are sampled without regard to any structure. This tendency of the bifactor model to flexibly fit any dataset better is related to the complexity of the model structure. Although in this case, the number of parameters can be used to correct for the overfitting of bifactor models (using traditional information-theoretic criteria such as the Aikaike Information Criterion [AIC] or Bayesian Information Criterion [BIC]), other research has shown this is not always the case (Bonifay & Cai 2017, Murray & Johnson 2013). For smaller samples and models similar in the number of parameters, the structural complexity of the model must also be accounted for (Bonifay & Cai 2017, Markon & Krueger 2004, Murray & Johnson 2013).

Numerous authors have noted that the greater flexibility of bifactor and other hierarchical models can be explained by means of their relationship with higher-order models. As was noted earlier, higher-order models can be reexpressed as a constrained form of hierarchical model, where direct loadings of superordinate factors on observed variables in the hierarchical model are constrained to be proportional to a product of subordinate and superordinate factor loadings under the higher-order model. These proportionality constraints (Gignac, 2016) illustrate how a parameter in one model can have different fit value from a parameter in another model: a loading in the hierarchical model is less constrained, and more flexible, than a loading in the higher-order model. In this regard, parameters are not all equal; the number of parameters per se, which forms the basis of most commonly used model selection statistics, can provide an incomplete summary of model flexibility.

Research on information-theoretic inference has led to the development of ways of quantifying model flexibility above and beyond number of parameters per se. Work on minimum description length (MDL) and normalized maximum likelihood (NML) in particular has shown that a model's complexity can be theoretically quantified in terms of how well a model fits data in general, considered over all possible datasets (Grunwald 2007, Rissanen 2007). The NML statistic, for example, quantifies a model's flexibility as the sum of the model's optimized likelihoods, taken over all possible datasets; the NML itself is the maximum likelihood in the observed sample divided by this sum. Intuitively, this is sensible: a model that is more flexible will tend to fit any dataset well, regardless of truth or plausibility, and will produce greater likelihoods on average, leading to a larger likelihood sum. NML therefore "corrects" an observed sample likelihood by its tendency to produce large likelihoods in any dataset in general. NML is closely related to BIC and other Bayesian fit indices (Grunwald 2007, Rissanen 2007), and can be approximated in various ways (Grunwald 2007, Nalisnick & Smyth 2017). Although not as well-researched, variants of NML for use with more other types of estimators (e.g., least squares estimators) have also been explored, and similar principles emerge (Rissanen 2003).

HIERARCHICAL MODELS, COVARIATES, AND PREDICTION

As hierarchical and bifactor models have become more ubiquitous in the literature, they have been incorporated more frequently into various longitudinal and predictive analyses, as well as models incorporating significant covariates (e.g., Greene & Eaton 2017, Kim & Eaton 2015, Martel et al. 2017, McLarnon et al. 2016). These analyses arise from the need to validate constructs suggested by bifactor models (e.g., Kim & Eaton 2015, Martel et al. 2017), as well as from their use in variance decomposition, such as in multi-informant measurement scenarios (Bauer et al. 2013).

Hierarchical and bifactor structures pose unique challenges in longitudinal and covariate modeling due to factors in these types of models having overlapping indicators and forced orthogonality. Exchangeability between factor correlations and cross-loadings can mask implied correlations between predictors, for example, which might have implications for collinearity-like effects in prediction, where uncertainty about specific effects, independent of general effects, is increased. Adding covariates, similarly, can implicitly induce correlations between factors that might deidentify hierarchical and bifactor models (Koch et al. 2017).

Koch and colleagues (Koch et al. 2017) have shown that treating factors of a bifactor model as an explanandum—that is, as being predicted by covariates—can have significant implications for identification and bias. For example, in the case where specific and general factors are both being explained by a set of common covariates, models need to be constructed in such a way as to ensure identification, as the covariates can imply correlations between the general and specific factors which may lead to misidentification problems. As already noted, subordinate factors of a bifactor model are best interpeted as residual factors relative to the superordinate variables; if they share covariates, this interpretation may become problematic. Even if in a bifactor measurement model different factors are treated as orthogonal, once covariates are introduced, those covariates might imply correlations between the factors that in turn might cause problems for model identification.

Other constraints are required even when specific factors alone are being modeled with covariates. Koch and colleagues (2017) also note that various parameter estimates, such as those of intercepts, loadings, and regression paths, can be biased depending on the nature of the bifactor model when exogenous covariates are present. Solutions are available depending on the scenario (Koch et al. 2017), but the issues involved are not necessarily intuitive and require care to address.

One relatively understudied question is the impact of using factors of a bifactor model as explanans, such as in predicting outcomes from bifactor variables. The complicating effects of multicollinearity on regression inferences are well-known, in that it increases uncertainty of coefficient estimates, as reflected in increased standard errors and associated loss of power. Bifactor structures raise similar issues with regard to prediction: in both cases there is a challenge of making inferences about specific residual effects net some dominant shared variance among the predictors. In the case of multicollinearity, the task is to isolate unique residual effects of each predictor; with hierarchical models, the task is to isolate residual effects that are nonetheless still shared among some subset of the predictors.

In some sense, the orthogonality of shared and residual factors in hierarchical models obscures potential effects on inferences about predictive and covariate effects. Although the factors are orthogonal, this orthogonality is forced, depending on a specific set of bifactor loadings, which are fallibly estimated and possibly even variant across time or subpopulations. To the extent that the loading assumptions are incorrect, unmodeled correlations between predictors might be induced, affecting conclusions about covariates and prediction depending on the analytic scenario.

Gonzalez and MacKinnon (2018) examined power, Type I error, and bias in the case of models with mediating subordinate bifactor variables, where specific factors have indirect predictive effects. In their simulations power to detect mediation increased, and bias in path estimates decreased, as specific and general factor loadings increased. Type I errors were generally low across conditions. However, the authors did not examine cases where general factors might have mediating effects in addition to, or instead of, the specific factors, so it is unclear how the simultaneous presence of the two types of effects affects inferences about each. Presumably when superordinate and subordinate factors each have independent mediating effects, disentangling them would be more challenging. Research on these types of questions is needed as interest in the criterion validity of bifactor variables increases, especially if unique associations of subordinate factors with covariates are not always evident (Martel et al. 2017).

CONCLUSIONS AND RECOMMENDATIONS

Although hierarchical and related models have been in development for over a century (Beaujean 2015, Reise 2012) and have become ubiquitous in contemporary modeling, in many ways their properties are only beginning to be understood. Long dismissed as equivalent to other superordinate models, hierarchical and bifactor models are now recognized as distinct and having a broad range of applications and utility. These models provide decomposition of observed variance into relatively independent components at various levels of abstraction, from the general to the specific, which is fundamental to statistical modeling across many domains.

Ensuring that a model is specified as intended, and interpreting a model as specified, is fundamental to any modeling endeavor. However, the history of hierarchical and bifactor modeling is clouded by misunderstandings about model equivalences and the types of parameter estimates, exactly, that are permissible in such models. These issues are further complicated by a surfeit of different EFA and CFA approaches to hierarchical model specification and estimation—some of which have only relatively recently been developed—as well as by issues such as empirical underidentification, where model estimability may be sample-dependent. As such, it is especially important for modelers to clarify whether or not a given bifactor structure is identified and specified as intended. A number of excellent resources on hierarchical model specification and identification are available (Eid et al. 2017, Green & Yang 2017, Koch et al. 2018), although further research in this area is needed, especially with hierarchical models including covariates and prediction.

Overfitting of hierarchical and bifactor models is a critical issue that has broad implications for modeling in general, well beyond hierarchical models per se. Emerging research in this area (Bonifay & Cai 2017), together with existing literature (Grunwald 2007, Rissanen 2007), indicates that the number of parameters alone cannot always be relied on as an index of model flexibility in model fit and selection statistics, as model structure impacts flexibility as well. Bifactor and other hierarchical models are very nearly as flexible as exploratory factor models, which are saturated in their parameters (Bonifay & Cai 2017). As a result, bifactor models tend to overfit. Researchers should therefore exercise care in interpreting better fit of a confirmatory bifactor model relative to another model, as bifactor models on average will tend to fit any dataset better than other confirmatory models regardless of the population "true" model.

Developments in model fit and selection are needed to adjust for bifactor model overfitting, as traditional methods based on the number of parameters alone are unlikely to entirely correct for the magnitude of overfitting. Promising alternative model selection statistics have been developed (Grunwald 2007, Rissanen 2007); although some of these are relatively challenging to compute and involve novel methods, others are more easily computable with quantities that are already routinely provided by current software (Bollen et al. 2014, Markon & Krueger 2004). Further research is needed to further develop such statistics and to explicate their performance characteristics with different types of models.

One practical approach to the overfitting problem is to review estimates obtained from bifactor model analysis, to consider whether they are consistent with the structure intended, or might be more consistent with another type of structure. The relative flexibility of bifactor models not only increases their fit in terms of likelihood, but also enables bifactor specifications to capture other types of structural scenarios in terms of the estimates themselves. A presumed bifactor model might be this in name only, and might be better characterized as a simple structure correlated factors model or other type of structure. It is important for researchers and consumers of research to consider the possibility that even if a model is specified as bifactor in nature, and fits better when adjusted for parametric complexity, a different model might be more verdical, and the parameter estimates obtained might reflect that model. In an important sense, confirmatory bifactor models are "quasi-exploratory" in nature, and should be approached as such.

SUMMARY POINTS

1. Bifactor and other hierarchical models represent superordinate structure in terms of orthogonal general and specific factors representing distinct nonnested components of shared variance among indicators. This contrasts with higher-order models, which represent superordinate structure in terms of specific factors that are nested in general factors, and correlated factors models, which represent superordinate structure in terms of correlations among subordinate factors.
2. Higher-order models can be approached as a constrained form of hierarchical model, where direct relationships between superordinate factors and observed variables in the hierarchical model are constrained to equal products of superordinate-subordinate paths and subordinate-observed variable paths.
3. Multiple exploratory factor analytic approaches to delineation of hierarchical structure are available, including rank-deficient transformations, analytic rotations, and targeted rotations. Among other things, these transformations and rotations differ in the number of factors being rotated, the nature of those factors, and how superordinate factor structures are approximated.
4. Misspecification or underspecification of confirmatory bifactor and hierarchical models can occur for multiple reasons. Problems with model identification may occur with specific patterns of homogeneity in estimated or observed covariances; if factors are allowed to correlate in inadmissible ways; or if covariate paths imply inadmissable correlations. Signs of model misspecification may be evident in anomalous estimates, such as loading estimates near boundaries or estimates that are suggestive of other types of models.
5. Common model fit statistics can overstate the fit of bifactor models, due to the tendency of bifactor and other hierarchical models to overfit to data in general, regardless of plausibility or population structure. Hierarchical models are similar to exploratory factor models in their expansiveness of fit,

and more expansive in fit than other confirmatory models, in general.

FUTURE ISSUES

1. Research is needed to determine how to best account for the flexibility of hierarchical models when comparing models and evaluating model fit, given that the relative flexibility of hierarchical models can only partly be accounted for by the number of parameters. Approaches based on minimum description length and related paradigms, such as Bayesian inference with reference priors, are promising in this regard.

2. More research is needed to clarify the properties of hierarchical structures when they are embedded in longitudinal models and models with covariates. As with challenges of multicollinearity in regression, parsing unique general and specific factor components of explanatory paths may be inferentially challenging in the presence of strongly related predictors, covariates, and outcomes.

3. More can be learned about specification and identification of hierarchical models, and the relationships between hierarchical models and other types of models such as exploratory factor models. Similarities in overfitting patterns between exploratory and hierarchical models, approaches to hierarchical structure through bifactor rotations, and patterns of anomalous estimates that are sometimes obtained with hierarchical models, point to significant relationships beween exploratory and hierarchical models. Further explication of model specification principles with hierarchical models would also help clarify appropriate structures to consider when evaluating models.

LITERATURE CITED

Abad FJ, Garcia-Garzon E, Garrido LE, Barrada JR. 2017. Iteration of partially specified target matrices: Application to the bi-factor case. *Multivar. Beh. Res.* 52:416-429.

Adcock CJ. 1964. Higher-order factors. *Br. J. Stat. Psychol.* 17:153-160.

Aguado J, Luciano JV, Cebolla A, Serrano-Blanco A, Soler J, Garcia-Campayo J. 2015. Bifactor analysis and construct validity of the Five Facet Mindfulness Questionnaire (FFMQ) in non-clinical Spanish samples. *Front. Psychol.* 6:404. doi:10.3389/fpsyg.2015.00404

Bauer DJ, Howard AL, Baldasaro RE, Curran PJ, Hussong AM. 2013. A tri-factor model for integrating ratings across multiple informants. *Psychol. Methods.* 18:475-493.

Beaujean AA. 2015. John Carroll's views on intelligence: Bi-factor vs. higher-order models. *J Intell.* 3:121-36.

Bollen KA, Harden JJ, Ray S, Zavisca J. 2014. BIC and alternative Bayesian information criteria in the selection of structural equation models. *Struct. Equ. Model.* 21:1-19.

Bonifay W, Cai L. 2017. On the complexity of item response theory models. *Multivar. Beh. Res*. 52:465-84.

Carroll JB, Schweiker RF. 1951. Factor analysis in educational research. *Rev. Educ. Res.* 21:368-88.

Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H. 2014. The p factor: One general psychopathology structure in the structure of psychiatric disorders? *Clin. Psychol. Sci.* 2:119-37.

DeMars CE. 2006. Application of the bi-factor multidimensional item response theory model to testlet-based tests. *J. Edu. Meas.* 43:145-68.

Eid M, Geiser C, Koch T, Heene M. 2017. Anomalous results in g-factor models: Explanations and alternatives. *Psychol. Methods.* 22:541-62.

Frisby CL, Beaujean AA. 2015. Testing Spearman's hypotheses using a bi-factor model with WAIS-

IV/WMS-IV standardization data. *Intell.* 51:79-97.

Galton F. 1883. *Introduction into human faculty and its development.* Macmillan: New York. 388 pp.

Gignac GE. 2016. The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intell.* 55:57-68.

Gonzales O, MacKinnon DP. 2018. A bifactor approach to model multifaceted constructs in statistical mediation analysis. *Educ. Psychol. Meas.* 78:5-31.

Green S, Yang Y. 2017. Empirical underidentification with the bifactor model: A case study. *Educ. Psychol. Meas.*: In press.

Greene AL, Eaton NR. 2017. The temporal stability of the bifactor model comorbidity: An examination of moderated continuity pathways. *Compr. Psychiatry.* 72:74-82.

Grunwald PD. 2007. *The minimum description length principle.* Cambridge, MA: MIT Press.

Gu H, Wen Z, Fan X. 2017. Examining and controlling for wording effect in a self-report measure: A monte carlo simulation study. *Struct. Equ. Model.* 24:545-55.

Holzinger KJ, Swineford F. 1937. The bi-factor method. *Psychom.* 2:41-54.

Humphreys LG. (1981). The primary mental ability. In *Intelligence and Learning*, ed. MP Friedman, JP Das, N O'Connor, pp. 87-102, New York: Plenum.

Jennrich RI, Bentler PM. 2011. Exploratory bifactor analysis. *Psychom.* 76:537-49.

Jennrich RI, Bentler PM. 2012. Exploratory bifactor analysis: The oblique case. *Psychom.* 77:442-54.

Kim H, Eaton NR. 2015. The hierarchical structure of common mental disorders: Connecting multiple levels of comorbidity, bifactor models, and predictive validity. *J. Abnorm. Psychol.* 124:1064-78.

Koch T, Holtmann J, Bohn J, Eid M. 2018. Explaining general and pecific factors in longitudinal multimethod, and bifactor models: Some caveats and recommendations. *Psychol. Methods.*: In press.

Krueger RF, Markon KE, Patrick CJ, Benning SD, Kramer MD. 2007. Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *J. Abnorm. Psychol.* 116:645-666.

Laceulle OM, Vollebergh WAM, Ormel J. 2015. The structure of psychopathology in adolescence: Replication of a general psychopathology factor in the TRAILS study. *Clin. Psychol. Sci.* 3:850-860.

Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. 2012. Is there a general factor of prevalent psychopathology during adulthood? *J. Abnorm. Psychol.* 121:971-77.

Lv J, Liu JS. 2014. Model selection principles in misspecified models. *J. R. Statist. Soc. B.* 76:141-67.

Mansolf M, Reise SP. 2016. Exploratory bifactor analysis: The Schmid-Leiman orthogonalization and Jennrich-Bentler rotations. *Multivar. Beh. Res.* 51:698-717.

Markon KE, Krueger RF. 2004. An empirical comparison of information-theoretic selection criteria for multivariate behavior genetic models. *Behav. Genet.* 34:593-610.

Martel MM, Pan PM, Hoffman MS, Gadelha A, do Rosário MC. 2017. A general psychopathology factor (p factor) in children: Structural model analysis and external validation through familial risk and child global executive function. *J. Abnorm. Psychol.* 126:137-48.

McLarnon MJ, Goffin RD, Schneider TJ, Johnston NG. 2016. To be or not to be: Exploring the nature of positively and negatively keyed personality items in high-stakes testing. *J Personal. Assess.* 98:480-90.

Murray AL, Johnson W. 2013. The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intell.* 41:407-22.

Naragon-Gainey K, Prenoveau JM, Brown TA, Zinbarg RE. 2016. A comparison and integration of structural models of depression and anxiety in a clinical sample: Support for and validation of the tri-level model. *J. Abnorm. Psychol.* 125:853-67.

Nalisnick E, Smyth P. 2017. Learning approximately objective priors. In *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence, Sydney, Australia, Aug. 11-15*, http://auai.org/uai2017/proceedings/papers/188.pdf

Patalay P, Fonagy P, Deighton J, Belsky J, Vostanis P, Wolpert M. 2015. A general psychopathology factor in early adolescence. *Br. J. Psychiatry.* 207:15-22.

Reise SP. The rediscovery of bifactor measurement models. *Multivar. Beh. Res*. 47:667-96.

Reise SP, Kim DS, Mansolf M, Widaman KF. 2016. Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg Self-Esteem Scale. *Multivar. Beh. Res.* 51:818-38.

Reise SP, Moore TM, Haviland MG. 2010. Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *J. Pers. Assess.* 92:544-559.

Rissanen J. 2003. Complexity of simple nonlogarithmic loss functions. *IEEE Transact. Inf Theory.* 49:476-84.

Rissanen J. 2007. *Information and complexity in statistical modeling.* New York: Springer.

Schmid J, Leiman JM. 1957. The development of hierarchical factor solutions. *Psychom.* 22:53-61.

Shevlin M, McElroy E, Bentall RP, Reininghaus U, Murphy J. 2017. The psychosis continuum: Testing a bifactor model of psychosis in a general population sample. *Schizophr. Bull.* 43:133-141.

Spearman CE. 1904a. The proof and measurement of association between two things. *Am. J. Psychol.* 15:72-101.

Spearman CE. 1904b. "General intelligence", objectively determined and measured. *Am. J. Psychol.* 15:201-93.

Stochl J, Khandaker GM, Lewis G, Perez J, Goodyer IM. 2015. Mood, anxiety, and psychotic phenomena measure a common psychopathological factor. *Psychol. Med.* 45:1483-1493.

Thurstone LL. 1944. Second-order factors. *Psychom.* 9:71-100.

Thomson GH. 1939. *The factorial analysis of human ability*. University of London Press: London. 395 pp.

Tomas JM, Oliver A. 1999. Rosenberg's Self-Esteem Scale: Two factors or method effects. Structural Equation Modeling. *Struct. Equ. Model.* 6:84-98.

Waller NG. 2017. Direct Schmid-Leiman transformations and rank-deficient loadings matrices. *Psychom.* : In press.

Wherry RJ, Winer BJ. 1953. A method for factoring large numbers of items. *Psychom.* 18:161-79.

Wherry RJ. 1959. Hierarchical factor solutions without rotation. *Psychom.* 24:45-51.

White H. 1982. Maximum likelihood estimation of misspecified models. *Econom.* 50:1-25.

Witthöft M, Fischer S, Jasper F, Rist F, Nater UM. 2016. Clarifying the latent structure and correlates of somatic symptom distress: A bifactor model approach. *Psychol. Assess.* 28:109-115.

Yung Y-F, Thissen D, McLeod LD. 1999. On the relationship between the higher-order model and the hierarchical factor model. *Psychom.* 64:113-28.

FIGURE CAPTIONS

Figure 1. Hierarchical and related models. (***a***) Spearman's (1904a, b) two-factor model, a precursor to hierarchical and bifactor models. The two-factor model includes a general factor (G) as well as systematic specific factors (S) and random error factors (e). Spearman's two-factor model cannot be estimated as originally formulated, but it established the idea of a superordinate general factor plus subordinate specific factors that account for systematic residual influences not accounted for by the general factor. (***b***) Hierarchical or bifactor model, which includes superordinate general factors (G) as well as subordinate specific factors (S); error factors are not shown. Bifactor models are a subtype of hierarchical model with one superordinate factor and multiple subordinate factors. The two-factor model and hierarchical model are examples of "top down" models, in that subordinate factors instantiate residual effects that are unexplained by the superordinate factor.

Figure 2. Higher-order and correlated factors models. (***a***) A simple-structure correlated factors model, where subordinate, lower-order factors correlate with one another. (***b***) Higher-order model, which includes superordinate general factors (G) as well as subordinate specific factors (S); error factors are not shown. In the higher-order model, superordinate factors account for correlations between subordinate factors. These types of models are examples of "bottom-up" models, in that superordinate structural features instantiate relationships between subordinate factors.

Figure 3. Density plots of the difference in log-likelihoods (lnL) between bifactor and correlated simple structure models as fit to simulated data. 1000 simulated data samples (N=250 each) were randomly generated under a correlated simple structure population model ("Simple Structure"), a bifactor population model ("Bifactor"), or an unstructured condition ("Random"). Correlated factor and

bifactor models were fit to each simulated dataset, and the difference in lnL was computed. The two structured population models comprised estimates reported by Caspi and colleagues (2014; Models A and B' in Table 1 of Caspi et al. 2014). In the unstructured condition, data were uniformly sampled from the ranges of the data used in that paper. Bifactor models tended to fit better by lnL in all conditions, even when the data had no structure. The estimated mean difference in lnL was 3.09 under a correlated factors population; 23.86 under a bifactor population, and 4.32 across random datasets. Using AIC, the expected mean difference in the two structured population conditions would be 3; using BIC, it would be 16.56.

Figure 1

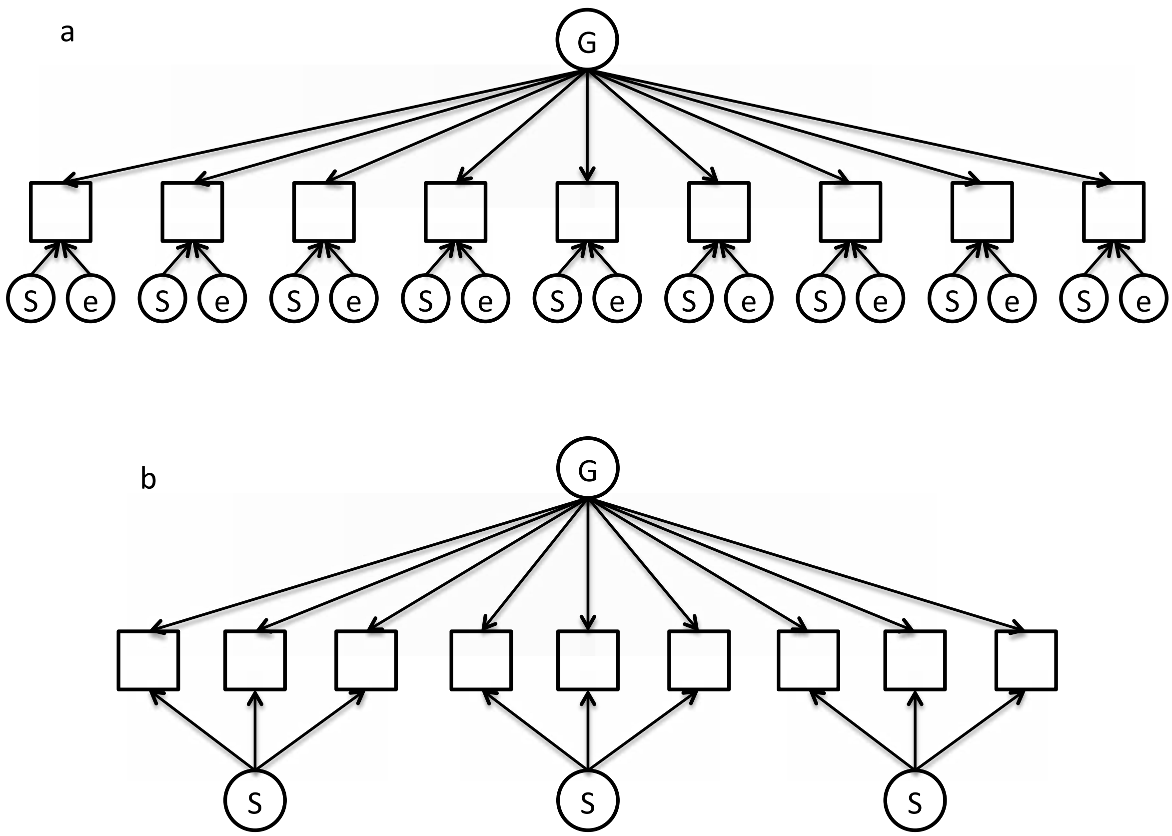


Figure 2

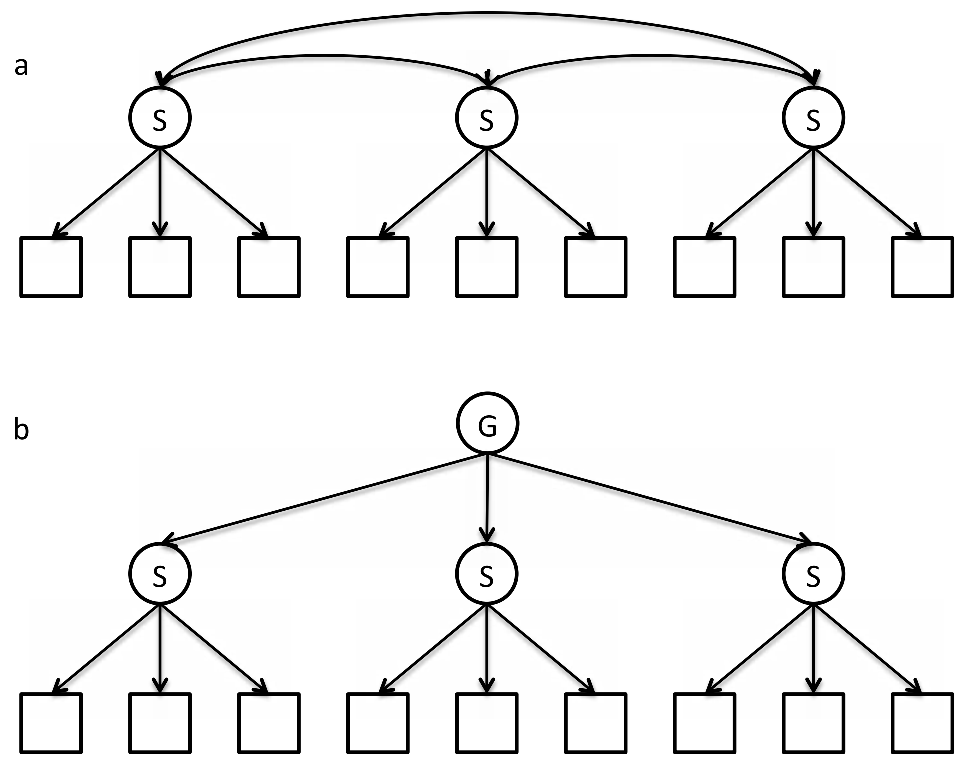
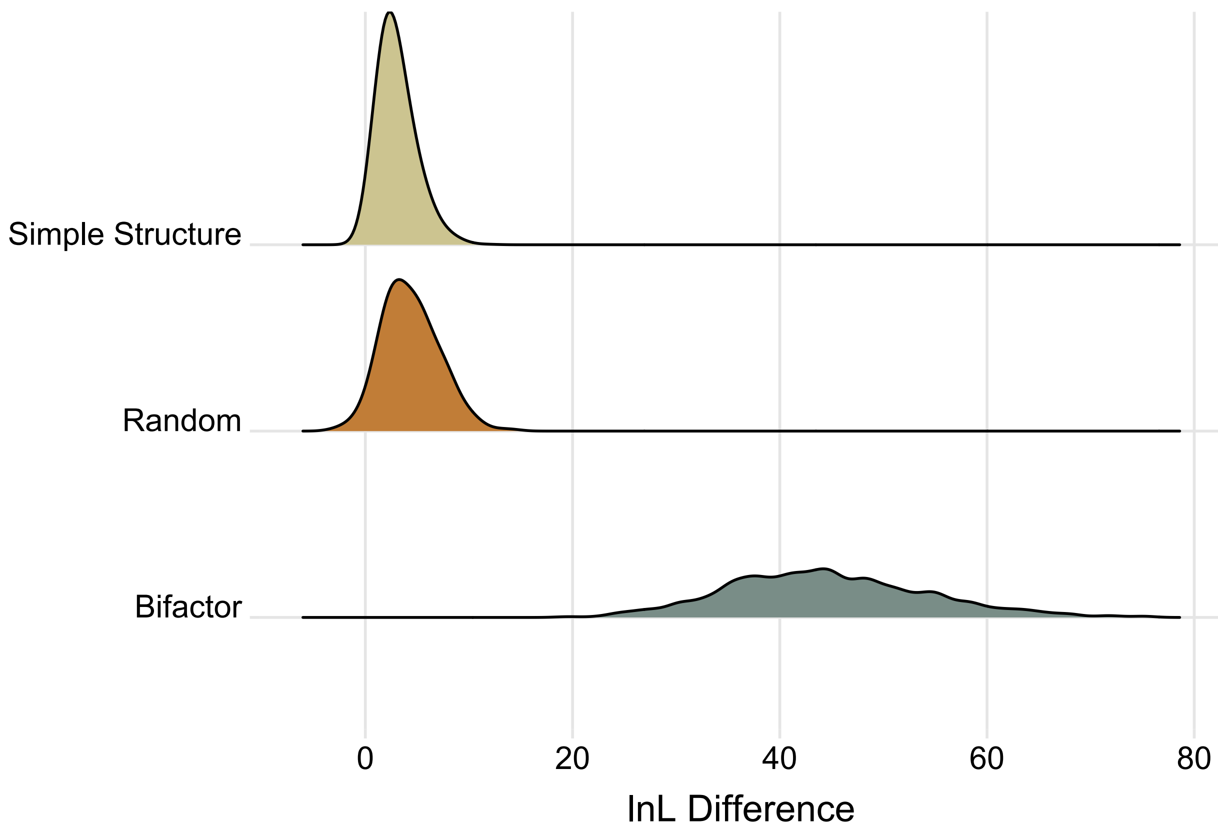


Figure 3


Simple Structure
Random
Bifactor
0
20
40
60
80
lnL Difference