Invited Paper: The Rediscovery of Bifactor Measurement Models

Steven P. Reise
University of California, Los Angeles

Steven P. Reise: reise@psych.ucla.edu

Abstract

Bifactor latent structures were introduced over 70 years ago, but only recently has bifactor modeling been rediscovered as an effective approach to modeling *construct-relevant* multidimensionality in a set of ordered categorical item responses. I begin by describing the Schmid-Leiman bifactor procedure (Schmid & Leiman, 1957), and highlight its relations with correlated-factors and second-order exploratory factor models. After describing limitations of the Schmid-Leiman, two newer methods of exploratory bifactor modeling are considered, namely, analytic bifactor (Jennrich & Bentler, 2011) and target bifactor rotations (Reise, Moore, & Maydeu-Olivares, 2011). In section two, I discuss limited and full-information estimation approaches to confirmatory bifactor models that have emerged from the item response theory and factor analysis traditions, respectively. Comparison of the confirmatory bifactor model to alternative nested confirmatory models and establishing parameter invariance for the general factor also are discussed. In the final section, important applications of bifactor models are reviewed. These applications demonstrate that bifactor modeling potentially provides a solid foundation for conceptualizing psychological constructs, constructing measures, and evaluating a measure's psychometric properties. However, some applications of the bifactor model may be limited due to its restrictive assumptions.

A bifactor structural model specifies that the covariance among a set of item responses can be accounted for by a single general factor that reflects the common variance running among all scale items, and group[1] factors that reflect additional common variance among clusters of items, typically, with highly similar content. It is assumed that the general and group factors all are orthogonal. Substantively, the general factor represents the conceptually broad "target" construct an instrument was designed to measure, and the group factors represent more conceptually narrow subdomain constructs.[2] The bifactor model, thus, appears ideally suited for representing the construct-relevant multidimensionality that arises in the responses to measures of broad constructs where multiple and distinct domains of item content are included to increase content validity (see, for example, Reise, Moore, & Haviland, 2010).

Although originally described over 70 years ago (Holzinger & Harman, 1938; Holzinger & Swineford, 1937), bifactor modeling has spent the last 50 years overshadowed by the numerous applications of Thurstone's correlated-factors model. It only is recently that bifactor models have been rediscovered as an important alternative structural representation of multidimensionality and a topic of research and application in item response theory (IRT) and structural equation modeling (SEM). Evidence of this renewed enthusiasm is abundant and comes in several forms, for example:

---

[1]Many authors refer to group factors as "specific" factors. I prefer to reserve the term specific for that part of an item's reliable variance not shared with other items. Generally, an item's specific variance cannot be separated from its error variance, and, thus, both are combined to form an item's uniqueness.

[2]Gustafsson and Aberg-Bengtsson (2010, p. 107) have emphasized that, "breath of influence on observed variables, rather than distance from observed variables, is what distinguishes broad and narrow factors."

a. Major personality and assessment journals now routinely include articles demonstrating applications of bifactor modeling (e.g., Ackerman, Donnellan, & Robins, 2012; Bados, Gomez-Benito, & Balaguer, 2010; Ebesutani et al., in press; Gibbons, Ruch, & Immekus, 2009; Gignac, Palmer, & Stough, 2007; Patrick, Hicks, Nichol, & Krueger, 2007);

b. Didactic articles recently have appeared arguing for the utility of bifactor models in resolving important problems in conceptualizing and measuring psychological constructs (Brunner, Nagy, & Wilhelm, in press; Cai, Yang, & Hansen, 2011; Chen, Hayes, Carver, Laurenceau, & Zhang, 2012; Gustafsson & Aberg-Bengtsson, 2010; Reise, Moore, & Haviland, 2010; Reise, Morizot, & Hays, 2007; Thomas, 2012);

c. Published psychometric articles now compare the bifactor to competing structural representations (Chen, West, & Sousa, 2006; Rijmen, 2010), provide solutions to challenging estimation problems (Cai, 2010c; Rijmen, 2009), demonstrate important extensions of bifactor modeling to computerized adaptive testing (Gibbons, et al., 2008), vertical scaling (Li & Lissitz, 2012), and assessing differential item functioning (Fukuhara & Kamata, 2011; Jeon, Rijmen, & Rabe-Hesketh, in press); and importantly,

d. User friendly software now is available that facilitates the estimation of parameters for a variety of latent variable models, including bifactor (e.g., *IRTPRO* 2.1, Cai, Thissen, & du Toit, 2011; *EQSIRT*, Wu & Bentler, 2011).

Despite the above contributions, many conceptual as well as technical issues in the application of bifactor models remain poorly understood in the psychometric and assessment communities. The primary goals of this review, thus, are to: a) provide insight into several of these issues, b) point out strengths and limitations of bifactor modeling, and c) call attention to topics in need of further research. To accomplish these goals, the remainder of this article is divided into three sections.

In the first section I describe exploratory approaches to bifactor modeling. At present, exploratory bifactor modeling is greatly underused by applied researchers. This is unfortunate because it is critically important to explore one's data thoroughly prior to proceeding to apply more restrictive, confirmatory models. Exploratory analyses allows researchers to identify potential modeling problems directly, rather than indirectly through post-hoc inspection of fit and modification indices after estimating a confirmatory model (see Browne, 2001, p. 124–125 for additional commentary). In the second section, I review confirmatory bifactor approaches arising from the factor analytic and IRT literatures, as well as competing models, such as the confirmatory second-order and correlated-factors models. I also address the topic of establishing general factor invariance. In the final section I review applications of bifactor modeling that address important problems in the psychometric evaluation of a measure. This is arguably the most important section, for without applications of substantive consequence, bifactor modeling would be of little contemporary interest.

For illustrative purposes, throughout I use a sample of 1,060 adolescents who responded to the 15 anxiety items from the Revised Child Anxiety and Depression Scale (*RCADS-15*) described in Ebesutani et al. (in press). For this report, I dichotomized all item responses (1 versus 2, 3, 4) to simplify analyses and to avoid reporting results that are redundant with Ebesutani et al. (in press). The *RCADS-15* was designed to be a short scale emphasizing the precise measurement of global anxiety but for content validity purposes, includes three items for each of the five diagnostic categories (separation anxiety disorder (*SAD*), generalized anxiety disorder (*GAD*), panic disorder (*PD*), social anxiety disorder (*SOC*), and obsessive-

compulsive disorder (*OCD*)). Abbreviated item content is shown in Table 1 and estimated tetrachoric correlations are provided in Table 2.

EXPLORATORY BIFACTOR MODELS

An exploratory bifactor approach to factor analysis was developed in a series of reports entitled "Preliminary reports on Spearman-Holzinger Unitary Trait Study" that were summarized in Holzinger and Swineford (1937). An elegant and simple bifactor estimation method – called the Schmid-Leiman orthogonalization (*SL*; Schmid & Leiman, 1957) – was introduced 20 years later (see also Schmid, 1957; Wherry, 1959). Since its introduction, the *SL* method has been the dominant approach to exploratory bifactor modeling, and this remains true today.

Although applications of exploratory correlated-factors analysis are common in psychology, reports of exploratory bifactor analysis are rare. In part, this may be attributable to: a) popular statistical software packages have not included the *SL* as part of their factor rotation options, and b) assessment researchers may not be aware that alternative exploratory rotations, such as a SL bifactor, are simply transformations of the familiar correlated-factors and second-order solutions. Accordingly, the aims of this first section are to: a) demonstrate the relations between exploratory correlated-factors, second-order, and the *SL*, b) describe limitations of the *SL* method, and c) call attention to recent innovations in exploratory bifactor modeling that potentially address these limitations.

The Equivalence of Exploratory Models

To understand the relation between correlated-factors, second-order, and *SL*, the tetrachoric correlation matrix for *RCADS-15* was submitted to a series of exploratory analyses using the *schmid* routine in the *psych* library (Revelle, 2012) available in the *R* 2.12.9 statistical package (*R* Software Development Core, 2012). For all analyses, minres extraction was used with oblimin rotation. In the left hand panel of Table 3 are the estimated loadings (top) and factor intercorrelations (bottom) for a familiar, five factor correlated-factors solution:

$$\widehat{R}=\Lambda_{15.5}(\phi_{5.5})\Lambda_{15.5}{}^{T}+\Theta_{15.15} \quad (1)$$

Where, *R* is a 15 by 15 model reproduced correlation matrix;   is a 15 by 5 loading matrix; φ is a 5 by 5 matrix of correlations among the primary factors;   a 15 by 15 diagonal matrix with first-order uniqueness on the diagonal. Results show that there is a fairly good independent cluster structure (McDonald, 1999) with most items loading strongly on only one of the five factors, and near zero otherwise. The primary factors are moderately correlated.

Equation 1 is the familiar, "default" statistical representation of multidimensional structure in psychology. Substantively, this model considers the trait of pathological childhood anxiety as multifaceted and consisting of five (correlated) primary traits. This structural model "hides" the common variance among the factors (and thus the items) in the φ matrix. As a consequence, this model is attractive for assessment researchers who desire to characterize individual differences through a profile of scores on relatively conceptually narrow constructs.

In the middle ( ) row of the left panel of Table 3 is shown a set of five second-order factor loadings (correlations) found by factor analyzing the phi matrix into a single common factor. In the present data, the five loadings of the primary factors on the higher-order factor are .79, .65, .69, .45 and .63, respectively. Thus, the percentages of primary factor variance explained by the second-order factor are, .62, .42, .48, .20, and .40, respectively. The

unexplained variance (disturbances) for the primary factors must then be .38, .58, .52, .80, and .60, respectively.

When these loadings are combined with the loadings of items on the primary factors, this is called a second-order model. Statistically, this model attempts to account for the correlations among primary factors by stipulating a single second-order factor. Thus, the second-order model allows individual differences on both a general trait and conceptually narrower subtraits to be recognized in the same model. Importantly, however, there are no direct relations between the second-order factor (general anxiety) and the primary trait indicators (items). Rather, the effect of anxiety on each item works indirectly through the five primary factors (traits).

Although the second-order model appears to be more substantively informative than the correlated-factors model, the differences are illusory. This second-order structure (Equation 2) is merely a re-expression of the correlations among the primary traits (see Equation 3) and, thus, the models in Equation 1 and 2 are equivalent.

$$\widehat{R} = \Lambda_{15.5}(\Gamma_{5.1}\Phi_{1.1}\Gamma_{1.5} + \Psi_{5.5})\Lambda_{5.15}{}^{T} + \Theta_{15.15} \quad (2)$$

$$(\Gamma_{5.1}\Phi_{1.1}\Gamma_{1.5} + \Psi_{5.5}) = \phi_{5.5} \quad (3)$$

Where, is a 5 by 1 matrix of loadings of the primary factors on the second-order; is the correlation matrix of second-order factors (1 in this simple case), a 5 by 5 matrix with disturbances on the diagonal and residual correlations among the primary factors on the off-diagonal.

Finally, I formed a 5 by 6 transformation matrix, **T**, with the first column equal to the loadings of the primary factors on the second-order factor, and the diagonal of the remaining matrix equal to the square root of the unique variance (disturbances) for the primary factors. In the present case, these values are: .62, .76, .72, .89, and .77. A bifactor structure then can be generated by post-multiplying the pattern matrix from the correlated-factors solution by **T** (see Equations 4 and 5). The results are shown in the right hand panel of Table 3. This transformation is the *SL*, which is nothing more than a reparameterization (orthogonalization) of the second-order exploratory solution.

$$SL = \Lambda_{15.5}T = \Lambda_{15.5}\left[\Gamma_{5.1}|\Psi_{5.5}^{1/2}\right] = \left[\Lambda_{15.5}\Gamma_{5.1}|\Lambda_{15.5}\Psi_{5.5}^{1/2}\right] \quad (4)$$

is a solution such that,

$$\widehat{R} = (SL)(SL)' + \Theta_{15.15} \quad (5)$$

In the *SL*, the common variance among all the items is represented as a general anxiety dimension, and narrower anxiety subdomains are represented as a set of five group factors that are orthogonal to each other and to the anxiety dimension. Consequently, group factors in the *SL* do not have the same interpretation as primary factors in the previous models – the latter reflects two sources of variance (general and group) and the former reflects only group. This separating out of sources of variance is a chief virtue of bifactor structural representations and underlies many applications, as described later in this report. Finally, in contrast to the second-order, in the *SL* items are influenced directly by both general and group factors.

Problems With The SL

There are two important points in the above equations and in the results in Table 3. First, although these exploratory models offer substantively different representations of the latent structure, they are functionally equivalent. In other words, they are a reparameterization of each other, and any multidimensional dataset with correlated primary factors, arguably, can be viewed though the lens of any of the three structural representations. Second, the relations among the models are clear in that, assuming perfect independent cluster structure where each item loads on only a single primary factor and has zero loadings otherwise:

    a. An item's loading on the general factor in the *SL* is a product of its loading on a primary factor, times the loading of the primary factor on the second-order factor;

    b. An item's loading on a group factor in the *SL* is a product of its loading on a primary factor, times the square root of the residual variance of the primary factor; and,

    c. If perfect independent cluster structure is violated, like in the *RCADS-15* data, and items cross-load on two or more factors in a correlated-factors solution, it likely will cross-load on the group factors in the *SL* solution as well. In turn, this can be highly problematic for the *SL*. Generally speaking, in the presence of non-zero cross-loadings, general factor loadings are overestimated and group factor loadings are underestimated (see Reise, Moore, & Haviland, 2010; Reise, Moore, & Maydeu-Olivares, 2011, for examples). The larger the cross-loadings of items on multiple group factors, the larger the degree of this distortion.

Beyond the problems caused by cross-loading items, a second important concern with the *SL* is that it contains proportionality constraints (see Yung, Thissen, & McLeod, 1999). Clearly, for items within a group factor, their loadings on the general and group factor are found by multiplying their loading on the primary by the same two constants: a) the loading of the primary factor on the second-order factor, and b) the square root of the residual variance of the primary factor, respectively. In turn, if the data have perfect independent cluster structure, then the ratio of the general to group factor loadings for all items within a group factor will be exactly the same (i.e., proportional).[3] Since this forced proportional pattern of loadings is unlikely to be true in a population, these constraints are a serious concern. For example, Brunner, Nagy, and Wilhelm (in press, p. 13) note, "the proportionality constraint limits the value of the higher-order factor model in providing insights into the relationships between general and specific abilities, on the one hand, and other psychological constructs, sociodemographic characteristics, or life outcomes, on the other … ."

Alternatives to the SL

Given the problems with cross-loadings and proportionality constraints noted above, it is important to consider contemporary approaches for estimating the parameters of exploratory bifactor models that do not impose proportionality constraints. The first method I consider is target bifactor rotations (Reise, Moore, & Maydeu-Olivares, 2011). The basic idea of a target rotation is for the researcher to *a priori* specify, based on preliminary data analyses or theory, a factor pattern matrix of specified (typically 0) and unspecified elements (? or + if must be positive). Factor extraction then is conducted as usual, but the extracted matrix is rotated to minimize the difference between the estimated factor pattern and the specified elements of the target factor pattern (see Browne, 2001, Equation 13, p. 124). Cai (2010a, p.

---

[3]When items load (or are set to load) only on one factor in the correlated-factors model, and zero otherwise, the proportionality of the loadings within group factors is obvious in the *SL* solution. For *RCADS-15* data (Table 2), which do display some small cross-loadings, the effects of the proportionality constraints are not obvious from the table.

49) suggests that the root-mean square standard deviation computed on the difference between the estimated pattern and the target pattern be used to judge the adequacy of the resulting solution.

In a Monte Carlo simulation, Reise, Moore, and Maydeu-Olivares (2011) generated dichotomous item response data from populations with known bifactor loading patterns. They then used a preliminary *SL* analysis to suggest how a target bifactor pattern should be specified. For example, if the *SL* loading was greater than .20, they marked that target pattern loading as an unspecified element, and if the *SL* loading was less than .20, they marked that target pattern loading as a specified zero. Then, using *MPLUS* (Muthén & Muthén, 2010), they evaluated how well target bifactor rotations were able to recover the known true population parameters. Of special note was that target bifactor rotations often were able to correctly estimate solutions where the items displayed cross-loadings on group factors.

A second alternative exploratory approach is analytic bifactor rotations (Jennrich & Bentler, 2011). These authors did not conduct a Monte Carlo investigation, but rather provided example applications of the bifactor rotation to data that had been previously analyzed using confirmatory factor methods. They found that results of the exploratory bifactor rotation technique appear to agree well with the published confirmatory factor results. Clearly, more research is needed on the strength and weaknesses of the bifactor rotation method, especially in terms of its ability to handle non-zero cross-loadings.

To illustrate these methods in the *RCADS-15* data, the left hand panel of Table 4 displays a target bifactor rotation and the right hand panel displays the analytic bifactor rotation. The target pattern for the target bifactor model was based on the theory that all items load on the general factor, and each item loads on a single group factor. Specifically, the target pattern had all ? (unspecified) elements in the first column. In columns 2 through 6, three items from a specific content domain had ? (unspecified) elements and 0s (specified) otherwise. The target bifactor model was estimated with *CEFA* 3.02 (Browne et al., 2008) using tetrachoric correlations, least squares extraction, and orthogonal rotation to a target. The analytic bifactor rotation was estimated using personal software, but note that it is an available feature of *EQS 6.2* (Bentler, 2006) and the *psych* library (Revelle, 2012).

Interestingly, the target bifactor rotation in the left panel of Table 4 appears very similar to the *SL* displayed in the right panel of Table 3. In contrast, the analytic bifactor model in the right panel of Table 4 is highly similar to the *SL* and target bifactor models, with the exception that Item #5 appears to be a pure marker of the general factor in the analytic bifactor solution. Item #5 has the highest average correlation with the other *RCADS-15* items, and this may in turn contribute to its high loading on the general in the analytic rotation. Nevertheless, without more research on the analytic bifactor procedure, and how it functions under diverse conditions, it is not immediately clear why this result occurs.

CONFIRMATORY BIFACTOR MODELS

In a confirmatory bifactor model, each item is allowed to load on a general factor, and only one group factor. All other loadings are fixed to zero, and all factors are specified to be orthogonal. In confirmatory bifactor models, the problem of proportionality constraints imposed by the *SL* are no longer of concern. However, the potential parameter distorting effects of forcing small cross-loadings to zero, and accommodating items with substantial cross-loadings on group factors, remains troublesome (see also Finch, 2011). This is one reason why I previously emphasized the necessity of inspecting the data structure carefully through an exploratory bifactor analysis prior to considering confirmatory modeling. Unfortunately, the practice of rushing to estimate a confirmatory bifactor model, and

cavalierly reporting a "good fit", or that the bifactor "fits better" than some nested model, is nearly universal in recently published reports.

This section is divided into three parts. First, I describe estimation approaches for confirmatory bifactor measurement models developed from two distinct latent variable modeling traditions – factor analysis and IRT. These approaches differ primarily in parameter estimation method (full-versus limited-information) and model evaluation methods used. Second, in contrast to the previous section, I demonstrate that in confirmatory mode, bifactor, second-order, and correlated-factors models form a nested hierarchy of alternative multidimensional structural representations. Third, I describe item parameter invariance conditions for bifactor models. The establishment of item parameter invariance is an important, but frequently overlooked aspect of exploring the appropriateness of a measurement model's applications.

Two Approaches to Parameter Estimation

Over the last 20 years, there has been increased interest in the development of factor analytic approaches appropriate for the analysis of ordered categorical (dichotomous or polytomous) item response data (Wirth and Edwards, 2007). To understand the factor analysis of ordinal variables, consider n dichotomously scored items, factored into one general (GEN) and three group (GR) factors. The ordinal factor analysis model assumes that the observed 0 or 1 item response is a discrete realization of a continuous and normally-distributed latent response process (x^*) underlying the items. A linear factor model can then be written as:

$$E(x_i^*|\theta_1..\theta_P)=x_i^*=\lambda_{iGEN}\theta_{1GEN}+\lambda_{iGR1}\theta_{GR1}+\lambda_{iGR2}\theta_{GR2}+\lambda_{iGRP3}\theta_{GRP3}+\varepsilon_i. \quad (6)$$

Where   are latent factor scores, $p = 1$ to P factors, and the are standardized factor loadings. To complete the model, an item threshold parameter ( ) needs to be estimated such that $x =$ 1, $if x^*$ $_i$ and $x_i = 0$, $if x^*$ $_i$. That is, individuals will endorse an item only if their response propensity is above the item's threshold. Thus, in ordinal factor analysis, both item loading and threshold parameters need to be estimated for each item.

One approach to implementing this model simply is to replace Pearson correlations with tetrachoric or polychoric correlations and then conduct a limited-information factor analysis (e.g., weighted least squares; see Wirth and Edwards, 2007). These estimators are called limited-information because only the mean and covariances among the items are used to estimate item parameters (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009). Knol and Berger (1991), for example, demonstrated that an ordinary least squares factor analysis of tetrachorics often can recover known item parameters just as well, if not better than, more complicated estimation methods (see also Finch, 2010; 2011).

The top portion of Table 5 displays the results of estimating a confirmatory bifactor model using robust maximum likelihood estimation with EQS (Bentler, 2006). All items were treated as categorical, and the tetrachoric correlation matrix was estimated by EQS. There are $(15 \times 14) / 2 = 105$ unique correlations and 30 estimated parameters (15 loadings on the group factor, 15 loadings on the general factor). Thus, 105 minus 30 leaves 75 degrees of freedom. The Satorra-Bentler (SB; Satorra & Bentler, 1994) chi-square is 137.70 on 75 DF, robust $CFI = .946$, robust $RMSEA$ is .053 (.047 to .059), and $SRMR$ is .048, indicating that the sample correlation matrix is well recovered. Note, however, that these fit indices do *not* include an evaluation of the estimated threshold parameters.

A second type of bifactor model estimation strategy has arisen from the IRT literature. Specifically, many applications of IRT are based on the marginal maximum likelihood (MML; Bock & Aitken, 1981) estimation method. This method often is referred to as "full-

information" item factor analysis because it uses the entire item response matrix as part of the calibration (Gibbons, & Hedeker, 1992). Only recently have highly efficient dimensionality reduction techniques become available (Cai, 2010abc) that greatly expand the utility of *MML* estimation to a wide variety of confirmatory bifactor models (Cai, Yang, and Hansen, 2011).

To illustrate a bifactor IRT model, Equation 7 displays the two-parameter bifactor model expressed in a logistic-metric.

$$P(x_i=1|\theta)=E(x_i|\theta)=\frac{\exp(z_i)}{1+\exp(z_i)} \quad (7)$$

Where, $z_i=\alpha_{iGEN}(\theta_{GEN})+\alpha_{iGR1}(\theta_{GR1})+\alpha_{iGR2}(\theta_{GR2})+\alpha_{iGR3}(\theta_{GR3})+\gamma_i$

In Equation 7, the probability of endorsing an item is determined by an individual's latent trait scores, , on the general and three group factors, and by item properties: a) the discrimination ($_{GEN}$) of the item on the general factor, b) the discrimination ($_{GRP}$) of the item on the group factor, and c) , a multidimensional intercept parameter reflecting an item's easiness (higher values reflect items with higher endorsement proportions).

Table 6 displays parameter estimates for the *RCADS-15* data for the two-parameter logistic bifactor model as output by *IRTPRO* (Cai, du Toit, & Thissen, 2011), using *MML* full-information estimation. In *IRTPRO* the fit of each item is judged using adjusted chi-square statistics developed by Orlando and Thissen (2000; 2003). In the present case, 2 of 15 items were judged not to fit, $p < .05$. *IRTPRO* also shows indices that reflect the degree of local dependence (Chen & Thissen, 1997), with larger values indicating higher residual correlations between item pairs after controlling for the latent trait(s). No large violations were found in the present data meaning that the bifactor model performs well in accounting for common variance.

Finally, three goodness-of-fit statistics based on the overall contingency table: chi-square, Pearson, (both *not* computable here due to a sparse contingency table), and M_2 (Maydeu-Olivares & Joe, 2005) are provided. Only the latter has been empirically supported by Monte Carlo studies. In the present data, M_2 was 123.77 on 75 degrees-of-freedom with $p < .001$. This M_2 value provides evidence that a model may fit acceptably under an SEM framework but be unacceptable under an IRT framework. The reported *RMSEA*, however, was .02, in agreement with the SEM results.

Equivalence of IRT and SEM Bifactor Models

The above results demonstrate that researchers interested in confirmatory bifactor models have two options for parameter estimation, full- and limited-information. However, it has long been recognized that the 2-parameter normal-ogive IRT model and the factor analytic model for ordinal item responses are equivalent for either dichotomous or polytomous data (Takane & de Leeuw, 1987; Kamata & Bauer, 2008). For this reason, contemporary software provides parameter estimates in both IRT and factor analytic metrics (e.g., *TESTFACT*, Bock et al., 2003; *IRTPRO*, Cai, Thissen, & du Toit, 2011; *MPLUS*, Muthén & Muthén, 2011). Specifically, it can be shown that the IRT parameters in Equation 7, after

conversion to a normal-ogive metric[4], can be transformed into the factor analytic parameters of Equation 6 and vice versa: for $p = 1,\ldots,P$ dimensions, slopes and loadings are,

$$\lambda_{ip} = \frac{\alpha_{ip}}{\sqrt{1 + \sum_{p=1}^{P} \alpha_{ip}^2}} \quad \alpha_{ip} = \frac{\lambda_{ip}}{\sqrt{\psi_i}} = \frac{\lambda_{ip}}{\sqrt{1 - \sum_{p=1}^{P} \lambda_{ip}^2}}. \quad (8)$$

and for thresholds and intercepts,

$$\tau_i = \frac{-\gamma_i}{\sqrt{1 + \sum_{p=1}^{P} \alpha_{ip}^2}} \quad \gamma_i = \frac{-\tau_i}{\sqrt{1 - \sum_{p=1}^{P} \lambda_{ip}^2}}. \quad (9)$$

It is not surprising then that in the *RCADS-15* analyses, if the IRT parameters in Table 6 are converted to a normal ogive metric (by dividing by 1.7) and then converted to factor analytic parameters, the estimates are very close to the Table 5 values (not shown). In this dataset, at least, it appears to make no difference which parameter estimation approach is adopted.

It is tempting to take the above equivalence too far, and, thus, I provide three cautions. First, the approaches are distinct in that each estimation method has characteristic weaknesses. A major limitation of the factor analytic approaches to modeling ordinal data is the well-documented challenges of estimating tetrachoric and polychoric correlations, especially in the presence of missing data. The MML approach, on the other hand, has trouble with the numerical integration involved with high-dimensional data. This problem is solved in bifactor models by collapsing the dimensionality down to two factors (Cai, Yang, and Hansen, 2011), but as a consequence, no full-information confirmatory bifactor estimation software that I am aware of, allows items to load on more than one group factor.

Second, Reise, Moore, and Maydeu-Olivares (2011) point out that in bifactor solutions, the interpretation of item parameters can differ greatly in the IRT and SEM solutions despite the fact they are equivalent. Consider two items where Item A has loading of .50 on the general and .70 on a group factor (communality = .74), and Item B has loading of .50 on the general and .30 on the group factor (communality of .34). These items appear to be equally strong markers of the general factor. In a normal-ogive IRT metric, however, application of Equation 8 reveals that the IRT slope on the general factor for items A and B are 0.98 and 0.61, respectively. In turn, multiplying these slopes by 1.7, the slopes on the general factor for items A and B are 1.66 and 1.04, respectively, if expressed in the logistic IRT metric. An assessment researcher would come to very different conclusions about the psychometric functioning of these items depending on whether they examined the factor analysis or IRT parameter estimates.

Finally, it is not safe to assume that approaches to model fit developed under the linear factor analysis tradition are easily generalizable to the evaluation of non-linear IRT models, and vise versa (see Maydeu-Olivares, Cai, & Hernandez, 2011, for details). Moreover, there is scant work in either SEM or IRT on evaluating the fit of bifactor models based on ordinal item responses. Well known SEM benchmarks for "acceptable fit" developed under the multivariate normality assumption in SEM are may not be helpful in judging the adequacy of a bifactor measurement model based on full-information estimation strategies. In fact, research that has evaluated the use of SEM fit indices in evaluating the unidimensionality

---

[4]IRT models were originally developed under a normal-ogive model to describe the relation between the latent variable and the probability of the item response. However, IRT item parameters are now routinely estimated based on a logistic model (Equation 6) rather than the normal-ogive model. To convert logistic model parameters to their normal-ogive model counter-parts, they need to be divided by 1.7.

assumption in IRT models has found them to be severely lacking (see Cook, Kallen, & Amtmann, 2009).

Nested Models

I now consider competing models that are nested within the bifactor model. Several scholars have advised that, only if the least restricted model (in this case the bifactor) is judged to fit the data (Yuan & Bentler, 2004), it is appropriate to consider whether applying a more restricted, nested model, significantly degrades that fit. For ease of presentation, I will confine the following discussion to specification and comparison of SEM models, but fit statistics could be calculated for analogous nested IRT models as well.

Correlated-Factors Model

The correlated-factors model can be derived from the bifactor by fixing the loadings in the bifactor general factor to zero and freeing the orthogonality constraint on the group factors. This model was estimated by specifying five latent variables (primary factors) with variance equal to 1.0, and three items loading freely on each latent variable. Factor inter-correlations were freely estimated. This model has 25 parameter estimates (15 loadings and 10 correlations) and, thus, $df = 80$. In the present data, the fit is excellent: *SB* chi-square = 119.88 (80 *df*, $p < .01$), robust *CFI* = .99, robust *RMSEA* = .022 (.013 – .029), and *SRMR* = .05. The Satorra-Bentler scaled chi-square difference test comparing the correlated factors to the bifactor model is 16.46 on 5 *df* ($p < .05$), indicating that the bifactor is a (statistically) better model.

Second-order Model

A nested alternative to the correlated-factors model is to place a measurement structure on the correlations between primary factors in an attempt to model and, thus, explain the correlations among the primary factors. In the present data, there are five primary factors so the second-order model is nested within the correlated-factors. To identify the second-order model, for each of the five primary factors, a loading was set to 1.0 for one item within a primary factor, and disturbances were freely estimated. This model has 20 parameter estimates (15 loadings and 5 disturbances) and, thus, $df = 85$. Again, using *EQS* robust maximum likelihood, the *SB* chi-square was 136.15 (85 *df*, $p < .01$), robust *CFI* = .987, robust *RMSEA* = .024 (.016 – .031), and *SRMR* = .056. Clearly, the bifactor, correlated-factors, and second-order models all provide an excellent fit to the correlation matrix. In practice any of these models can be applied with confidence, and the goals of the study dictate model preference.

Finally, in a unidimensional model, each item is allowed to load on a single latent variable. This is the "default" measurement model used in both IRT and SEM (typically after parceling items, however). After fixing the factor variance to 1.0, there are 15 estimated loadings and the $df = 90$. It is nested within the bifactor because it can be derived by setting all group factor loadings in the bifactor to zero (or simply eliminating them). In the *RCADS-15* data, the maximum likelihood parameter estimates in the unidimensional model resulted in a *SB* chi-square of 500.59 (90 *df*, $p < .01$), robust *CFI* = .899, robust *RMSEA* = .066 (.060 – .071), and *SRMR* = .092. In contrast to the multidimensional models considered above, the unidimensional model appears not to be a plausible candidate model. In a subsequent section, however, I will use bifactor modeling to reconsider just how "unacceptable" a unidimensional representation of the multidimensional *RCADS-15* data is.

Invariance of the General Factor

Psychological traits of broad theoretical importance influence behavior in diverse domains – that is what makes them interesting to study and demands that trait measures include heterogeneous item content representing multiple domains of trait manifestation. For example, Chen, West, and Sousa (2006) describe a content heterogeneous self-report measure of health-related quality of life, a construct theorized to influence "cognition, vitality, mental health, and disease worry" (p. 189). Reise, Moore, and Haviland (2010) evaluate an observer-report measure of alexithymia, which in turn, is proposed to influence individual differences in five domains, including being emotionally distant, being psychologically uninsightful, having excessive health worries, lacking humor, and cognitive and behavioral rigidity.

A chief goal of applying a confirmatory bifactor model to item response data resulting from the administration of complex trait measures is to estimate a model such that the parameter estimates on the general factor accurately reflect the relations between items and the general construct of interest (health related quality of life, alexithymia) while controlling for the biasing effects of multidimensionality caused by content diversity. To successfully achieve this goal, the general factor in the bifactor model must validly reflect the common variance running among all the items in a measure. Unfortunately, didactic articles that inform applied researchers regarding the conditions facilitating the correct identification of the general factor are scant. In this section, I therefore raise the critical issue of item parameter invariance in confirmatory bifactor models. To accomplish this objective, however, I first describe the concept of parameter invariance in a unidimensional measurement model.

Latent variable measurement models in both IRT and SEM sometimes have implications for measurement that are at odds with conventional "best practices." For example, the inclusion of content diverse trait indicators that exhaust the range of trait manifestations is often touted as virtuous because it increases a measure's content and thus construct validity. On the other hand, this so-called best practice does not hold if a construct is represented as a single latent variable rather than as a summed score composite. For example, it is well known that if item response data are unidimensional (one common trait explains the correlations among items), the common latent variable can be properly identified with three indicators. Moreover, it does not make any difference what content domain those three items are selected from.

In other words, if the data are truly unidimensional and a unidimensional latent variable measurement model is proposed to fit the data, content representativeness does not make any difference in defining the latent variable, or in the value of the estimated item parameters (Bollen & Lennox, 1991). This item parameter *invariance* property is of profound importance in both SEM and IRT. In the latter, for example, all important applications of a unidimensional IRT model, such as computerized adaptive testing and differential item functioning analysis, depend critically on item parameter invariance.

The concept of item parameter invariance extends to bifactor models. Just as a unidimensional measurement model has parameter invariance when data meet its assumptions, so too does the bifactor if the data are bifactor in the population. Stated differently, given a set of item parameter estimates for a bifactor model that is 30 items with six 5-item group factors, the assumptions of invariance implies that the same general and group factor loadings would result if, for example, only a subset of 15 items with three 5-items group factors were estimated. The applied consequence of this is important. Specifically, establishing item parameter invariance is a critical step in arguing for the validity of the bifactor model. A researcher would be hard pressed to argue for applications of bifactor modeling (see below) if one could not demonstrate, at the very least, that roughly

the same general factor is being measured regardless of which subset of item content domains (group factors) are included. Accordingly, in Table 7 are shown estimates of the general factor for several possible three group factor *RCADS-15* confirmatory bifactor models. These results indicate that the general factor does change slightly according to which group factors are included in the model. Researchers interested in measuring anxiety using the *RCADS-15* are thus advised to use all 15 items.

APPLICATIONS OF BIFACTOR MODELING

Technical innovations seldom lead to substantive application unless researchers are convinced the approach offers something of value. Thus, in this final section, I describe four important psychometric applications of bifactor modeling[5]: a) partitioning item response variance into general versus group factor sources, b) determining the degree to which item response data are unidimensional versus multidimensional, c) estimating the degree to which raw scale scores reflect a single common source, and d) evaluating the viability of subscale scores after variance due to the general factor has been controlled for. Although these procedures can be justifiably applied to either exploratory or confirmatory bifactor solutions, below I work exclusively from the preferred confirmatory perspective.

Partitioning Item Variance

Psychological constructs, and item response data resulting from measures designed to assess such constructs, often are proposed to have a "hierarchical" or "multifaceted" structure (Brunner, Nagy, & Wilhelm, in press; Chen, et al., 2012). One meaning of these terms is that psychological traits affect behavior across heterogeneous behavioral domains. To the degree that a measure includes multiple items from these heterogeneous domains, item response data will have, at least, two *common* sources of variance; one, the factor affecting all items (reflecting the conceptually broad general trait) and a second affecting subsets of content homogeneous items (reflecting conceptually narrow subdomains).

In these cases, factor analyses often will reveal that the data are not strictly unidimensional and that multidimensional models, such as the correlated factors, second-order, or bifactor, provide a better account of the correlations among the items. Due to the orthogonality of general and group factors, however, it is only the latter model that allows researchers to easily partition item response variance into two common sources. In turn, this partitioning can be invaluable in evaluating and refining an existing instrument and in furthering understanding of a trait's structure.

For example, Simms, Gros, Watson, and O'Hara (2008) used confirmatory bifactor modeling to explore the relative contribution of general and group factors in affecting responses to the Inventory of Depression and Anxiety Symptoms (Watson et al., 2007). This instrument includes 76 psychiatric symptoms that are further classified into 13 subdomains. Interestingly, they found that the common variance shared by most symptoms could be partitioned roughly equally between general and group factors. There were some content domains (e.g., dysphoria), however, that primarily were markers of the general factor and others (e.g., appetite problems) primarily reflecting a group factor.

Findings for the *RCADS-15* data are similar. Inspection of Table 5 reveals that the common variance for items such as #1, #4, #7, and #9 is approximately equally accounted for by the general and group factors. Items #3 (fear of crowds) and #9 (panics), on the other hand, are

---

[5]One major application of bifactor modeling is to separate out general and group factors when a researcher is interested in their independent contribution to the prediction of a criterion. This topic was covered extensively in Chen, West, and Sousa (2006) and will not be reviewed here.

predominantly markers of the general factor. Item #11 (worries about other's opinion) displays the opposite pattern, loading primarily on the social phobia group factor. Item #10 (worries about poor performance) does not appear to be a good marker of either the general or the social phobia group factor. Finally, note that sets of items with high loadings on group factors may signal too much content similarity. Items #11 and #12, for example, with very high loadings of .78 and .67 on the social phobia group factor, may be redundant. This is an important consideration – we want the group factors to reflect a conceptually narrow psychological trait and not be a mere artifact of asking the same question repeatedly in slightly different ways.

A Unidimensionality Statistic

It is well established that unidimensional IRT model parameters appear to be reasonably robust even if the data are multidimensional, as long as there is a "strong general factor." What exactly this means empirically and how to identify it are not so well established, however. Thus, it is not surprising that in the IRT literature, there are dozens of proposed procedures for evaluating when multidimensional item response data are "unidimensional enough" (e.g., ratio of 1st to 2nd eigenvalue, residuals after extracting a single factor), and several proposed methods of evaluating the degree of unidimensionality (see, for example, research on the DETECT index (Zhang & Stout, 1999).

I will not review those procedures here, but rather note that if multidimensional item response data are consistent with a bifactor structure, there is a simple approach to indexing the degree of unidimensionality. Specifically, the explained common variance (*ECV*) can be defined as the ratio of variance explained by the general factor divided by the variance explained by the general plus the group factors. In the bottom portion of Table 5 are shown the variance explained (sum of squared loadings) by the common factors in the *RCADS-15* data. These values lead to an *ECV* index of *.54* reflecting that common variance is about equally spread across general and group factors in these data. Generally speaking, the higher *ECV*, the "stronger" the general factor relative to the group factors and thus, the more confidence a researcher has in applying a unidimensional measurement model to multidimensional data. Unfortunately, however, no benchmark values for *ECV* can be proposed for determining when the relative general factor strength is high enough so that it is safe to apply unidimensional models to multidimensional (bifactor) data, because the relation between *ECV* and parameter bias is moderated by the structure of the data.

Specifically, Reise, Scheines, Widaman, & Haviland, et al. (in press), working from a factor analytic model, demonstrated that if item response data are bifactor, and those data are forced into a unidimensional model, structural parameter bias (which depends on loading bias) is a function of the relative strength of the general to group factors (*ECV*), which in turn, is moderated by the percentage of uncontaminated correlations (*PUC*). Generally speaking, when *PUC* is very high (> .90), even low *ECV* values can lead to unbiased parameter estimates.[6] To understand the *PUC* index, consider *RCADS-15* where there are $(15 \times 14) / 2 = 105$ unique correlations. The correlations for items within a group factor are contaminated by both general and group factor variance, and there are $[(3 \times 2) / 2] \times 5 = 15$ of those. The correlations among items from different group factors reflect general factor variance only, and there are $105 – 15 = 90$ uncontaminated correlations and, thus, *PUC* is $90/105 = .86$ – a high value.

---

[6]Unbiased is defined in this context as when the parameter estimates in the unidimensional model are the same as the general factor in the bifactor model.

In Tables 5 and 6, for example, are displayed estimated parameters in the *RCADS-15* when a unidimensional SEM and IRT (logistic metric) are fit, respectively. Observe that in either model, despite the massive multidimensionality of the *RCADS-15* (good fit to a 6-factor model!), the parameter estimates in the unidimensional model are reasonably consistent with those on the general factor in the bifactor. This is evidence the latent variable in the unidimensional model is the same as the general factor in the bifactor model, and thus, the *RCADS-15* item response may indeed be "unidimensional enough" for unidimensional IRT or SEM model application. This result is surprising given the relatively modest *ECV* value and the fact that the confirmatory unidimensional model did not provide an adequate fit to the data, as demonstrated earlier (see Reise et al., in press, for further discussion).

Moreover, understanding *ECV* and *PUC* computed within the context of a bifactor model has implications beyond the prediction and understanding of the biasing effect of forcing bifactor data into a unidimensional measurement model. Moreover, these concepts are important to understand for scale constructors developing measures of broadband, multifaceted constructs. If a researcher has in mind an "essentially unidimensional" but broadband trait measure, then high *PUC* value is desired in order to diminish the biasing effects of the group factors. For example, a 30-item test with 10 3-item group factors, yields a *PUC* of .93. In contrast, for the same 30-item measure with 3 10-item group factors, *PUC* would only be .69. In this latter case, only if the *ECV* is very high, can multidimensional data be modeled using unidimensional models without high degrees of parameter bias.

Interpreting a Scale Score

Before proceeding, an important distinction between the (uni)dimensionality of the data and the interpretability of raw scores needs to be made. If item response data are strictly unidimensional, then raw scores can be unambiguously interpreted as reflecting variation on a single latent variable -- the degree to which observed score variance is due to one *common* source of variance. What is not so commonly known is that the presence of multidimensionality, per se, does not necessarily muddle the interpretability of a unit-weighted composite score, nor does it automatically demand the creation of subscales. Thus, researchers must make a distinction between the degree of unidimensionality in the data, and the degree to which total scores reflect a single common variable.

As noted above, a model-based index of unidimensionality is the percent of common variance due to the general factor. The *ECV* index is not of much value, however, for making judgments about the degree to which raw scores reflect a common dimension. The reason is that a measure with a single weak common factor would be perfectly unidimensional but would produce raw scores that reflect mostly error. Thus, to judge the degree to which composite scale scores are interpretable as a measure of a single common factor, we need a related index called coefficient omega hierarchical (McDonald, 1999; Zinbarg et al., 2005).

$$\omega_H = \frac{\left(\sum \lambda_{iGEN}\right)^2}{\left(\sum \lambda_{iGEN}\right)^2 + \left(\sum \lambda_{iGRP_1}\right)^2 + \left(\sum \lambda_{iGRP_2}\right)^2 .... + \left(\sum \lambda_{iGRP_P}\right)^2 + \sum \theta_i^2} \quad (10)$$

Where *p* represents each common factor (general and group), and θ_i^2 is an item's error variance. Generally speaking, ω_H increases as a function of scale length, the average size of loadings on the general factor, and *PUC* value.

Omega hierarchical is an appropriate model-based reliability index when item response data are consistent with a bifactor structure. The index is simply the sum of the factor loadings on the general factor, squared, divided by the (modeled) variance of scale scores. These values also are shown in the bottom portion of Table 5. Notice that the sum of the denominator terms (60.06, 1.39, 2.34, 1.90, 2.82, 1.56 and 7.09) add up to 77.16, which in this case, is equal to the sum of the elements in the observed tetrachoric correlation matrix.[7]

Using the reported values in Table 5, in the *RCADS-15* data ω_H is estimated to be .79. This value can be contrasted with omega ω (Lucke, 2005) shown in Equation 11, which is also a model-based reliability estimate. The ω index is analogous to coefficient alpha and is affected by all sources of common variance. In the present data, ω is estimated to be .91.

$$\omega_H = \frac{\left(\sum \lambda_{iGEN}\right)^2 + \left(\sum \lambda_{iGRP_1}\right)^2 + \left(\sum \lambda_{iGRP_2}\right)^2 .... + \left(\sum \lambda_{iGRP_P}\right)^2}{\left(\sum \lambda_{iGEN}\right)^2 + \left(\sum \lambda_{iGRP_1}\right)^2 + \left(\sum \lambda_{iGRP_2}\right)^2 .... + \left(\sum \lambda_{iGRP_P}\right)^2 + \sum \theta_i^2} \quad (11)$$

When ω_H is high, composite scores predominantly reflect a single common source even when the data are multidimensional. Gustafsson and Aberg-Bengtsson (2010) show how in large scale aptitude testing, despite concerns that the tests are "multidimensional," scores still are dominated by the general factor.

The Viability of Subscales

When item response data have a multidimensional structure (e.g., correlated factors), the standard practice in psychological research remains the reporting of coefficient alpha for a total scale score and for subscale scores. However, if a bifactor model has been fit to the data, the logic of coefficient omega hierarchical can be extended to the estimation of subscale reliability, controlling for the effects of the general factor. As such, the mathematics of coefficient omega hierarchical can be an invaluable tool in judging whether it is reasonable to report subscale scores (see also, Gignac et al., 2007).

In the following illustration, I use the term omega subscale (ω_S) to clarify that it is a reliability estimate for a residualized subscale – one that controls for that part of the reliability due to the general factor. For comparison, first I compute a model-based reliability estimate (ω) for each of the *RCADS-15* subscales using Equation 11 but apply it to one subset of items at a time. For example, for the *SAD* subscale, the sum of the general factor loadings squared is 2.79, the sum of the group factor loadings squared is 1.39, and the sum of the error variances is 2.53. Coefficient ω for the subscale is thus .62:

$$\omega = \frac{2.79 + 1.39}{2.79 + 1.39 + 2.53} = .62$$

Now I ask, what would the reliability of subscale scores be if the effects of the general factor were removed? This easily can be found by removing the effect of the general factor from the numerator only. For the *SAD* subscale, ω_S is .21.

$$\omega_S = \frac{1.39}{2.79 + 1.39 + 2.53} = .21$$

---

[7]Note, researchers who use a variance-covariance matrix in place of a tetrachoric or polychoric correlation matrix in their analyses must also make sure that the factor loadings in Equation 10 are unstandardized versions.

Coefficient   for the remaining four subscales are: .66, .67, .62, and, .66 respectively. Coefficient $_S$ for the remaining four subscales are: .32, .26, .44, and .22, respectively. These values make clear that if both total scores and subscales were to be formed, the interpretation of the subscales as precise indicators of unique constructs is extremely limited – very little reliable variance exists beyond that due to the general factor.

SUMMARY and LIMITATIONS

In closing this article, it is important to briefly note some important concerns and limitations. First, not all researchers agree that a bifactor model is an appropriate representation of the structure of item response data or, more importantly, psychological traits. A quote in Bagby, Taylor, Quilty, and Parker (2007, p. 258) in reference to Gignac et al.'s (2007) application of a bifactor model to their alexithymia measure, is consistent with this view, "we challenge Gignac et al.'s unheralded and largely unsupported use of a nested model." Moreover, Vanheule, Desmet, Groenvynck, Rosseel, and Fontaine (2008, p. 180) state, in reference to application of a bifactor model to a popular depression measure, "We believe that the inclusion of a G factor which loads on all items is problematic: It is difficult to interpret what this G factor measures, or to implement it in research and practice."

The former concern is hard to address. It appears that the authors assume that multidimensionality must be operationalized through a correlated-factors model, failing to recognize that in a confirmatory framework, the correlated-factors model is nested under the bifactor. Therefore, the bifactor likely will always be better supported in terms of model-to-data fit than a correlated factors model, given the same pattern of constraints. The latter concern is much easier to address; the general factor in the bifactor model represents the single source of common variance running through all the items on a measure, and it is easy to interpret as representing the psychological construct the instrument was likely created to measure. In fact, bifactor modeling is one solution to the interpretative mess that often is created when researchers force multidimensional item response data into a unidimensional measurement model. In such cases, the latent variable, indeed, may reflect a hodgepodge of differentially weighted and psychologically distinct sources of variance.

A second commonly heard objection to a bifactor model is its rigidity in regard to the orthogonality constraint among the general and group factors. In addressing this issue, note that at the least, group and general factors must be orthogonal. Without this constraint, group factors would no longer be interpretable as residualized factors – sources of common variance beyond that explained by the general factor. On the other hand, if the group factors are allowed to correlate with each other (see Jennrich & Bentler, in press), this suggests the presence of additional and unmodeled general factors. Moreover, when group factors are allowed to correlate, implementing any of the applications reviewed above would be challenging. Thus, any gains in fit that may be observed by allowing group factors to correlate among each other ultimately may be offset by losses in model interpretability and applicability.

A third and more persuasive criticism of a bifactor model, in my judgment, is that the model may be too overly restrictive to accurately reflect the structure of item response data in the population. This criticism can be applied to almost any confirmatory model, but for the bifactor, it is especially relevant, because the accuracy of the parameter estimates depends on the constraints being accurate. For example, consider a model for "realistic" data generation proposed by MacCallum and Tucker (1991) where items may be influenced by one or more "dominant" factors but also by dozens of smaller common factors. Under this more realistic model of population data structure, a completely orthogonal confirmatory structure demanding that all items load on one general and one group factor, and have a zero

loading otherwise is unlikely to reflect the structure of real-world psychological data. Recent attempts to better integrate confirmatory and exploratory modeling into one common analytic framework may address this concern in the future, however (Asparouhov & Muthén, 2009).

Finally, it is important to note that bifactor modeling is not the appropriate analytic tool for all types of psychological measures. The model appears best suited for the psychometric analysis of those assessment instruments where the researcher expects a response to primarily reflect a strong common trait, but there is multidimensionality caused by well defined clusters of items from diverse subdomains. Measures that have been shown to fit confirmatory correlated-factors and second-order models are good candidates for consideration of bifactor modeling. On the other hand, for measures with highly homogeneous item content, or measures that were not originally developed with a clear blueprint to include at least three items from at least three contain domains, bifactor modeling is likely to be a challenge.

In closing, the above limitations and concerns need to be weighed against the potential advantages of bifactor models. In my view, the bifactor structural model, which views the variance in trait indicators as being influenced by both general and group sources of variance, provides a strong foundation for understanding psychological constructs and their measurement. Most importantly, I argue that the demonstrations described earlier suggest that bifactor modeling can importantly inform scale construction practices, and the evaluation of a measure's psychometric properties, including the critical evaluation of the necessity of creating and scoring subscales.

Acknowledgments

References

Ackerman RA, Donnellan MB, Robins RW. An item response theory analysis of the narcissistic personality inventory. Journal of Personality Assessment. 2012; 94 141-15.

Asparouhov T, Muthén B. Exploratory structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal. 2009; 16:397–438.

Bados A, Gomez-Benito J, Balaguer G. The state-trait anxiety inventory, trait version: does it really measure anxiety? Journal of Personality Assessment. 2010; 92:560–567. [PubMed: 20954057]

Bagby RM, Taylor GJ, Quilty LC, Parker JDA. Reexamining the factor structure of the 20-item Toronto Alexithymia Scale: Commentary on Gignac, Palmer, and Stough. Journal of Personality Assessment. 2007; 89:258–264. [PubMed: 18001226]

Bentler, PM. EQS structural equations program manual. Encino, CA: Multivariate Software, Inc; 2006.

Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters. Application of an EM algorithm. Psychometrika. 1981; 46:443–459.

Bock, RD.; Gibbons, R.; Schilling, SG.; Muraki, E.; Wilson, DT.; Wood, R. TESTFACT 4.0 [Computer software and manual]. Lincolnwood, IL: Scientific Software; 2003.

Bollen K, Lennox R. Conventional wisdom on measurement: A structural equations perspective. Psychological Bulletin. 1991; 110:305–314.

Browne MW. An overview of analytic rotation in exploratory factor analysis. Multivariate Behavioral Research. 2001; 35:111–150.

Browne, MW.; Cudeck, R.; Tateneni, K.; Mels, G. CEFA: Comprehensive exploratory factor analysis, version 3.02 [Computer software and manual]. 2008. Retrieved from http://quantrm2.psy.ohio-state.edu/browne/

Brunner M, Nagy G, Wilhelm O. A tutorial on hierarchically structured constructs. Journal of Personality. (in press).

Cai L. High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. Psychometrika. 2010a; 75:33–57.

Cai L. Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. Journal of Educational and Behavioral Statistics. 2010b; 35:307–335.

Cai L. A two-tier full-information item factor analysis model with applications. Psychometrika. 2010c; 75:581–612.

Cai, L.; Thissen, D.; du Toit, SHC. IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling [Computer software]. Lincolnwood, IL: Scientific Software International; 2011.

Cai L, Yang JS, Hansen M. Generalized full-information item bifactor analysis. Psychological Methods. 2011; 16:221–248. [PubMed: 21534682]

Chen FF, West SG, Sousa KH. A comparison of bifactor and second-order models of quality of life. Multivariate Behavioral Research. 2006; 41:189–225.

Chen FF, Hayes A, Carver CS, Laurenceau JP, Zhang Z. Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. Journal of Personality. 2012; 80:219–251. [PubMed: 22092195]

Chen WH, Thissen D. Local dependence indices for item pairs using item response theory. Journal of Educational and Behavioral Statistics. 1997; 22:265–289.

Cook KF, Kallen MA, Amtmann D. Having a fit: Impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. Quality of Life Research. 2009; 4:447–460. [PubMed: 19294529]

Ebesutani C, Reise SP, Chorpita BF, Ale C, Regan J, Young J, Higa-McMillan C, Weisz J. The revised child anxiety and depression scale – short version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor. Psychological Assessment. (in press).

Finch H. Item parameter estimation for the MIRT model: Bias and precision of confirmatory factor analysis-based models. Applied Psychological Measurement. 2010; 34:10–26.

Finch H. Multidimensional item response theory parameter estimation with nonsimple structure items. Applied Psychological Measurement. 2011; 35:67–82.

Forero CG, Maydeu-Olivares A, Gallardo-Pujol D. Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. Structural Equation Modeling. 2009; 16:625–641.

Fukuhara H, Kamata A. A bifactor multidimensional item response theory model for differential item functioning analysis on testlet-based items. Applied Psychological Measurement. 2001; 35:604–602.

Gibbons RD, Hedeker DR. Full-information item bi-factor analysis. Psychometrika. 1992; 57:423–436.

Gibbons RD, Rush AJ, Immekus JC. On the psychometric validity of the domains of the PDSQ: An illustration of the bi-factor item response theory model. Journal of Psychiatric Research. 2009; 43:401–410. [PubMed: 18554611]

Gibbons RD, Weiss DJ, Kupfer DJ, Frank E, Fagiolini A, Grochocinski VJ, Bhaumik DK, Stover A, Bock RD, Immekus JC. Using computerized adaptive testing to reduce the burden of mental health assessment. Psychiatric Services. 2008; 59:361–368. [PubMed: 18378832]

Gignac GE, Palmer BR, Stough C. A confirmatory factor analytic investigation of the TAS-20: Corroboration of a five-factor model and suggestions for improvement. Journal of Personality Assessment. 2007; 89:247–257. [PubMed: 18001225]

Gustafsson, JE.; Aberg-Bengtsson, L. Unidimensionality and the interpretability of psychological instruments. In: Embretson, SE., editor. Measuring Psychological Constructs. Washington, DC: American Psychological Association; 2010. p. 97-121.

Holzinger KJ. Relationships between three multiple orthogonal factors and four bifactors. Journal of Educational Psychology. 1938; 29:513–519.

Holzinger KJ, Harman HH. Comparison of two factorial analyses. Psychometrika. 1938; 3:45–60.

Holzinger KJ, Swineford F. The bi-factor method. Psychometrika. 1937; 2:41–54.

Jennrich RI, Bentler PM. Exploratory bi-factor analysis. Psychometrika. 2011; 76:537–549. [PubMed: 22232562]

Jennrich RI, Bentler PM. Exploratory bi-factor analysis: the oblique case. Psychometrika. (in press).

Jeon M, Rijmen F, Rabe-Hesketh S. Modeling differential item functioning using a generalization of the multiple-group bifactor model. Journal of Educational and Behavioral Statistics. (in press).

Kamata A, Bauer DJ. A note on the relation between factor analytic and item response theory models. Structural Equation Modeling. 2008; 15:136–153.

Knol DL, Berger MPF. Empirical comparison between factor analysis and multidimensional item response models. Multivariate Behavioral Research. 1991; 26:457–477.

Li Y, Lissitz RW. Exploring the full-information bifactor model in vertical scaling with construct shift. Applied Psychological Measurement. 2012; 36:3–20.

Lucke JF. The   and the   of congeneric test theory: An extension of reliability and internal consistency to heterogeneous tests. Applied Psychological Measurement. 2005; 29:65–81.

MacCallum RC, Tucker LR. Representing sources of error in the common factor model: Implications for theory and practice. Psychological Bulletin. 1991; 109:502–511.

Maydeu-Olivares A, Joe H. Limited and full information estimation and testing 2n contingency tables: A unified framework. Journal of the American Statistical Association. 2005; 100:1009–1020.

Maydeu-Olivares A, Cai L, Hernandez A. Comparing the fit of item response theory and factor analysis models. Structural Equation Modeling. 2011; 18:333–356.

McDonald, RP. Test theory: A unified treatment. Mahwah, NJ: Lawrence Erlbaum associates; 1999.

Muthén, LK.; Muthén, BO. Mplus User's Guide. 6th ed.. Los Angeles, CA: Muthén & Muthén; 1998–2010.

Orlando M, Thissen D. Likelihood-based item-fit for dichotomous item response theory models. Applied Psychological Measurement. 2000; 24:50–64.

Orlando M, Thissen D. Further investigation of the performance of S- ^{2}: An item fit index for use with dichotomous item response theory models. Applied Psychological Measurement. 2003; 27:289–298.

Patrick CJ, Hick BM, Nichol PE, Krueger RF. A bifactor approach to modeling the structure of the psychopathy checklist-revised. Journal of Personality Disorders. 2007; 21:118–141. [PubMed: 17492917]

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: 2012.

Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. Journal of Personality Assessment. 2010; 92:544–559. [PubMed: 20954056]

Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Quality of Life Research. 2007; 16:19–31. [PubMed: 17479357]

Reise SP, Moore TM, Maydeu-Olivares A. Targeted bifactor rotations and assessing the impact of model violations on the parameters of unidimensional and bifactor models. Educational and Psychological Measurement. 2011; 71:684–711.

Reise SP, Scheines R, Widaman KF, Haviland MG. The Effects of Multidimensionality on Structural Coefficients in Structural Equation Modeling. Educational and Psychological Measurement. (in press).

Revelle, W. Psych: Procedures for psychological, psychometric, and personality research. R package version 1.1-10. 2012. Retrieved from http://personalityproject.org/r/psych.manual.pdf.

Rijmen, F. Efficient full information maximum likelihood estimation for multidimensional IRT models (Tech Report NO. RR-09-03). Princeton, NJ: Educational Testing Service; 2009.

Rijmen F. Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. Journal of Educational Measurement. 2010; 47:361–372.

Rindskopf D, Rose T. Some theory and applications of confirmatory second-order factor analysis. Multivariate Behavioral Research. 1988; 23:51–67.

Satorra, A.; Bentler, PM. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye, A.; Clogg, CC., editors. Latent variable analysis: Applications for development research. Thousand Oaks, CA: Sage; 1994. p. 399-419.

Schmid J. The comparability of the bi-factor and second-order factor patterns. Journal of Experimental Education. 1957; 25:249–253.

Schmid J, Leiman JM. The development of hierarchical factor solutions. Psychometrika. 1957; 22:53–61.

Simms LJ, Gros DF, Watson D, O'Hara W. Parsing the general and specific components of depression and anxiety with bifactor modeling. Depression and Anxiety. 2008; 25:E34–E46. [PubMed: 18027844]

Takane Y, de Leeuw J. On the relationship between item response theory and factor analysis of discretized variables. Psychometrika. 1987; 52:393–408.

Thomas ML. Rewards of bridging the divide between measurement and clinical theory: Demonstration of a bifactor model for the brief symptom inventory. Psychological Assessment. 2011; 24:101–113. [PubMed: 21767026]

Watson D, O'Hara MW, Simms LJ, Kotov R, Chmielewski M, McDade-Montez EA, Gamez W, Stuart S. Development and validation of the inventory of depression and anxiety symptoms (IDAS). Psychological Assessment. 2007; 19:253–268. [PubMed: 17845118]

Wirth RJ, Edwards MC. Item factor analysis: Current approaches and future directions. Psychological Methods. 2007; 12:58–79. [PubMed: 17402812]

Wherry RJ. Hierarchical factor solutions without rotation. Psychometrika. 1959; 24:45–51.

Wu, EJC.; Bentler, PM. EQSIRT – A User-Friendly IRT Program. Encino, CA: Multivariate Software, Inc; 2011.

Vanheule S, Desmet M, Groenvynck H, Rosseel Y, Fontaine J. The factor structure of the Beck depression inventory II: An evaluation. Assessment. 2008; 15:177–187. [PubMed: 18182478]

Yuan KH, Bentler PM. On chi-square difference and z tests in mean and covariance structure analysis when the base model is misspecified. Educational and Psychological Measurement. 2004; 64:737–757.

Yung YF, Thissen D, McLeod LD. On the relationship between the higher-order factor model and the hierarchical factor model. Psychometrika. 1999; 64:113–128.

Zang J, Stout W. The theoretical DETECT index of dimensionality and its application to approximate simple structure. Psychometrika. 1999; 64:231–249.

Zinbarg RE, Revelle W, Yovel I, Li W. Cronbach's alpha, Revelle's , and McDonald's H. Their relations with each other and two alternative conceptualizations of reliability. Psychometrika. 2005; 70:123–133.

Table 1

RCADS-15 Abbreviated Item Content and Proposed Disorder

1	SAD	Scared if I have to sleep on own
2	SAD	Afraid of on my own at home
3	SAD	Afraid of crowded places
4	GAD	Something will happen to family
5	GAD	Something bad will happen
6	GAD	I think about death
7	PD	Tremble or shake
8	PD	Suddenly become dizzy or faint
9	PD	Suddenly get a scared feeling
10	SOC	Worry when done poorly
11	SOC	Worry other people think of me
12	SOC	Fool of myself in in front of people
13	OCD	Have to think special thoughts
14	OCD	Do things over and over again
15	OCD	Do things in just the right way

Note SAD = sepration anxiety disorder, GAD = generalized anxiety disorder; PD = panic disorder; SOC = social phobia; OCD = obsessive-compulsive disorder

Table 2

RCADS-15 Pearson and Tetrachoric Correlation Matrices

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		.35	.25	.18	.21	.04	.16	.11	.24	.07	.13	.16	.23	.13	.18
2	.54		.28	.14	.21	.09	.19	.14	.26	.10	.11	.15	.24	.19	.21
3	.39	.44		.27	.18	.09	.24	.17	.24	.08	.09	.12	.26	.19	.21
4	.33	.27	1		.38	.28	.17	.10	.17	.14	.11	.19	.17	.13	.20
5	.36	.40	.33	.6		.33	.20	.16	.28	.15	.23	.19	.20	.17	.33
6	.07	.16	.16	.45	.45		.07	.11	.18	.10	.15	.17	.12	.09	.16
7	.26	.32	.38	.33	.33	.51		.34	.37	.06	.07	.07	.26	.20	.24
8	.20	.26	.30	.21	.32	.36	.34		.26	.02	.07	.08	.22	.08	.17
9	.38	.42	.39	.31	.48	.30	.44	.55		.10	.13	.17	.27	.21	.28
10	.11	.17	.14	.24	.25	.16	.03	.11	.57		.16	.15	.03	.01	.10
11	.23	.22	.17	.20	.39	.26	.17	.13	.26	.46		.46	.10	.04	.16
12	.29	.29	.22	.33	.33	.29	.15	.13	.29	.16	.70		.15	.09	.20
13	.36	.38	.41	.30	.34	.38	.37	.41	.19	.15	.17	.26		.29	.41
14	.21	.32	.32	.23	.27	.14	.14	.33	.41	.01	.07	.16	.44		.30
15	.28	.36	.35	.33	.51	.40	.31	.40	.51	.10	.16	.20	.34	.61	
M	.40	.38	.31	.77	.71	.67	.31	.21	.41	.59	.78	.78	.45	.55	.60
S	.49	.45	.46	.42	.46	.47	.46	.40	.49	.49	.41	.41	.50	.50	.49

Note Pearson correlations are above the diagonal and tetrachoric are below. Tetrachoric correlations were estimated using the *psych* library (Revelle, 2012).

Table 3

RCADS-15 Exploratory Five Correlated-Factors and Schmid-Leiman Bifactor Models.

Item	Five Correlated-Factors					Schmid-Leiman					
	SAD	GAD	PD	SOC	OCD	ANX	SAD	GAD	PD	SOC	OCD
1	**.74**	.04	-.01	.05	-.03	.48	**.58**	.03	-.01	.04	-.02
2	**.63**	.05	.04	.03	.09	.54	**.49**	.04	.03	.03	.05
3	**.37**	.01	.20	.00	.16	.50	**.29**	.01	.14	.00	.10
4	.12	**.65**	.02	-.04	.00	.49	.09	**.50**	.01	-.04	.00
5	.08	**.78**	.05	.01	.05	.63	.06	**.59**	.04	.01	.03
6	-.15	**.67**	-.02	.06	.01	.35	-.12	**.51**	-.02	.05	.01
7	.00	-.01	**.83**	-.03	.02	.57	.00	-.01	**.60**	-.03	.01
8	-.06	.02	**.69**	.05	-.01	.46	-.05	.01	**.50**	.05	-.01
9	.16	.14	**.50**	.07	.06	.61	.13	.10	**.36**	.06	.03
10	.05	.21	.00	**.23**	-.09	.21	.04	.16	.00	**.20**	-.06
11	-.03	.02	.05	**.85**	-.06	.37	-.02	.02	.03	**.76**	-.04
12	.07	-.02	-.05	**.80**	.10	.43	.05	-.02	-.03	**.71**	.06
13	.09	-.09	.08	.00	**.72**	.62	.07	-.07	.06	.00	**.44**
14	.09	.00	.01	-.09	**.55**	.46	.07	.00	.00	-.08	**.34**
15	-.08	.13	-.01	.06	**.77**	.66	-.06	.10	-.01	.06	**.47**
	.79	.65	.69	.45	.63						
φ	1										
	.39	1									
	.44	.44	1								
	.31	.46	.21	1							
	.51	.49	.59	.32	1						

Note. ANX = broad anxiety; SAD = sepration anxiety disorder, GAD = generalized anxiety disorder; PD = panic disorder; SOC = social phobia; OCD = obsessive-compulsive disorder

Table 4

RCADS-15 Target and Analytic Exploratory Bifactor Rotation Models.

Item	Target Bifactor Rotation						Analytic Bifactor Rotation					
	ANX	SAD	GAD	PD	SOC	OCD	ANX	SAD	GAD	PD	SOC	OCD
1	.51	**.62**	–.02	–.03	.04	–.03	.52	**.57**	.03	.04	.02	.01
2	.59	**.40**	–.04	.00	.02	.04	.57	**.45**	.08	.01	.01	.06
3	.54	**.22**	–.04	.10	–.04	.10	.50	**.28**	.02	.14	.03	.14
4	.56	.02	**.45**	–.05	–.01	–.06	.63	.04	**.77**	.04	.00	.07
5	.66	.03	**.48**	.03	.09	.00	.90	.21	**.00**	.16	.08	.28
6	.35	–.11	**.59**	.02	.12	.02	.47	.24	**.18**	.12	.09	.10
7	.62	–.07	–.11	**.56**	–.14	.00	.54	.02	.03	**.71**	.11	.07
8	.40	.03	.08	**.58**	.02	.07	.44	.02	.06	**.44**	.03	.05
9	.64	.09	.06	**.31**	.02	.05	.63	.10	.08	**.32**	.03	.05
10	.40	–.17	–.01	–.20	**.16**	–.30	.26	.01	.09	.02	**.19**	.10
11	.36	.01	.06	.03	**.87**	–.02	.44	.04	.11	.04	**.64**	.10
12	.45	.06	.09	–.08	**.61**	.05	.46	.05	.06	.06	**.81**	.08
13	.57	.11	–.03	.13	–.03	**.50**	.57	.09	.02	.09	.05	**.53**
14	.46	.03	–.04	–.01	–.11	**.37**	.43	.05	.02	.07	.08	**.37**
15	.66	–.09	.04	.00	.04	**.45**	.67	.11	.09	.01	.01	**.43**

Note. ANX = broad anxiety; SAD = sepration anxiety disorder, GAD = generalized anxiety disorder; PD = panic disorder; SOC = social phobia; OCD = obsessive-compulsive disorder

Table 5

Confirmatory Bifactor and Unidimensional SEM Parameter Estimates for RCADS-15

Item	1	GEN	GRP1	GRP2	GRP3	GRP4	GRP5	1	UNI
1	0.25	.52	.46					.72	0.25	.53
2	0.58	.59	.50					.62	0.58	.60
3	0.50	.56	.22					.80	0.50	.56
4	−0.74	.51		.50				.70	−0.74	.57
5	−0.54	.68		.49				.54	−0.54	.71
6	−0.42	.36		.54				.76	−0.42	.43
7	0.50	.57			.64			.51	0.50	.60
8	0.82	.46			.46			.76	0.82	.50
9	0.23	.68			.28			.67	0.23	.69
10	−0.22	.25				.23		.94	−0.22	.26
11	−0.79	.38				.78		.49	−0.79	.42
12	−0.78	.45				.67		.58	−0.78	.48
13	0.13	.61					.48	.64	0.13	.64
14	−0.12	.46					.33	.82	−0.12	.48
15	−0.25	.67					.44	.60	−0.25	.69
( )²		4.23	.51	.78	.70	1.11	.54			
( )²		60.06	1.39	2.34	1.90	2.82	1.56			

Note is a threshold, is a factor loading, and is a square root of the error variance. Sum of error variances is 7.09; $ECV = .54$; Coefficient $= .91$; Coefficient $H = .78$; $PUC = 86\%$.

Table 6

Confirmatory Bifactor and Unidimensional IRT Parameter Estimates for *RCADS-15*

Item		GEN	GRP1	GRP2	GRP3	GRP4	GRP5	UNI
1	−0.61	1.27	1.25				−0.49	1.05
2	−1.54	1.60	1.24				−1.25	1.33
3	−1.07	1.25	0.44				−1.03	1.20
4	1.82	1.29		1.21			1.54	1.23
5	1.72	2.17		1.55			1.33	1.76
6	0.95	0.80		1.23			0.78	0.79
7	−1.80	2.04			2.31		−1.09	1.34
8	−1.83	1.07			0.96		−1.64	1.07
9	−0.58	1.74			0.68		−0.55	1.66
10	0.38	0.42				0.37	0.37	0.43
11	2.93	1.41				2.96	1.47	0.84
12	2.37	1.32				1.94	1.52	1.00
13	−0.34	1.66					1.26	−0.28	1.43
14	0.24	0.91					0.68	0.22	0.91
15	0.71	1.89					1.26	0.58	1.61

Note: is an intercept, *GEN* is the slope for the general factor, *GRP*=1…5 are slopes for the five group factors, and *UNI* is the slope in the unidimensional model.

Table 7

Invariance Analysis of General Factor Loadings for *RCADS-15*

Item	Domain	GEN	GEN	GEN	GEN	GEN	GEN	GEN
1	SAD	.52	.54				.52	.51
2	SAD	.59	.59				.60	.54
3	SAD	.56	.56				.60	.46
4	GAD	.51	.58	.66				.66
5	GAD	.68	.71	.80				.77
6	GAD	.36	.44	.57				.53
7	PD	.57	.63	.51	.65	.66		
8	PD	.46	.53	.46	.55	.54		
9	PD	.68	.72	.62	.69	.70		
10	SOC	.25		.32	.20			.35
11	SOC	.38		.49	.36			.54
12	SOC	.45		.51	.43			.58
13	OCD	.61			.70	.70		
14	OCD	.46			.51	.53		
15	OCD	.67			.72	.67		

Note. SAD = sepration anxiety disorder, GAD = generalized anxiety disorder; PD = panic disorder; SOC = social phobia; OCD = obsessive-compulsive disorder. The *GEN* values are the Loadings on the general factor when different subsets of items are analyzed under a bifactor model.