Exploratory Bifactor Measurement Models in Vocational Behavior Research

Casey Giordano, Deniz S. Ones, and Niels G. Waller

University of Minnesota

Kevin C. Stanek

preValio, LLC

Casey Giordano, Deniz S. Ones, and Niels G. Waller, Department of Psychology, University of Minnesota; Kevin C. Stanek, preValio, LLC.

Correspondence concerning this article should be addressed to Casey Giordano or Deniz S. Ones, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455. E-mails: giord023@umn.edu or deniz.s.ones-1@tc.umn.edu.

This research was partially supported by the Auke Tellegen Fellowship in Applied Personality Assessment awarded to Casey Giordano.

Abstract

We provide an overview of and guidance for applying exploratory bifactor models to vocational research. First, we describe bifactor models and highlight their potential and actual applications in vocational psychology. Second, we review the theoretical bases of bifactor models and offer methodological guidance to correctly implement and interpret these models in practice. Third, we estimate a bifactor model in two vocational datasets to illustrate the concepts reviewed in this manuscript. The resulting models highlight novel insights in careers research (e.g., developmental performance feedback and personality [conscientiousness] modeling) that are made possible by leveraging bifactor measurement models. Overall, this manuscript provides a useful introduction to bifactor models to facilitate vocational behavior scholars and practitioners in thoughtfully producing and consuming bifactor models in their own research.

Exploratory Bifactor Measurement Models in Vocational Behavior Research

The field of psychology is in the midst of a bifactor model renaissance. Evincing this renaissance, Reise's (2012) rediscovery of bifactor models has quickly become a citation classic, amassing over 700 citations[1] in a few years—eclipsing that of Schmid and Leiman's (1957) seminal article more than half a century after its publication. This precipitous resurgence of bifactor models has spurred rapid methodological advancements—particularly in the domain of exploratory bifactor analysis (Giordano & Waller, 2020). Exploratory bifactor analyses can contribute to and support the refinement of multidimensional psychological theories dominating vocational behavior research of the past half century.

The purpose of this manuscript is to introduce exploratory bifactor analysis to the broader audience of vocational behavior researchers and to provide them with the necessary tools to apply bifactor models to their own work. To achieve these goals, we first provide a conceptual overview of bifactor models. Next, we illustrate the utility and broad applicability of exploratory bifactor models with an emphasis on career and vocational behavior research domains. Then, using two datasets, we demonstrate how a researcher might leverage bifactor models to answer important substantive questions. To facilitate the proper application of the ideas discussed in this manuscript, the online supplemental materials include the statistical code and data used to estimate exploratory bifactor models and interpret the results. Through this manuscript, we hope to aid vocational behavior researchers to correctly (a) understand the strengths and limitations of

[1] Reise (2012) has been cited 711 times according to Web of Science, accessed on February 18th, 2020.

exploratory bifactor estimation methods, (b) estimate exploratory bifactor models in their own research, and (c) interpret results from a bifactor measurement model.

Background, Applicability, and Applications of Bifactor Measurement Models

Background

Exploratory bifactor analysis refers to a class of models within the broader exploratory factor analytic domain (for historical reviews, see Carroll, 1993, Ch. 2; Giordano & Waller, 2020). At its core, factor analysis is a data-driven approach to modeling latent factors as determinants of observed data (e.g., responses to career satisfaction items). To use a linear regression analogy, latent factors are the independent variables that predict variation in the dependent variables (i.e., observed variables). Holzinger's (1935, 1936, 1937, 1945; Holzinger & Swineford, 1937) bifactor model was designed to account for variance in observed variables from the effects of three types of latent factors: (a) a general factor that influences all observed variables, (b) multiple group factors[2] that influence a subset of the observed variables, and (c) a set of uniqueness factors, each one of which captures variance unique to an observed variable. To visualize an example bifactor measurement model, see the path diagram in Figure 1. Conceptually, the bifactor model represents a marriage between Spearman's (1904) general factor model and Thurstone's (1934, 1947) multiple factors model. In most applications of the bifactor model, all latent factors are mutually orthogonal (uncorrelated; for an exception, see Jennrich & Bentler, 2012).

---

[2] Many bifactor applications call these factors "specific factors." In this paper, we defer to Holzinger's (1937) nomenclature of group factors because, in factor analysis, "specific factors" are a component of uniqueness factors.

Insert Figure 1 here

As a class of models within exploratory factor analysis, it is helpful to distinguish between two types of bifactor models. Namely, constrained, hierarchical bifactor models and unconstrained, non-hierarchical bifactor models. These model types differ in the dimensionality of (a) the bifactor loadings matrix and (b) the estimated factor scores. Although hierarchical and non-hierarchical bifactor solutions can produce similar patterns of factor loadings (Giordano & Waller, 2020), the aforementioned dimensionality differences have important ramifications for both theoretical interpretations and practical applications (we address these topics more fully in later sections).

Hierarchical and Non-Hierarchical Bifactor Models

The main difference between hierarchical and non-hierarchical bifactor models is the presence or absence of additional model constraints (Yung, Thissen, & McLeod, 1999; see also, Gignac, 2016). To understand the implications of these constraints, it is helpful to understand how a constrained, hierarchical bifactor model is estimated. Traditionally, hierarchical bifactor models have been conceptualized as re-expressions of higher-order, common factor models (e.g., Schmid & Leiman, 1957; Thomson, 1951; Thurstone, 1947). Higher-order factor analyses seek to explain first-order factor correlations (i.e., from an obliquely-rotated, correlated-factors model) by one or more higher-order factors. When one higher-order factor sufficiently accounts for the lower-order factor correlations, the higher-order factor is often called a general factor. Structurally, the general factor (or multiple general factors) putatively influences the lower-order latent factors, which in turn influence the observed variables. Stated differently, the general factor has an *indirect* (i.e., mediated) effect on each observed variable. When re-expressing a higher-order model as a hierarchical bifactor model, these indirect effects result in a constrained bifactor loadings matrix that is *rank-deficient* (Waller, 2018). In common parlance, this means that the general factor loadings are an exact linear combination of the group factor loadings. In contrast to hierarchical models, the general factor in non-hierarchical bifactor models has a *direct* (i.e., non-mediated) effect on each observed variable. Thus, non-hierarchical bifactor models produce unconstrained bifactor loadings matrices that are *full-rank* in the sense that the general factor loadings are (statistically and theoretically) independent from the group factor loadings.

By including hierarchical and non-hierarchical solutions under the bifactor moniker, this paper diverges from some researchers that only consider non-hierarchical bifactor solutions in their bifactor classification scheme (e.g., Murray & Johnson, 2013)[3]. We include both models under our bifactor umbrella for two reasons. First, we previously described bifactor models as those including (a) a general factor (or multiple general factors), (b) multiple group factors, and (c) uniqueness factors. No stipulations were made about constraints on the estimated bifactor loadings structure. Second, empirically ascertaining whether one's observed data adhere to a hierarchical or a non-hierarchical bifactor model can be exceedingly difficult (e.g., Giordano & Waller, 2020; Greene et al., 2019; Mulaik & Quartetti, 1997; Rindskopf & Rose, 1988, Yang, Spirtes, Scheines, Reise, & Mansolf, 2017). Related to this latter point, previous research has quantitatively examined the similarity of hierarchical and non-hierarchical solutions. For

[3] Several researchers use the term 'hierarchical bifactor model' to refer to what we call an unconstrained non-hierarchical bifactor model (e.g., Gignac, 2008, 2016). Whereas in this paper we use the term 'hierarchical' to reference a multi-order factor solution (i.e., a hierarchy of factors), the other use of 'hierarchical' refers to a breadth factor (i.e., a factor that influences many variables).

example, in a simulation comparing 162 non-hierarchical bifactor models to their closest hierarchical analog, Giordano and Waller (2020) found a median congruence coefficient[4] of $.995$ ($min = .97$). In other words, on average, the non-hierarchical bifactor pattern and its hierarchical counterpart were virtually indistinguishable in these simulations.

Whereas the empirical differences between (constrained) hierarchical and (unconstrained) non-hierarchical bifactor models may be small, the theoretical differences can be substantial. Namely, the relationship between latent factors and observed variables differs across the two models. In the constrained hierarchical model, the general factor's influence on the observed variables is mediated through the first-order factors. Alternatively, in the unconstrained non-hierarchical model, the general factor directly influences the observed variables. In substantive terms, these models differ in the causal pathways (representing mediated and direct effects) between factors and observed variables. Moreover, the two models also differ in how the group factors are defined. Specifically, group factors in the hierarchical model are residualized, lower-order factors that are created by partialling out the effects of the general factor from the original, correlated lower-order factors. In contrast to that approach, the group factors in the non-hierarchical model are directly defined to be orthogonal to the general factor (cf. Abad et al., 2017; Giordano & Waller, 2020).

Taken together, due to the combination of (a) minor empirical differences and (b) meaningful theoretical differences between the two models, we believe that deciding whether to estimate a hierarchical or a non-hierarchical bifactor model should not be based on fit indices

---

[4] A congruence coefficient (Lorenzo-Seva & ten Berge, 2006) is a common index for measuring the similarity between two factor solutions. Congruence coefficient values (that range from 0 to 1) over .95 are suggestive that the two solutions are functionally equivalent (Lorenzo-Seva & ten Berge, 2006).

alone. Rather, we recommend that researchers draw upon theoretical frameworks and domain knowledge when deciding between the two types of bifactor models (see also, Preacher, Zhang, Kim, & Mels, 2013). Exemplifying this approach,[5] Beaujean (2015) reviewed Carroll's (e.g., 1993) highly influential work on the structure of cognitive abilities and argued that the cognitive abilities domain is best represented by a non-hierarchical bifactor model (cf. McGrew, 2005 for an opposing view). Other researchers (e.g., Digman, 1997; Stanek & Ones, 2018) have suggested that many Big Five personality traits are consistent with higher-order factor models and, thus, could be profitably modelled by hierarchical bifactor models.

Fitting Exploratory Bifactor Models to Vocational Data

One reason for the renewed interest in both constrained and unconstrained bifactor measurement models is that both models produce orthogonal factors that are (often) easily interpretable. Among these orthogonal factors, in most substantive domains, the general factor often accounts for the lion's-share of the covariation among the observed variables (e.g., items, item clusters, scales, etc.). After partialling out the effects of the general factor, the residual correlations are presumably due to the group factors (Holzinger & Swineford, 1937). In vocational research, group factors could represent substantive constructs (e.g., facets of a construct), methodological factors (e.g., positive and negative valence from positively and negatively worded items), contextual domain effects (e.g., the same construct manifesting in work, educational, personal life), or temporal influences (e.g., developmental stages of a construct; aging effects) on the observed variables. To illustrate these ideas, Figure 1 shows how

[5] We provide these examples of hierarchical versus non-hierarchical representations purely for illustrative purposes. It is beyond the scope of this manuscript to make definitive claims about the appropriateness of specific bifactor models in substantive domains.

the bifactor model can be applied to account for the many potential sources of variance in vocational data.

To date, bifactor models have been overwhelmingly applied to datasets wherein group factors represent systematic construct variation beyond that of the general factor. This has been particularly true in the domain of cognitive ability (e.g., Carroll, 1993; Cucina & Byle, 2017). Though such study designs serve an important role in advancing our understanding of various construct domains, bifactor models can be fruitfully applied to a broader range of study designs. For example, Levin (1973) applied a bifactor model to a multitrait-multimethod study design where four leadership criteria were assessed by self, peer, and observer reports. Modeling the rating sources as group factors allows one to specifically parse out (a) the shared variance across rating sources (i.e., a general leadership factor) and (b) the unshared variance that is unique to each source (e.g., source-specific perspectives in ratings). The resulting bifactor model clearly showed that observer ratings are more strongly influenced by a general leadership factor. In contrast, self-ratings were influenced more heavily by source-specific perspective effects with a comparatively weaker influence from the general leadership factor.

Like the leadership domain, most, if not all, constructs in vocational research are multidimensional. A common yet important question in multidimensional variable domains is the relative strength of the general factor (e.g., the strength of a general satisfaction factor) versus the strength of more narrowly defined group factors (e.g., satisfaction with advancement, pay, meaningfulness, and work-life balance). Such questions are easily addressed via bifactor measurement models that partition latent factor variance into uncorrelated general and group factors. For example, a 2019 issue of the *Journal of Vocational Behavior*, examined this question in three studies pertaining to the Psychology of Working Theory (Duffy, Blustein, Diemer, &

Autin, 2016). These studies investigated, in three separate countries, how work can impact the fulfillment of one's basic human needs. In Italy, Portugal, and Brazil, the general 'decent work' factor accounted for 59%, 52%, and 65% of the total observed scale variance, respectively (Di Fabio & Kenny, 2019; Ferreira, et al., 2019; Ribeiro, Teixeira, & Ambiel, 2019). The cumulative effect of the 'decent work' subdimensions—that is, safe conditions, access to healthcare, adequate compensation, free time and rest, and complementary values—accounted for 36%, 41%, and 30% of the variance for Italy, Portugal, and Brazil, respectively. Thus, whereas the 'decent work' subdimensions are all positively correlated (i.e., a strong general factor is present), there is meaningful differentiation among the group factors. We note that none of the three studies theoretically justified whether a constrained (hierarchical) or unconstrained (non-hierarchical) bifactor model would fit better.

Another application of bifactor modeling in career and vocational psychology examined the differentiation of occupational interests (Toker & Ackerman, 2012). Specifically, Toker and Ackerman were concerned with science, technology, engineering, and mathematics (STEM) students and investigated how STEM students differ in their interest for complex careers. These authors applied a bifactor model with a general 'complexity interest' factor as well as group factors representing numerical complexity, symbolic complexity, spatial complexity, and idea complexity. Further analysis of the original factor solution (see our online supplement) found that the general 'complexity interest' factor accounted for 80% of the total observed score variance whereas the remaining group factors collectively accounted for 18% of the total observed variance. Simply put, in this example, 80% of the variance in an observed scale-score was comprised of general factor variance. Some authors would claim that values in this range are *prima facie* evidence for a unidimensional measure (Rodriguez, Reise, & Haviland, 2016).

Importantly, however, group factors have stronger effects in their associated subscale scores but these effects shrink in total scale scores by nature of (a) adding items from unrelated group factors and (b) a general factor that impacts all items.

The ‘decent work’ and ‘occupational complexity interest’ domains are only two demonstrations of the utility for applying bifactor measurement models. Many prominent variables and criteria of interest are also multidimensional: job performance (e.g., J. P. Campbell & Wiernik, 2015; Viswesvaran & Ones, 2000), organizational citizenship (e.g., LePine, Erez, & Johnson, 2002), transformational leadership (e.g., Judge & Piccolo, 2004), emotional labor (e.g., Morris & Feldman, 1996), burnout (e.g., Demerouti, Bakker, Varkadou, & Kantas, 2003), job satisfaction (e.g., Locke, 1969), employability (e.g., Fugate, Kinicki, & Ashforth, 2004), career success (e.g., Arthur, Khapova, & Wilderom, 2005), quality of life (Chen, West, & Sousa, 2006), and career adaptability (e.g., Zacher, 2014), among numerous others. Likewise, assessments of key explanatory variables are often multidimensional and are thus well represented by bifactor measurement models. Examples include many personality assessments (e.g., McCrae & Costa, 2004; Stanek & Ones, 2018), interest measures (e.g., D. P. Campbell & Holland, 1972), cognitive ability tests (e.g., Carroll, 1993), affect scales (e.g., Watson, Clark, & Tellegen, 1988), situational/contextual characteristics (e.g., Rauthmann, Gallardo-Pujol, Guillaume, et al., 2014), inventories of occupational constraints and demands (e.g., Karasek, 1979), measures of organizational support constructs (e.g., Rhoades & Eisenberger, 2002), and many others. Each of these domains are well suited for bifactor modeling.

The extant literature suggests that bifactor measurement models can aid in understanding subdimensions of hierarchical construct domains. However, there are many other ways in which to conceptualize and model the group factors in a bifactor model (see Figure 1). For example, group factors can represent method-specific effects in a multimethod study design (e.g., Levin, 1973; McAbee & Connelly, 2016), an approach we take in Example 1 below. Other method effects, such as positively and negatively worded items, can be modeled to partition variance into a substantive general factor of the focal construct and group factors associated with the potentially contaminating effects of item keying (e.g., "I enjoy my work environment" and "I loathe my work tasks" are oppositely-keyed items of job satisfaction). Developmental effects might also be modeled with bifactor models. Consider measures of vocational interests in adolescence, adulthood, and older age. Longitudinal interest data from these developmental stages can be modeled to identify, for example, a general social interest factor alongside life-stage-limited social interests (i.e., group factors corresponding to each life stage). Life domain or context effects can also constitute group factors (e.g., Stanek, Ones, & McGue, 2017). To provide another example in the interest domain, previous research has found that "vocational, leisure, and family interests of adults are strongly intercorrelated" (Gaudron & Vautier, 2007, p. 568), even after accounting for a common methods factor. When applied to, say, realistic interests (e.g., D. P. Campbell & Holland, 1972), a bifactor model could provide insights into the amount of variation that is due to the global realistic interest factor as well as specific group factors, such as realistic vocational, realistic leisure, and realistic family interests. Does the general 'realistic interest' factor account for the most variance or do people meaningfully differentiate their interests according to specific contexts? Here, a bifactor model can be leveraged to advance developmental and individual difference theories of vocational interests. These are just a few examples of novel bifactor applications to address unanswered substantive questions in psychological domains.

Implementing Bifactor Measurement Models: Methodological Decisions and Their Consequences

Estimating Exploratory Bifactor Models

Like other multivariate analyses, a bifactor analysis requires numerous methodological choices that can influence the quality of the obtained solution. Although some choices might not meaningfully alter the obtained pattern of bifactor loadings, other choices during model estimation can prominently impact obtained bifactor solutions. Here we highlight a few key methodological decisions and issues relevant to bifactor modeling.

Deciding the Number of Latent Factors to Model. An influential, early decision in the bifactor modeling process is deciding on the number of latent factors to model (Preacher, Zhang, Kim, & Mels, 2013). The consequence for misidentifying the number of latent factors results in one of two errors: (a) over-extraction (i.e., extracting and modeling too many factors) and (b) under-extraction (i.e., extracting and modeling too few factors). Over-extraction yields less parsimonious solutions that tend to split meaningful factors into two or more weakly-determined factors (Auerswald & Moshagen, 2019; Fava & Velicer, 1992)[6]. Importantly, the detrimental effects of over-extraction are exacerbated as factor loadings and sample sizes decrease (Fava & Velicer, 1992). In contrast to correlated-factors models—wherein items typically load onto one factor—bifactor models tend to have lower factor loadings because an item's primary loading is bifurcated into loadings on a general factor and one or more group factors. Compared to over-extraction, under-extraction leads to more severely biased factor loadings, which has

[6] This effect is demonstrated in Example 2 later in the manuscript, where extracting a third group factor cleaved the 'prudent work-orientation' factor into two separate factors (i.e., prudence and work-orientation).

downstream effects, such as distorting the estimated factor scores (Wood, Tataryn, & Gorsuch, 1996).

Prior to conducting a factor analysis, it is recommended that researchers jointly consider theoretical perspectives and empirical procedures for determining the number of latent factors to retain (Preacher, Zhang, Kim, & Mels, 2013). Theoretical insights into a variable domain help decide how to model its structure (e.g., hierarchical versus non-hierarchical) and may even give a plausible range for the number of factors to extract (e.g., five personality factors in the Big Five model of personality; e.g., Digman, 1997). Empirical procedures are data-driven approaches to determine an optimal number of latent factors to model. However, different empirical procedures applied to the same dataset often result in different suggestions—this is exemplified in both datasets later in the manuscript. Moreover, a recent simulation study of dimensionality assessment found that “no single approach displayed the highest accuracy in all conditions” (Auerswald & Moshagen, 2019, p. 487).

When estimating the dimensionality of a dataset, researchers should seek converging evidence from theoretical insights and multiple empirical procedures (Auerswald & Moshagen, 2019). To estimate the number of factors to model, most methods implement a decision rule based on eigenvalues—properties of the sample-based correlation matrix (e.g., Braeken & van Assen, 2017). The most popular approach—and the default for many programs, such as SPSS—is to retain all factors associated with eigenvalues greater than one. Although popular, this decision rule has low accuracy and frequently leads to over-extraction (Auerswald & Moshagan, 2019; Cliff, 1988; Hayton, Allen, Scarpello, 2004, Preacher, et al. 2013). A recent and related method relies on the theoretical sampling distributions of eigenvalues to improve the ‘eigenvalues greater than one’ rule. This approach is named the Empirical Kaiser Criterion

(EKC; Braeken & van Assen, 2017). Aside from EKC, two other methods can accurately detect the correct number of factors to retain in multiple factor models. They are the parallel analysis (PA; Hayton et al., 2004; Horn, 1965) and comparison data (CD; Ruscio & Roche, 2012) techniques. Briefly, these methods compare sample-based eigenvalues to eigenvalues obtained from computer-generated datasets. Specifically, PA generates random data with no underlying factor model (i.e., a null model) whereas CD generates non-random data with an underlying factor model that is comparable to the sample-based data. In general, if two of these methods (e.g., EKC, PA, and CD) agree on the number of latent factors, there is a good chance they have converged on the correct number of factors (Auerswald & Moshagen, 2019).

Deciding Which Exploratory Bifactor Procedure to Use. With the rapid advancements in exploratory bifactor analysis, researchers have numerous methodological options at their disposal for estimating a bifactor solution. For simplicity, the competing methods can be distinguished on two dimensions (see Table 1). The first dimension is the analytic strategy used (i.e., *how* a bifactor pattern is obtained) with categories of (a) hybrid approaches, (b) target rotations, and (c) analytic bifactor rotations. In this context, hybrid approaches are generally conducted in two stages to obtain bifactor parameter estimates. For instance, in the Schmid-Leiman (SL; Schmid & Leiman, 1957) method, one first conducts a higher-order factor analysis and then re-expresses the higher-order parameter estimates into a constrained (hierarchical) bifactor pattern. Similarly, target rotation methods have applied either partially- (Abad et al., 2017; Browne, 2001) or fully-specified (Waller, 2018) target matrices (i.e., a factor rotation toward a supplied target structure, like a bifactor structure) to obtain a bifactor solution. Lastly, analytic bifactor rotations (Jennrich & Bentler, 2011, 2012, 2013) can be used to rotate the factors from an exploratory factor analysis directly to a bifactor pattern.

The second dimension characterizing bifactor estimation methods (Table 1) concerns the type of model that is ultimately obtained. These procedures can be divided into those that estimate either a (constrained) hierarchical or (unconstrained) non-hierarchical bifactor model. Although a thorough discussion of each method and its underlying mechanics is beyond the scope of this manuscript (cf. Abad et al., 2017; Giordano & Waller, 2020), in what follows we briefly describe the popular approaches—highlighting their benefits and drawbacks—for estimating exploratory bifactor measurement models.

To estimate exploratory bifactor measurement models, a prominent analytic strategy is that of analytic rotations. Until recently, no analytic rotations (e.g., varimax, oblimin, promax) were capable of directly estimating a bifactor solution. Recent authors have addressed this gap by extending the quartimin and geomin rotation criteria to recover non-hierarchical (i.e., unconstrained and full-rank) bifactor models (Jennrich & Bentler, 2011, 2012, 2013). These rotations are known as the bifactor quartimin and bifactor geomin rotations; These rotations should not be confused with their non-bifactor analogues that are intended to find simple structure in the traditional factor analysis paradigm. Unfortunately, two comprehensive studies have found that bifactor quartimin and bifactor geomin rotations are among the least accurate methods for estimating exploratory bifactor measurement models (Abad et al., 2017; Giordano & Waller, 2020).

The SL procedure—and its modern cognate, the Direct Schmid-Leiman (DSL; Waller, 2018) procedure—provides an alternative to estimating a bifactor model by an analytic rotation. The SL and DSL procedures both estimate a hierarchical (i.e., constrained and rank-deficient) bifactor model. Most applications of the SL procedure transform a second-order model with one general factor but there is no theoretical limit to the number of higher-order levels that can be transformed (Schmid & Leiman, 1957; Yung, Thissen, & McLeod, 1999). As a consequence, SL can estimate general factors at different hierarchical levels when a sufficient number of lower-order factors exist to ensure model identification.[7] For example, solutions with one general factor need at least three lower-order factors to yield an identified second-order solution (Ledermann, 1937). Notably, this shortcoming of the SL method is not shared with the DSL procedure because DSL utilizes a target rotation. Thus, if one's data are best represented by two group factors, a DSL approach will become the optimal estimation method because the SL approach will yield biased parameter estimates in the bifactor measurement model.

Best Performing Exploratory Bifactor Analysis Methods. Of the available methods to estimate exploratory bifactor measurement models, three methods seem to outperform the rest: (a) SL, (b) DSL, and (c) iterated Schmid-Leiman target rotation (SLi; Abad et al., 2017). In a comprehensive Monte Carlo simulation (Giordano & Waller, 2020), SL and DSL were best able to recover *both* hierarchical and non-hierarchical population models. These two methods, however, were not equally accurate in recovering bifactor solutions. In simplified terms, comparing SL versus DSL is akin to the optimal-weight versus unit-weight argument from the multiple regression literature (e.g., Schmidt, 1971). Namely, DSL applies unit weights to obtain a bifactor pattern and thus should be superior to SL in terms of cross-validation accuracy when sample sizes are small (e.g., $n < 500$). Alternatively, SL obtains a hierarchical bifactor pattern

---

[7] To check whether a sufficient number of variables (e.g., test items, lower-order factors) exist to produce an identified model in exploratory factor analysis, Ledermann's (1937) inequality can be applied. Specifically, let $k \leq \frac{2p+1-\sqrt{8p+1}}{2}$ where k is the maximum number of factors that are identified and p is the number of observed variables. This formula is symmetric such that the minimum number of variables, p, needed to identify the number of latent factors, k, is quantified where $p \geq \frac{2k+1+\sqrt{8k+1}}{2}$. See also the 'Ledermann' function in the *fungible* R library (Waller, 2019).

through optimal weights and therefore is more accurate than DSL when sample sizes are large (e.g., $n > 500$). SLi, as a method yielding non-hierarchical and unconstrained bifactor patterns, often surpasses SL and DSL in large samples when cross-loadings are present (e.g., Figure 17 of the supplemental materials in Giordano & Waller, 2020). Nevertheless, a prominent limitation of the SLi method is its notable tendency for finding bifactor patterns that markedly diverge from the true population values—even in large samples (e.g., $n = 2{,}000$). Taken together, SL should be applied in studies with large sample sizes and DSL should be applied in studies with small sample sizes. SLi can be applied under both conditions but researchers should be aware of its tendency to produce nonsensical solutions with some datasets (see Figures 1 and 2 and the online supplement of Giordano & Waller, 2020).

Limitations of Target Rotations in Exploratory Bifactor Analyses. The generally good performance of target rotations when estimating bifactor models comes with an important caveat—namely, target rotations often find a desired structure (e.g., a bifactor structure) regardless of the data generating model (e.g., Hurley & Cattell, 1962). In the present context, if the data generating model is an orthogonal factor pattern *without* a general factor, the DSL (and related methods, such as the Direct Bifactor) method will likely find (erroneously) a bifactor pattern with a general factor. This shortcoming of bifactor target rotations (i.e., DSL) is not shared by the SL approach.

Limitations of the Schmid-Leiman method. Whereas target rotation methods for estimating bifactor models can produce misleading results if the data do not adhere to a bifactor structure, the SL method has its own drawback. When estimating the higher-order model within the SL procedure, the factor structure can be obliquely rotated an infinite number of ways without changing the fit of the estimated solution. In the exploratory factor analysis literature this issue is known as rotational indeterminacy (Mulaik, 2010, ch. 10). Different oblique rotations (cf. Browne, 2001) apply different criteria for finding simple structure (Thurstone, 1947) pattern matrices, and thus, different rotation methods can produce notably different factor correlation matrices. All else being equal, rotations that yield larger factor correlations will find stronger general factor saturations in an SL transformation. To illustrate the practical implications of rotational indeterminacy[8], we applied 1,001 different oblique rotations (cf. Crawford & Ferguson, 1970; see also, Browne, 2001) to the dataset from Example 1. Each rotation was plotted against the estimated general factor saturation from an SL procedure (see Figure 2). Simply put, rotations that seek factor loadings patterns in which each variable loads onto as few factors as possible (i.e., minimizing variable complexity) will often fail to recover indicator cross-loadings. Consequently, when estimated cross-loadings are biased towards zero the estimated factor correlations are upwardly biased, as is the estimated general factor saturation. Thus, oblique rotations that do not penalize cross-loadings—such as (non-bifactor) geomin (Hattori, Zhang, & Preacher, 2017; Yates, 1987)—may be preferred.

Insert Figure 2 here

Interpreting Exploratory Bifactor Models

[8] Another potential implication of rotational indeterminacy in the SL procedure is the downstream effect on other methods incorporating SL procedures to obtain bifactor measurement models. Specifically, the SLi and SLt methods initiate estimation using starting values obtained from an SL solution. Thus, differences in the SL starting values (i.e., SL solutions from different oblique rotations) may result in differences in the final parameter estimates.

Once a bifactor model is estimated, researchers can begin to interpret the estimated parameters. In this section we briefly discuss how bifactor models can be leveraged to better understand the underlying structure of multidimensional data in two ways. First, in the context of bifactor measurement models (Rodriguez et al., 2016), we describe indices designed to assess the relationships between observed variables (e.g., scale items, homogenous item parcel scores) and latent factors. Second, we introduce difficulties that are unique to bifactor models in relating estimated latent factors to external variables (i.e., using factor scores; Grice, 2001).

Relating Latent Factors to Observed Variables. A bifactor measurement model is a useful tool for partitioning variance into uncorrelated latent factors, particularly when modeling multidimensional indicators in a given construct domain (Reise, 2012; Rodriguez et al., 2016). The utility of bifactor models—compared to correlated-factor models—is readily apparent in viewing the estimated factor loadings. All factor loadings estimates are regression coefficients relating the latent factors (i.e., the independent variables) to the observed variables (i.e., the dependent variables). Much like in multiple regression with uncorrelated independent variables, in a bifactor solution with uncorrelated factors, these regression coefficients are equivalent to zero-order correlations (Holzinger, 1937). Moreover, squared factor loadings in a bifactor model (i.e., squared correlation coefficients) represent the proportion of variance in the observed variable that is accounted for by a given factor. Alternatively, in a correlated-factors model, the factor loadings are standardized regression weights (i.e., not zero-order correlations) and must therefore be interpreted as such. In short, factor loadings in bifactor models are simpler to interpret.

In orthogonal models, the relation between one variable and one factor is captured by the factor loading. To represent the collective effect of the general and group factors on a given variable, a researcher can calculate the communality (h^2) for each variable. With standardized factor loadings, communality values reflect the proportion of observed variable variance that is collectively due to the common factors. The remaining variance (captured by the uniqueness factors) is a combination of measurement error and specific factor influences that is not shared with other variables. From the communality values, another closely related index can be calculated to understand the dimensionality of the obtained bifactor solution. Namely, an item's explained common variance (I-ECV) index (Reise et al., 2010; Rodriguez et al., 2016; see also, ten Berge & Sočan, 2004). In essence, I-ECV represents the proportion of item communality that can be ascribed to the general factor. When examined in tandem, h^2 and I-ECV values let a researcher see (a) how saturated each item is with group and general factor variance and (b) how much of that latent factor saturation is due to the general factor. In other words, these indices provide useful insight into the dimensionality of each item in a bifactor measurement model. In later sections, we estimate bifactor models in two datasets to illustrate the computation and interpretation of the h^2 and I-ECV indices.

Relating Latent Factors to Observed Scale Variance. Whereas the previous section described various methods to conceptualize the relationship between each observed variable and one or more common factors, this section is concerned with indices that assess how factors account for variance in the summed (standardized) scale scores. To illustrate the difference between these ideas, recall Toker and Ackerman's (2012) examination of complexity interests in STEM students. Whereas h^2 and I-ECV will quantify factor saturation for any given scale item, we need different indices to quantify factor saturation across item combinations (such as items forming a subscale). These indices represent model-based reliability indices (e.g., Rodriguez et al., 2016; Zinbarg, Revelle, Yovel, & Li, 2005). Note that we use the term 'model-based reliability' to differentiate these measures from traditional reliability indices (i.e., the ratio of true score variance to observed score variance). Specifically, model-based reliability focuses on aspects of the true scores that are due to the common factors (for more details, see Rodriguez et al., 2016).

One of the more prominent model-based reliability indices that is based on common factor models is called coefficient omega (ω; McDonald, 1999; Rodriguez et al., 2016; Zinbarg et al., 2005; see also Ferrando & Lorenzo-Seva, 2018). This index represents the ratio of *common* factor variance (i.e., aggregated across the general and group factors) to the observed variance of the *unit-weighted* total score (computed from standardized item scores). Although unit-weighted sum scores are a suboptimal method for estimating factor scores (Grice, 2001; Grice & Harris, 1998), they are the most commonly applied method for estimating factor scores. Because ω is interpreted in the context of unit-weighted sum scores, it is therefore well-suited for applications of bifactor measurement models that rely on unit-weighted scores (for a comparable model-based reliability index using optimally-weighted scoring, see Ferrando & Lorenzo-Seva, 2018; Rodriguez et al., 2016).

Variations of ω can also be computed to better understand how individual factors or a combination of factors relate to the sum scores (e.g., reflecting a subscale; cf. Rodriguez et al., 2016). For instance, omega hierarchical ω_h reflects the proportion of the total observed score variance that is attributed to the general factor. Thus, the square-root of ω_h represents the correlation between the general-factor factor scores and the observed sum scores (when the item scores have been standardized). Moreover, the ratio of ω_h over ω indicates how much latent factor variance (i.e., general and group factor variance) is due to the general factor. As this latter value approaches 1.0 (its maximum) the estimated model approaches a unidimensional structure.

Another notable modification to ω is called ω hierarchical subscale (ω_{hs}). This index represents the unique portion of *subscale* score variance that is due to the associated group factor. Importantly, when computing ω_{hs}, the bifactor loadings matrix is subset to only include those variables that are included in the subscale of interest. Taken together, ω and its cognates inform a researcher on the relative strength of factors in relation to (either overall or subscale) observed scores. For a review of these indices, the reader may consult Rodriguez et al. (2016).

Relating Latent Factors to External Variables. If a researcher is interested in relating factors from an exploratory, non-hierarchical bifactor measurement model to an external variable, they must rely on estimated factor scores as imperfect proxies of the true factor scores (Grice, 2001; Grice & Harris, 1998; Tucker, 1971). Importantly, as demonstrated by Steiger (1979), correlations between true factor scores and an external variable can differ markedly from the associated correlations obtained when using estimated factor scores. Unfortunately, as described more fully below, the use of estimated factor scores from hierarchical bifactor models is fraught with challenging psychometric obstacles.

Estimated factor scores represent an individual's predicted score on each of the modeled factors (e.g., a person's level of general cognitive ability on a cognitive ability test). Importantly, the most pervasive application of estimated factor scores is when a researcher sums all items exhibiting salient loadings on a particular factor (e.g., factor loadings $\geq |.30|$). These unit-weighted scores fail to consider that (a) some variables are better indicators of the latent factors than other variables (i.e., differences in their factor scoring weights; Grice, 2001) and (b) some variables are influenced by multiple group factors causing inflated correlations among the factor score estimates (due to the correlated error variance that results from using unit-weighted estimates). Consequently, unit-weighted factor score estimates "may be highly correlated even when the factors are orthogonal and they will be less valid representations of the factors in comparison with the refined factor scores [e.g., Thurstone's regression-based estimates]" (Grice, 2001, p. 434). Unit-weighted sum scores are therefore generally considered poor estimates of factor scores (Grice, 2001; Grice & Harris, 1998) unless researchers are working with small samples.

It merits comment that there is virtually no literature on estimating factor scores for constrained hierarchical bifactor models. Thus, in this section, we illustrate some problems in estimating factor scores that are unique to constrained (hierarchical) exploratory bifactor models. To understand these problems, it is informative to first consider the difference between an individual's *true* factor score and their *estimated* factor score. Theoretically, all individuals have a true standing on all latent factors (e.g., their cognitive ability scores, realistic interest scores, job satisfaction scores), although their exact standing is both unknown and unknowable in research contexts. Consequently, these factor scores must be estimated. Unfortunately, differences between true and estimated factor scores can be large (e.g., when few items define a factor and factor loadings are low; Guttman, 1955). When this occurs, the correlations between the *estimated* factor scores and external criteria may present a distorted picture of how the *true* factor scores relate to the external criteria (Steiger, 1979).

In bifactor measurement models, estimated and true factor scores can differ in multiple ways. One important divergence occurs in constrained hierarchical bifactor models. Specifically, because the factor loadings in these models are rank-deficient, the estimated factor scores (with the exception of unit-weighted scores; but see Table 4 for cautionary notes on using unit-weighted scores in hierarchical bifactor models) are also rank-deficient. Moreover, due to this property, some factor score estimates (e.g., ten Berge, Krijnen, Wansbeek, & Shapiro, 1999) cannot be calculated. In constrained hierarchical bifactor models, the estimated loadings on any factor (i.e., general or group) can be perfectly predicted from the estimated factor loadings from the remaining factors. Moreover, due to the rank-deficiency of the factor loadings matrix, the estimated factor scores on any factor can be perfectly reproduced from the estimated scores on the remaining factors. This problem of perfect collinearity has two practical ramifications when relating estimated bifactor scores (from constrained, hierarchical models) to external variables. First, due to the constraints in the bifactor loadings pattern, factor scoring methods cannot yield uncorrelated factor score estimates. Thus, although both the constrained (hierarchical) and unconstrained (non-hierarchical) bifactor models are composed of orthogonal factors, the estimated group and general factor scores in the former model will necessarily be correlated. Second, statistical analyses with estimated factor scores (e.g., via Thurstone's or Harman's method; cf. Grice, 2001) from constrained, hierarchical bifactor models may be inestimable due to the multicollinearity of the estimated scores. For example, multiple regression models with estimated factor scores from constrained, hierarchical bifactor models as predictors cannot isolate the unique effects of the (theoretically orthogonal) predictors (i.e., the estimated general and group factor scores) due to the aforementioned rank-deficient property. Non-hierarchical bifactor models do not include these problematic constraints and thus their estimated factor scores will not be collinear (i.e., perfectly correlated) in empirical applications. Note that in (unconstrained) non-hierarchical bifactor models, it is possible (though not always desirable) to compute orthogonal estimated factor scores (e.g., ten Berge et al., 1999; see also, McDonald & Burr, 1967; Tucker, 1971).

Estimated factor scores in all bifactor models have several drawbacks that merit consideration. The most salient of these drawbacks is the problem of factor score indeterminacy

(Guttman, 1955; Steiger, 1979; Wilson, 1928). Simply put, factor score indeterminacy means that factor scores cannot be uniquely calculated, although they can be uniquely estimated (Wilson, 1928). In more simple terms, "for any factor scores…satisfying the factor model, there exists also a different set of factor scores…, which also satisfy the model" (Steiger & Schönemann, 1978, p.151). In practice, not only are true factor scores unknowable, estimated factor scores from one method can differ from those obtained by another method (Grice & Harris, 1998). For instance, unit-weighted factor scores can produce notably different estimates than those obtained from other factor score estimation methods (cf. Grice, 2001).

Example Explorations of Bifactor Models in Vocational Behavior

In the previous section, we reviewed several important decisions (and their consequences) when fitting bifactor models to vocational data. In this section, to exemplify the concepts described previously, we fit constrained bifactor models to two vocational behavior datasets. In the first example, we illustrate a bifactor model of rater effects in the measurement of developmental performance feedback ratings (Hoffman, Lance, Bynum, & Gentry, 2010). In the second example, we illustrate a bifactor model to better understand the dimensional structure of conscientiousness (e.g., Hogan & Ones, 1997; Roberts, Chernyshenko, Stark, & Goldberg, 2005).

Example 1: Bifactor Modeling Rater Effects in Developmental Performance Feedback

Organizations often provide developmental performance feedback to employees using 360° evaluation systems (i.e., collecting ratings of a focal individual from multiple unique perspectives; Craig & Hannum, 2006). Such ratings are frequently used in employee development efforts (Smither, London, & Reilly, 2005). Multirater feedback systems rest on the premise that raters from different perspectives provide complementary insights into the performance of the ratees. Thus, a bifactor model is the perfect vehicle for disentangling a general performance factor from group factors reflecting rater-specific vantage points (e.g., supervisor, peer, subordinate, and self). To illustrate this idea, we reanalyzed published multisource, developmental ratings of managerial performance (Hoffman, et al., 2010).

Sample Description. To provide developmental feedback about a manager's performance, Hoffman and colleagues (2010) obtained data from a multisource performance feedback assessment tool called BENCHMARKS (Lombardo, McCauley, McDonald-Mann, & Leslie, 1999). Managers were rated on scales measuring three performance dimensions: (a) technical performance, (b) interpersonal performance, and (c) leadership. Ratings were obtained from the following sources: (a) two supervisor ratings, (b) two peer ratings, (c) two subordinate ratings, and (d) self-ratings. In total, 22,420 managers were assessed with a combined total of 156,940 raters. Hoffman et al. (2010, p. 129-130) described the managerial sample as consisting "primarily of White (76%) male (64%) college graduates (88%)" with an average age of 42. The BENCHMARKS instrument included 115 items. For these analyses, we used aggregated scale-level data with one rater per source yielding 12 scores (4 rater perspectives [sources] × 3 scales [performance dimensions] = 12 factor indicators). To align our research with applied best practices to minimize interrater measurement error (Ones, Viswesvaran, & Schmidt, 2008; Viswesvaran, Ones, & Schmidt, 1996), sources with multiple raters were combined into composites (e.g., both supervisor ratings were composited into a general supervisor rating).

Bifactor Modeling. To estimate a bifactor measurement model, we employed a mixture of rational/theoretical and empirical modeling strategies. This approach originates from the contemporary philosophies of factor analysts wherein "model selection is not intended to find the true model but rather is intended to find a parsimonious model that gives reasonable fit"

(Preacher, Zhang, Kim, & Mels, 2013, p. 52). Empirically, we relied on scree plots (Cattell, 1966) and the Empirical Kaiser Criterion (EKC; Braeken & van Assen) to help identify a plausible number of factors to retain. Both methods suggested the presence of four factors. Among the two types of bifactor models that have been discussed in this manuscript, we fit a constrained (hierarchical) bifactor model as this model is better aligned with the hierarchical relations among the developmental ratings.

The predicted dimensionality of the Hoffman et al. (2010) performance data is easily surmised. All raters assessed managers on the same, highly-correlated performance dimensions (see the online supplement). Thus, the four rating perspectives should be correlated to the extent that they all measure managerial performance. Moreover, each rating may be associated with systematic variance that is unique to each rating source (e.g., rating biases, unique performance insights). These combined influences on managerial performance can be modeled as a second-order factor model with four correlated factors (rating sources). These correlated factors are in turn influenced by a higher-order, general (performance) factor. As described by Schmid and Leiman (1957; see also Thomson, 1951; Thurstone, 1947), one can transform this higher-order model into a constrained, hierarchical bifactor model with a single general factor and four orthogonal group factors, each representing a rating perspective effect. In other words, the correlations among ratings of managerial performance are a function of (a) the manager's true general performance and (b) idiosyncratic perspective effects (e.g., boss or subordinate perspectives). Given the uncharacteristically large sample size for these data, we applied the SL (Schmid & Leiman, 1957) procedure to generate a constrained, hierarchical bifactor model of managerial performance.[9] To aid in the interpretation of this model, we computed communality (h^2) values, I-ECV indices, and several variants of coefficient ω (McDonald, 1999; Rodriguez et al., 2016; Zinbarg et al., 2005).

Results. Table 2 contains the estimated bifactor measurement model for the developmental performance ratings (Hoffman et al., 2010). Note that this model included one general and four group factors. The number of group factors is consistent with the recommendations of the scree and EKC plots. The results shown in Table 2 suggest that these group factors represent perspective effects (i.e., boss, peer, subordinate, and self-rated effects) on the managerial ratings.

Insert Table 2 here

As shown in Table 2, the factor loadings (λ) on the general performance dimension were substantially lower ($.21 \leq \lambda \leq .59$) than the primary loadings for the rater-perspective factors: boss ratings ($.76 \leq \lambda \leq .86$), peer ratings ($.65 \leq \lambda \leq .78$), subordinate ratings ($.63 \leq \lambda \leq .77$), and self-ratings ($.78 \leq \lambda \leq .90$). As expected, there are virtually no cross-loadings present in the estimated bifactor model. Interestingly, ratings by peers and subordinates produced factor loadings on the general performance factor ($.51 \leq \lambda \leq .59$) that were systematically larger than those generated by either the boss ($.44 \leq \lambda \leq .48$) or self-report ratings ($.21 \leq \lambda \leq .28$). Moreover, within each rating perspective, there was a consistent trend in relative factor loading sizes:

[9] When estimating the constrained bifactor model, we extracted unweighted (ordinary) least squares factor loadings. The first-order factor solution was subsequently rotated using an oblique geomin rotation from 100 random starting configurations (cf. Rozeboom, 1992) and a geomin tuning parameter set to .01 (cf. Hattori et al., 2017).

technical performance > interpersonal performance > leadership. Here, the reader should recall that bifactor loadings can be interpreted as correlations. Therefore, technical performance ratings are more highly correlated with both the general performance factor and the rater-perspective effects than interpersonal and leadership performance behaviors. Leadership ratings were the least highly correlated with the general performance factor.

The two right-most columns of Table 2 display the communalities and the I-ECV values. Communalities for the factor indicators ranged from .66 ≤ h^2 ≤ .97, meaning that, collectively, the latent factors accounted for between 66% to 97% of the observed indicator variance. I-ECV values ranged from .07 ≤ I-ECV ≤ .40. Thus, 7% to 40% of the reliable performance ratings variance was attributed to general performance with the remaining 60% or more due to perspective effects. Moreover, I-ECV values suggested that self-ratings (.07 ≤ I-ECV ≤ .11) were prominently lower in general (performance) factor saturation than boss (.24 ≤ I-ECV ≤ .27), peer (.36 ≤ I-ECV ≤ .40), and subordinate (.36 ≤ I-ECV ≤ .40) ratings. Although factor scores were not (and could not be) computed in this dataset, Guttman's (1955) factor determinacy index (ρ) was computed for each factor. The general factor was less determinant (ρ = .78) than the boss (.92), peer (.86), subordinate (.86) and self-ratings (.95).

For the model reported in Table 2, The coefficient ω model-based reliability index was high (ω = .96), suggesting that the general and group factors collectively accounted for about 96% of the (unit-weighted) sum score variance. Moreover, the general performance factor alone accounted for 56% ($\omega_h = .56$) of the sum score variance. Taken together, the general performance factor represents the majority (58%) of all common factor variance (i.e., the ratio of ω to ω_h) and the rater-perspective effects (i.e., the group factors) accounted for the remaining (42%) common factor variance.

When partitioning variance at the subscale level (i.e., ω_{hs}), group factors associated with the boss, peer, subordinate, and self-report perspectives each accounted for 73%, 60%, 59%, and 84% of the variance, respectively. Simply put, performance ratings from any one perspective are predominately unrelated to general (overall) managerial performance. Specifically, ratings from bosses, peers, subordinates, and the self only share 27%, 40%, 41%, and 16% (respectively) of their variance with the general performance factor. Comparing ω_h to ω_{hs} highlights the utility of multisource feedback ratings. Namely, ratings from any one perspective are unreliable and therefore insufficient to assess overall managerial performance. Nevertheless, reliability of performance ratings quickly increases as more perspectives are combined together.

Implications. Our re-analyses of the Hoffman et al. (2010) performance evaluations provided novel insights into single-source versus multiple-source ratings of managerial performance. Specifically, our ω_{hs} analyses demonstrated the relative contributions of rater perspectives on the overall observed variance. These results suggest that a substantial 84% of the observed variance in self-reported performance ratings is unrelated to the general performance factor. In contrast, across our modeled rating perspectives, results suggest that subordinate raters (followed closely by peer raters) have the lowest perspective-specific effects (59%). Subordinate ratings of performance are less contaminated with source-perspective effects and have among the highest correlations (i.e., factor loadings) with the general performance factor. Moreover, subordinate and peer managerial performance ratings are more strongly influenced by the general performance factor (i.e., higher I-ECV values) than either boss or self-reported perspectives. These results imply that subordinate and peer raters are the best single-source raters of managerial performance for developmental purposes.

In summary, although each rating source of managerial performance is predominately influenced by perspective-specific effects, the results of our (constrained) bifactor analysis suggests that a general performance factor accounted for the lion's share (56%) of variance in the collective multisource feedback ratings. This latter finding is novel to the present article. Hoffman et al. (2010) reported a variance accounted for index for each rating perspective and averaged across these values to summarize their results. They found that, on average, a general performance factor accounted for 3% of the variance in their models. However, averaging across raters fails to consider prominent psychometric concepts. Namely, that when combining parallel assessments of the same constructs, true score variance accumulates faster than error score variance. In this vein, a grand mean will appreciably underrepresent the overall general performance factor saturation across parallel assessments compared to the present findings that are based on the full bifactor model. In practice, the differences between the present findings and those of the published findings translate into different recommendations about multisource developmental ratings. The small grand mean value reported by Hoffman et al. (2010) suggests that multisource ratings are an expensive and inefficient undertaking. However, our resulting bifactor analyses suggest that multiple rating sources provide developmentally informative and more accurate insights into employee performance.

Example 2: Bifactor Modeling of Conscientiousness Inventories

Conscientiousness is a potent predictor of workplace behaviors and outcomes (Roberts et al., 2005; Wilmot & Ones, 2019). Moreover, it is perhaps the best personality determinant of training and educational performance (Connelly & Ones, 2010; Poropat, 2009). Furthermore, conscientiousness has been implicated as a determinant of satisfaction and well-being at work

(Seltzer, Ones, & Tatar, 2017) and health more generally (Bogg & Roberts, 2004). Thus, the impact of conscientiousness on vocational preparation and performance is notable.

At its core, conscientiousness refers to a person’s tendency to “follow rules and prioritize non-immediate goals” (DeYoung, 2015, p. 45). Individuals high in conscientiousness are often described as hardworking, orderly, responsible, self-controlled, and rule-abiding (Stanek & Ones, 2018). Of relevance for the present manuscript, conscientiousness is also a multidimensional construct (Hogan & Ones, 1997; Stanek & Ones, 2018). A number of empirical studies have sought to identify its lower level structure (e.g., DeYoung, Quilty, & Peterson, 2007; Roberts et al., 2005) though, currently, there is no consensus on the number and nature of the lower order traits (Roberts, Lejuez, Krueger, Richards, & Hill, 2014). Along this vein, we estimated an exploratory bifactor model of 11 conscientiousness facet scales to elucidate the dimensional structure of this domain. We use these data to illustrate aforementioned problems that can arise when estimating factor scores for constrained (hierarchical) bifactor models.

Sample and Data Description. Conscientiousness facet scales were administered to 761 undergraduate students at a large, Midwestern university. Participants were recruited online through the University’s research participant pool. Participants completed the entire study online. The sample was fairly typical for a Midwestern collegiate sample and was primarily composed of White (75.0%) females (68.2%) with an average age of (21.0, *SD* = 2.9). The remaining participants identified as Asian/Pacific Islander (13.0%), multi-racial (4.5%), Black (3.4%), or Hispanic/Latino (2.6%).

In order to represent conscientiousness facets (i.e., subdimensions) that have appeared in various conceptualizations of this domain (e.g., Roberts et al., 2005), multiple scales assessing all known facets of conscientiousness were administered to the sample. Eleven conscientiousness facet scales—achievement striving, cautiousness, dutifulness, industriousness, orderliness, persistence, responsibility, traditionalism, and virtue—were selected from the International Personality Item Pool according to the work of Hough and Ones (2002), Roberts et al. (2005), and Stanek and Ones (2018) to form a content valid representation of the conscientiousness facets. Participants rated how accurately each item described them on a five-point scale (1 = “Very Inaccurate” to 5 = “Very Accurate”). Attention checks were used, and careless responders were excluded from analyses.

Bifactor Modeling. A series of hierarchical bifactor models were applied to evaluate the structure of conscientiousness. Prior to performing these analyses, we ran several preliminary analyses (i.e., scree and EKC plots) to determine the latent dimensionality underlying the data. We also considered prior theoretical work in this domain to decide on the optimal number of group factors to include in the bifactor model. Prior work (over several decades) has supported views (e.g., Digman, 1997) about the hierarchical nature of conscientiousness (Stanek & Ones, 2018). Most recently, DeYoung and colleagues (DeYoung, 2015; DeYoung, Quilty, & Peterson, 2007) have presented empirical and theoretical support for two subdimensions of conscientiousness: orderliness and industriousness. These lower-order factors encompass various facets of conscientiousness that are influenced by a general conscientiousness factor. Once a constrained bifactor model was estimated, factor scores were estimated for all 761 students. Although in hierarchical bifactor models the estimated factor scores are not linearly independent—meaning that the estimated scores on one factor can be perfectly reproduced from the estimated scores on the remaining factors—for didactic purposes, we estimated factor scores for this example. Specifically, we estimated: (a) unit-weighted factor scores, more commonly known as sum scores, and (b) Thurstone's (1947) regression-based factor scores (Grice, 2001; McDonald & Burr, 1967; Tucker, 1971).

Results. Table 3 contains the estimated bifactor measurement model of the 11 conscientiousness subscales. Scree and EKC plots jointly recommended the extraction of three factors. However, prior theory strongly suggested that two factors are best able to explain variation in the lower-order conscientiousness factors (DeYoung, 2015). Thus, two constrained bifactor models were estimated, a DSL bifactor model with two group factors and an SL bifactor model with three group factors.[10] Both models, subjectively speaking, were equally interpretable. However, in the three-group-factor solution, the first group factor (and the items loading onto it) was cleaved in two. This produced two weakly-determined group factors, each marked by only two observed variables. Thus, the theoretically supported, two-group-factor DSL solution was retained (see Table 3). Interested readers can consult the online supplement to see the conscientiousness bifactor model with three group factors.

Insert Table 3 here

In the conscientiousness bifactor measurement model, factor loadings (λ) on the general conscientiousness factor ranged from small to moderately large ($.29 \leq \lambda \leq .58$). Using a common, through arbitrary, cutoff to identify which items saliently load onto each factor (i.e., $\lambda \geq |.30|$),

[10] An SL procedure is inappropriate in cases where fewer than three group factors are present. In a (pre-transformed) higher-order model, the higher-order factor must influence at least three first-order factors to uniquely determine the factor loadings. If two lower-order factors are present, factor loadings on the higher-order factor will be biased which, in turn, will bias the SL bifactor loadings parameters. A DSL procedure directly estimates a constrained bifactor model without first conducting a higher-order factor model and therefore does not suffer from these biases.

the conscientiousness subscales could be categorized under the two group factors. The first group factor was related to the following subscales (with salient group factor loadings in parentheses): diligence ($\lambda = .70$), achievement ($\lambda = .67$), persistence ($\lambda = .64$), industriousness ($\lambda = .44$), virtue ($\lambda = .43$), deliberateness ($\lambda = .41$), and cautiousness ($\lambda = .31$). We interpreted this group factor as prudent work orientation. The second group factor was related to the following subscales: dutifulness ($\lambda = .83$), traditionalism ($\lambda = .40$), and responsibility ($\lambda = .36$). We interpreted this group factor as conformity. Interestingly, in this sample, the orderliness scale had relatively weak loadings on all factors, though its largest loading was on the general conscientiousness factor ($\lambda = .37$), with weaker loadings on both the first ($\lambda = .29$) and second ($\lambda = .22$) group factors.[11]

The subscales varied greatly in how variance was partitioned across the factors. Communalities ranged considerably ($.25 \leq h^2 \leq .99$) with the dutifulness ($h^2 = .99$) subscale being almost entirely comprised of latent factor variance (i.e., general conscientiousness variance and conformity group factor variance). Alternatively, orderliness ($h^2 = .27$) and traditionalism ($h^2 = .25$) shared less than 30% of their observed variance with the three latent conscientiousness factors. This suggests that for both orderliness and traditionalism, there may be other latent personality factors (e.g., neuroticism for orderliness, and openness for traditionalism) accounting for reliable variance beyond conscientiousness. For example, Connelly and colleagues (2014) found that traditionalism is related to *both* low openness and high conscientiousness.

[11] In the conscientiousness bifactor model with three group factors, the prudent work orientation factor was bifurcated into two factors: prudence and work orientation. The latter work orientation factor appears to be fully in line with that industriousness aspect proposed by DeYoung (2015) and colleagues (DeYoung et al., 2007).

Turning to the I-ECV index, the 11 conscientiousness subscales had a somewhat narrow range in their general factor saturation ($.31 \leq$ I-ECV $\leq .52$). Specifically, of the common factor variance, the general conscientiousness factor accounted for 41% (diligence), 40% (achievement), 42% (persistence), 42% (industriousness), 51% (virtue), 49% (deliberateness), 52% (cautiousness), 52% (orderliness), 31% (dutifulness), 34% (traditionalism), and 49% (responsibility) of the various conscientiousness facet scales. Note that these I-ECV values must be considered in conjunction with the communality values. For example, the general conscientiousness factor only accounted for roughly 14% of the observed variance in the orderliness scale (i.e., 52% of 27%).

The estimated conscientiousness bifactor model with two group factors accounted for 90% of the observed total variance (ω = .90). The general conscientiousness factor accounted for nearly half of all observed variance (ω_h = .46) but over half (51%) of the latent factor variance (i.e., the ratio of ω_h to ω). At the subscale level (i.e., ω_{hs}), the first (prudent work orientation) and second (conformity) group factors each accounted for roughly 37% and 54% of the observed subscale variance, respectively.

To illustrate problems associated with estimated factor scores in constrained bifactor models, we estimated factor scores for the conscientiousness data using the unit-weighted and Thurstone's (1947) regression-based scoring methods for the 761 subjects. Table 4 contains the correlations between (a) unit-weighted factor score estimates, (b) regression-based factor score estimates, and (c) estimated scores on a given factor across the two estimation methods. Both methods produced highly intercorrelated factor score estimates but the correlations between estimated factor scores were notably higher (in absolute value) for the unit-weighted estimates than the regression-based estimates. Namely, for the unit-weighted estimates, the general factor scores correlated $r = .97$ and .67 with the first and second group factors, respectively, and the estimated factor scores for the two group factors intercorrelated $r = .51$. Recall that the general and group factors in this model are orthogonal, so observed correlations of .97 and .67 between the general and group factors are highly biased. The regression-based factor score estimates of the general factor correlated $r = .63$ and .52 with prudent work orientation and conformity group factors (respectively) whereas these group factor score estimates were negatively correlated ($r = -.34$). Across the estimation methods, factor scores were highly—but not perfectly—correlated. Thus, particularly for the group factors, estimated factor scores from one method can appreciably differ from estimates from another method. We remind the reader, however, that no factor scoring method is fully appropriate for the hierarchical bifactor model due to the aforementioned constraints on its factor loadings.

Insert Table 4 here

Implications. Applying a constrained bifactor model to 11 conscientiousness subscales provided insights into the dimensional structure of conscientiousness. From the obtained bifactor loadings matrix, it is apparent that several scales described as conscientiousness facets are only moderately correlated with the general conscientiousness factor. Particularly, as indicated by its communality, only 27% of the variance in the orderliness scale is related to conscientiousness and its subdimensions of prudent work orientation and conformity. Moreover, some subscales (e.g., dutifulness) predominately measure group factor variance. In practical terms, this implies that administering a diligence subscale will yield scores mostly reflecting a general conscientiousness factor whereas a dutifulness subscale will yield scores mostly reflecting the conformity subdimension of conscientiousness. Importantly, not all conscientiousness subscales are exchangeable.

Based on theoretical perspectives from the extant literature, we can begin to describe the content domain from the resulting conscientiousness dimensional structure. Namely, the conscientiousness general factor appears to reflect the tendency for people to prioritize long-term goals over immediate gratification (see also Connelly, Ones, Hülsheger, 2018; DeYoung, 2015). The group factor that we labeled prudent work orientation is further distinguished from general conscientiousness by the diligent effort directed to achieving goals, and it roughly corresponds to the industriousness aspect of conscientiousness, with an added element cautiousness (Connelly et al., 2018; DeYoung, 2015). Conformity emerged as the second group factor in our bifactor analyses of the 11 conscientiousness facet scales. This factor uniquely focuses on maintenance of social order, a socially-directed orderliness factor that helps protect long-term goals. Taken together, these group factors appear to reflect the two defining characteristics of conscientiousness. Namely, a person's tendency to "follow rules [conformity] and prioritize non-immediate goals [prudent work orientation]" (DeYoung, 2015, p. 45).

The estimated factor scores from the constrained conscientiousness bifactor model illustrate an important shortcoming of this type of model. Specifically, the use of sum scores (or other factor scoring estimators) as estimated factor scores can produce highly misleading results. Notice in Table 4 that, when estimating factor scores via unit-weighted sum scores, the general conscientiousness estimated factor scores are (slightly) more highly correlated with the prudent work orientation estimated factor scores ($r = .97$) than to the regression-based estimates of the general conscientiousness factor ($r = .96$). Thus, if a researcher estimates subjects' standings on, say, prudent work orientation via sum scores, they would be incorrect to claim that these scores are orthogonal to the general conscientiousness factor—despite the researcher estimating an orthogonal bifactor model!

Conclusion

In this manuscript, we provided vocational behavior researchers a brief overview of both hierarchical and non-hierarchical exploratory bifactor measurement models. We highlighted potential uses of exploratory bifactor models in vocational psychology, we described the best practices (and the statistical code to implement these practices) for estimating and interpreting bifactor models, and we illustrated these concepts in real-world examples of innovative and useful applications of hierarchical bifactor models. Along the way, we also noted important caveats and areas for future research. In short, we believe that exploratory bifactor models, when appropriately applied, hold great promise for aiding vocational behavior researchers in more clearly disentangling multidimensional sources of variance to better understand their research questions.

References

Abad, F. J., Garcia-Garzon, E., Garrido, L. E., & Barrada, J. R. (2017). Iteration of partially specified target matrices: Application to the bi-factor case. *Multivariate Behavioral Research, 52*(4), 416–429. https://doi.org/10.1080/00273171.2017.1301244

Arthur, M. B., Khapova, S. N., & Wilderom, C. P. (2005). Career success in a boundaryless career world. *Journal of Organizational Behavior, 26*(2), 177-202. https://doi.org/10.1002/job.290

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods, 24*(4), 468–491. https://doi.org/10.1037/met0000200

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence, 3*(4), 121-136. https://doi.org/10.3390/jintelligence3040121

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: a meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin, 130*(6), 887. https://doi.org/10.1037/0033-2909.130.6.887

Braeken, J., & van Assen, M. A. L. M. (2017). An empirical Kaiser criterion. *Psychological Methods, 22*(3), 450-466. http://dx.doi.org/10.1037/met0000074

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*(1), 111–150. https://doi.org/10.1207/s15327906mbr3601_05

Campbell, D. P., & Holland, J. L. (1972). A merger in vocational interest research: Applying Holland's theory to Strong's data. *Journal of Vocational behavior, 2*(4), 353-376. https://doi.org/10.1016/0001-8791(72)90012-7

Campbell, J. P., & Wiernik, B. M. (2015). The modeling and assessment of work performance. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 47-74. https://doi.org/10.1146/annurev-orgpsych-032414-111427

Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. Cambridge University Press.

Cattell, R. B. (1966). The Scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276. https://doi.org/10.1207/s15327906mbr0102_10

Chen, F. F., West, S., & Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin, 103*(2), 276--279. https://doi.org/10.1037/0033-2909.103.2.276

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological bulletin, 136*(6), 1092. https://doi.org/10.1037/a0021212

Connelly, B. S., Ones, D. S., Davies, S. E., & Birkland, A. (2014). Opening up openness: A theoretical sort following critical incidents methodology and a meta-analytic investigation of the trait family measures. *Journal of Personality Assessment, 96*(1), 17-28. https://doi.org/10.1080/00223891.2013.809355

Connelly, B., Ones, D. & Hülsheger, U. (2018). Personality in industrial, work and organizational psychology: theory, measurement and application. In D. S. Ones, N. Anderson & C. Viswesvaran (Eds.) The sage handbook of industrial, work and organizational psychology (pp. 320-365). 55 City Road, London: SAGE Publications Ltd https://doi.org/10.4135/9781473914940.n13

Craig, S. B., & Hannum, K. (2006). Research update: 360-degree performance assessment. *Consulting Psychology Journal: Practice and Research, 58*(2), 117. https://doi.org/10.1037/1065-9293.58.2.117

Crawford, C. B. & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika, 35*(321). https://doi.org/10.1007/bf02310792

Cucina, J., & Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence, 5*(3), 27. https://doi.org/10.3390/jintelligence5030027

Demerouti, E., Bakker, A. B., Vardakou, I., & Kantas, A. (2003). The convergent validity of two burnout instruments: A multitrait-multimethod analysis. *European Journal of Psychological Assessment, 19*(1), 12. https://doi.org/10.1027//1015-5759.19.1.12

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5), 880. https://doi.org/10.1037/0022-3514.93.5.880

DeYoung, C. G. (2015). Cybernetic Big Five theory. *Journal of Research in Personality, 56*, 33-58. https://doi.org/10.1016/j.jrp.2014.07.004

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*(6), 1246. https://doi.org/10.1037//0022-3514.73.6.1246

Di Fabio, A., & Kenny, M. E. (2019). Decent work in Italy: Context, conceptualization, and assessment. *Journal of Vocational Behavior, 110*, 131-143. https://doi.org/10.1016/j.jvb.2018.10.014

Duffy, R. D., Blustein, D. L., Diemer, M. A., & Autin, K. L. (2016). The psychology of working theory. *Journal of Counseling Psychology, 63*(2), 127. https://doi.org/10.1037/cou0000140

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research, 27*(3), 387-415. https://doi.org/10.1207/s15327906mbr2703_5

Ferrando, P. J. & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement, 78*(5), 762-780. https://doi.org/10.1177/0013164417719308

Ferreira, J. A., Haase, R. F., Santos, E. R., Rabaça, J. A., Figueiredo, L., Hemami, H. G., & Almeida, L. M. (2019). Decent work in Portugal: Context, conceptualization, and assessment. *Journal of Vocational Behavior, 112*, 77-91. https://doi.org/10.1016/j.jvb.2019.01.009

Fugate, M., Kinicki, A. J., & Ashforth, B. E. (2004). Employability: A psycho-social construct, its dimensions, and applications. *Journal of Vocational behavior, 65*(1), 14-38. https://doi.org/10.1016/j.jvb.2003.10.005

Gaudron, J. P., & Vautier, S. (2007). Analyzing individual differences in vocational, leisure, and family interests: A multitrait-multimethod approach. *Journal of Vocational Behavior, 70*(3), 561-573. https://doi.org/10.1016/j.jvb.2007.01.004

Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: G as superordinate or breadth factor? *Psychological Science Quarterly, 50*, 21–43.

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55*, 57–68. http://dx.doi.org/10.1016/j.intell.2016.01.006

Giordano, C., & Waller, N. G. (2020). Recovery of bifactor structures: A comparison of seven methods. *Psychological Methods, 25*(2), 143–156. https://doi.org/10.1037/met0000227

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., . . . Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of Abnormal Psychology, 128*(7), 740-764. http://dx.doi.org/10.1037/abn0000434

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*(4), 430. https://doi.org/10.1037/1082-989X.6.4.430

Grice, J. W., & Harris, R. J. (1998). A comparison of regression and loading weights for the computation of factor scores. *Multivariate Behavioral Research, 33*(2), 221–247. https://doi.org/10.1207/s15327906mbr3302_2

Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology, 8*(2), 65–81. https://doi.org/10.1111/j.2044-8317.1955.tb00321.x

Hattori, M., Zhang, G., & Preacher, K. J. (2017). Multiple local solutions and geomin rotation. *Multivariate Behavioral Research, 52*(6), 720–731. https://doi.org/10.1080/00273171.2017.1361312

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191–205. https://doi.org/10.1177/1094428104263675

Hogan, J., & Ones, D. S. (1997). Conscientiousness and integrity at work In R. Hogan, J. Johnson, & S. Briggs (Eds.) *Handbook of Personality Psychology* (pp. 849-870). Academic Press. https://doi.org/10.1016/b978-012134645-4/50033-0

Hoffman, B., Lance, C. E., Bynum, B., & Gentry, W. A. (2010). Rater source effects are alive and well after all. *Personnel Psychology, 63*(1), 119–151. https://doi.org/10.1111/j.1744-6570.2009.01164.x

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185. https://doi.org/10.1007/bf02289447

Holzinger, K. J. (1935). Introduction to bi-factor theory; solid and hollow staircase patterns for sets of data from mooseheart (No. 5). University of Chicago.

Holzinger, K. J. (1936). Recent research on unitary mental traits. *Journal of Personality, 4*(4), 335-343. https://doi.org/10.1111/j.1467-6494.1936.tb02038.x

Holzinger, K. J. (1937). Student manual of factor analysis. University of Chicago Press.

Holzinger, K. J., & Swineford, F. (1937). The Bi-factor method. *Psychometrika*, *2*(1), 41–54. 10.1007/BF02287965

Holzinger, K. J. (1945). Spearman as I knew him. *Psychometrika, 10*(4), 231-235. https://doi.org/ 10.1007/BF02288890

Hough, L. M., & Ones, D. S. (2002). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of Industrial, Work and Organizational Psychology*, Vol. 1. Personnel psychology (pp. 233-277). Thousand Oaks, CA: Sage Publications Ltd. https://doi.org/10.4135/9781848608320.n13

Hurley, J. R., & Cattell, R. B. (1962). The Procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science, 7*(2), 258-262. https://doi.org/10.1002/bs.3830070216

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika, 76*(4), 537–549. https://doi.org/10.1007/s11336-011-9218-4

Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika, 77*(3), 442–454. https://doi.org/10.1007/s11336-012-9269-1

Jennrich, R. I., & Bentler, P. M. (2013). Erratum to: Exploratory bi-factor analysis. *Psychometrika, 78*(3), 556–556. https://doi.org/10.1007/s11336-013-9346-0

Judge, T. A., & Piccolo, R. F. (2004). Transformational and transactional leadership: a meta-analytic test of their relative validity. *Journal of Applied Psychology, 89*(5), 755. https://doi.org/10.1037/0021-9010.89.5.755

Karasek Jr, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly, 24*(2), 285-308. https://doi.org/10.2307/2392498

Ledermann, W. (1937). On the rank of the reduced correlational matrix in multiple-factor analysis. *Psychometrika, 2*(2), 85-93. https://doi.org/10.1007/bf02288062

Levin, J. (1973) Bifactor analysis of a multitrait-multimethod matrix of leadership criteria in small groups. *The Journal of Social Psychology, 89*(2), 295-299. https://doi.org/10.1080/00224545.1973.9922602

LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: a critical review and meta-analysis. *Journal of Applied Psychology, 87*(1), 52. https://doi.org/10.1037//0021-9010.87.1.52

Locke, E. A. (1969). What is job satisfaction? *Organizational Behavior and Human Performance, 4*(4), 309-336. https://doi.org/10.1016/0030-5073(69)90013-0

Lombardo, M. M., McCauley, C. D., McDonald-Mann, D., & Leslie, J. B. (1999). BENCHMARKS® developmental reference points. Greensboro, NC: Center for Creative Leadership.

Lorenzo-Seva, U., & ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57-64. https://doi.org/10.1027/1614-2241.2.2.57

McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The Trait-Reputation-Identity Model. *Psychological Review, 123*, 569-591. https://doi.org/10.1037/rev0000035

McCrae, R. R., & Costa Jr, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences, 36*(3), 587-596. https://doi.org/10.1016/s0191-8869(03)00118-1

McDonald, R. P. (1999). Test theory: A unified treatment. Taylor & Francis.

McDonald, R. P., & Burr, E. (1967). A comparison of four methods of constructing factor scores. *Psychometrika, 32*(4), 381–401. https://doi.org/10.1007/bf02289653

McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, Present, and Future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, 136–181. The Guilford Press.

Morris, J. A., & Feldman, D. C. (1996). The dimensions, antecedents, and consequences of emotional labor. *Academy of Management Review, 21*(4), 986-1010. https://doi.org/10.5465/amr.1996.9704071861

Mulaik, S. A. (2010). Foundations of Factor Analysis. Boca Raton, FL: CRC Press.

Mulaik, S. A., & Quartetti, D. A. (1997). First order or higher order general factor? *Structural Equation Modeling: A Multidisciplinary Journal, 4*(3), 193–211. https://doi.org/10.1080/10705519709540071

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41*(5), 407–422. https://doi.org/10.1016/j.intell.2013.06.004

Ones, D., Viswesvaran, C., & Schmidt, F. (2008). No New Terrain: Reliability and Construct Validity of Job Performance Ratings. *Industrial and Organizational Psychology, 1*(2), 174-179. doi:10.1111/j.1754-9434.2008.00033.x

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322. https://doi.org/10.1037/a0014996

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research, 48*(1), 28-56. https://doi.org/10.1080/00273171.2012.710386

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology, 107*(4), 677. https://doi.org/10.1037/a0037250

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*, 544–559. http://dx.doi.org/10.1080/00223891.2010.496477

Rhoades, L., & Eisenberger, R. (2002). Perceived organizational support: a review of the literature. *Journal of Applied Psychology, 87*(4), 698. https://doi.org/10.1037//0021-9010.87.4.698

Ribeiro, M. A., Teixeira, M. A. P., & Ambiel, R. A. M. (2019). Decent work in Brazil: Context, conceptualization, and assessment. *Journal of Vocational Behavior, 112*, 229-240. https://doi.org/10.1016/j.jvb.2019.03.006

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23*(1), 51–67. https://doi.org/10.1207/s15327906mbr2301_3

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 58*(1), 103–139. https://doi.org/10.1111/j.1744-6570.2005.00301.x

Roberts, B. W., Lejuez, C., Krueger. R. F., Richards, J. M., Hill, P. L. (2014) What is Conscientiousness and how can it be assessed? *Developmental Psychology 50*(5), 1315-1330. https://doi.org/10.1037/a0031109

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137. https://doi.org/10.1037/met0000045

Rozeboom, W. W. (1992). The glory of suboptimal factor rotation: Why local minima in analytic optimization of simple structure are more blessing than curse. *Multivariate Behavioral Research, 27*(4), 585–599. https://doi.org/10.1207/s15327906mbr2704_5

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment, 24*(2), 282-292. https://doi.org/10.1037/a0025697

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53–61. https://doi.org/10.1007/bf02289209

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement, 31*(3), 699-714. https://doi.org/10.1177/001316447103100310

Seltzer, B. K., Ones, D. S., & Tatar, A. (2017). Using personality facets to understand the nature of personality-satisfaction relationships: Findings from meta-analytic bifactor latent modeling. *Career Development International, 22*(5), 477–506. https://doi.org/10.1108/cdi-08-2017-0141

Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology, 58*(1), 33-66. https://doi.org/10.1111/j.1744-6570.2005.514_1.x

Spearman, C. (1904). "General intelligence," objectively determined and measured. The *American Journal of Psychology, 15*(2), 201–292. https://doi.org/10.2307/1412107

Stanek, K. C., & Ones, D. S. (2018). Taxonomies and compendia of cognitive ability and personality measures relevant to Industrial, Work, and Organizational Psychology. In *The*

SAGE Handbook of Industrial, Work and Organizational Psychology (Vol. 1). Sage London, United Kingdom. https://doi.org/10.4135/9781473914940.n14

Stanek, K. C., Ones, D. S., & McGue, M. (2017). Counterproductive behaviors: Relations across life domains, etiology, and implications for applied practice. *International Journal of Selection and Assessment 25*, 111-124. https://doi.org/10.1111/ijsa.12164

Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's some interesting parallels. *Psychometrika, 44*(2), 157–167. https://doi.org/10.1007/bf02293967

Steiger, J. H., & Schönemann, P. H. (1978). A history of factor indeterminacy. In S. Shye (Ed.), *Theory Construction and Data Analysis in the Behavioral Sciences*, 136-178. San Francisco, CA: Jossey-Bass.

ten Berge, J. M., Krijnen, W. P., Wansbeek, T., & Shapiro, A. (1999). Some new results on correlation-preserving factor scores prediction methods. *Linear Algebra and its Applications, 289*(1-3), 311-318. https://doi.org/10.1016/s0024-3795(97)10007-6

ten Berge, J. M. F. & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika, 69*(4), 613-625. https://doi.org/10.1007/BF02289858

Thomson, G. H. (1951). The factorial analysis of human ability. Houghton Mifflin.

Thurstone, L. L. (1934). The Vectors of Mind. *Psychological Review, 41*(1), 1–32. https://doi.org/10.1037/h0075959

Thurstone, L. L. (1947). Multiple-factor analysis; a development and expansion of the vectors of mind. University of Chicago Press.

Toker, Y., & Ackerman, P. L. (2012). Utilizing occupational complexity levels in vocational interest assessments: Assessing interests for STEM areas. *Journal of Vocational Behavior, 80*(2), 524–544. https://doi.org/10.1016/j.jvb.2011.09.001

Tucker, L. R. (1971). Relations of factor score estimates to their use. *Psychometrika, 36*(4), 427–436. https://doi.org/10.1007/bf02291367

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8*(4), 216-226. https://doi.org/10.1111/1468-2389.00151

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*(5), 557–574. https://doi.org/10.1037/0021-9010.81.5.557

Waller, N. G. (2018). Direct Schmid–Leiman transformations and rank-deficient loadings matrices. *Psychometrika, 83*(4), 858–870. https://doi.org/10.1007/s11336-017-9599-0

Waller, N. G. (2019). *Fungible: Psychometric functions from the Waller lab.* R package version 1.93.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology, 54*(6), 1063. https://doi.org/10.1037//0022-3514.54.6.1063

Wherry, R. J. (1959). Hierarchical factor solutions without rotation. *Psychometrika, 24*(1), 45–51. https://doi.org/10.1007/bf02289762

Wilmot, M. P., & Ones, D. S. (2019). A century of research on conscientiousness at work. *Proceedings of the National Academy of Sciences, 116*(46), 23004-23010. https://doi.org/10.1073/pnas.1908430116

Wilson, E. B. (1928). On hierarchical correlation systems. *Proceedings of the National Academy of Sciences of the United States of America, 14*(3), 283. https://dx.doi.org/10.1073%2Fpnas.14.3.283

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1*(4), 354–365. https://doi.org/10.1037//1082-989x.1.4.354

Yang, R., Spirtes, P., Scheines, R., Reise, S. P., & Mansolf, M. (2017). Finding pure submodels for improved differentiation of bifactor and second-order models. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(3), 402-413. https://doi.org/10.1080/10705511.2016.1261351

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis.* Albany, NY: SUNY Press.

Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*(2), 113–128. https://doi.org/10.1007/bf02294531

Zacher, H. (2014). Career adaptability predicts subjective career success above and beyond personality traits and core self-evaluations. *Journal of Vocational Behavior, 84*(1), 21-30. https://doi.org/10.1016/j.jvb.2013.10.002

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_h: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7

Tables

Table 1

Exploratory Bifactor Analysis Methods

	Analytic Strategy		
Model type	Hybrid methods	Target rotation	Analytic bifactor rotation
Hierarchical	• **Schmid-Leiman** (1957) • Wherry (1959)	• **Direct Schmid-Leiman** (Waller, 2018)	No Methods Currently Available
Non-hierarchical	• Holzinger and Swineford (1937)	• **Direct Bifactor** (Waller, 2018) • **Schmid-Leiman target rotation** (Reise, Moore, & Haviland, 2010) • **Iterated Schmid-Leiman target rotation** (Abad et al., 2017)	• Bifactor Quartimin (Jennrich & Bentler, 2011, 2013) • Bifactor Geomin (Jennrich & Bentler, 2012)

Note: Bolded methods have been found to accurately recover the loadings matrix of bifactor measurement models (Giordano & Waller, 2020).

Table 2

Schmid-Leiman Bifactor Solution of the Multisource Performance Ratings

		Group factors				Item indices	
	Performance	Boss	Peer	Subordinate	Self	h^2	I-ECV
Boss Ratings							
Technical	.48	.86	-.01	-.01	.00	.97	.24
Interpersonal	.48	.78	.01	.02	-.01	.84	.27
Leadership	.45	.76	.00	.00	.02	.77	.26
Peer Ratings							
Technical	.58	.00	.78	-.02	.01	.95	.36
Interpersonal	.56	.00	.72	.01	-.02	.83	.38
Leadership	.53	.01	.65	.03	.02	.70	.40
Subordinate Ratings							
Technical	.59	-.01	-.02	.77	.01	.94	.36
Interpersonal	.56	.01	.02	.70	-.02	.80	.39
Leadership	.51	.01	.01	.63	.01	.67	.40
Self-Ratings							
Technical	.28	.01	.00	.00	.90	.89	.09
Interpersonal	.28	.00	.02	.03	.79	.71	.11
Leadership	.21	.00	-.01	-.02	.78	.66	.07

Note: h^2 = indicator communality; I-ECV = item (indicator) explained common variance.

Table 3

Direct Schmid-Leiman Bifactor Solution of the Conscientiousness Subscales

		Group Factors		Item indices	
	Conscientiousness	PWO	Conformity	h^2	I-ECV
Diligence	.58	.70	.02	.82	.41
Achievement	.55	.67	.01	.74	.40
Persistence	.56	.64	.05	.73	.42
Industriousness	.37	.44	.03	.33	.42
Virtue	.51	.43	.25	.51	.51
Deliberateness	.42	.41	.14	.36	.49
Cautiousness	.41	.31	.24	.32	.52
Orderliness	.37	.29	.22	.27	.52
Dutifulness	.55	.03	.83	.99	.31
Traditionalism	.29	.04	.40	.25	.34
Responsibility	.41	.22	.36	.35	.49

Note: PWO = Prudent work orientation; h^2 = indicator communality; I-ECV = item (indicator) explained common variance.

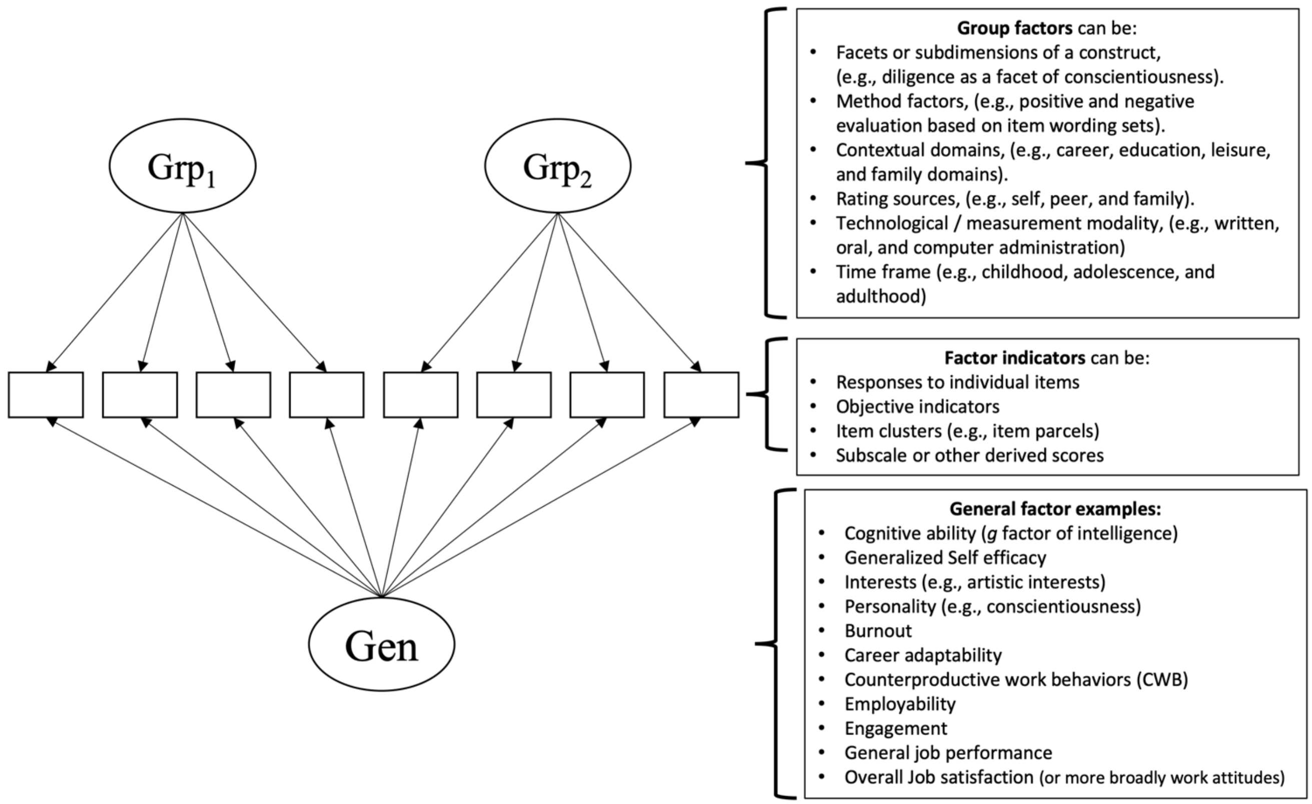
Table 4

Intercorrelations Between Factor Score Estimates in the Conscientiousness Bifactor Model

	Conscientiousness	Prudent work orientation	Conformity
Conscientiousness	(.96)	.63	.52
Prudent work orientation	.97	(.85)	-.34
Conformity	.67	.51	(.74)

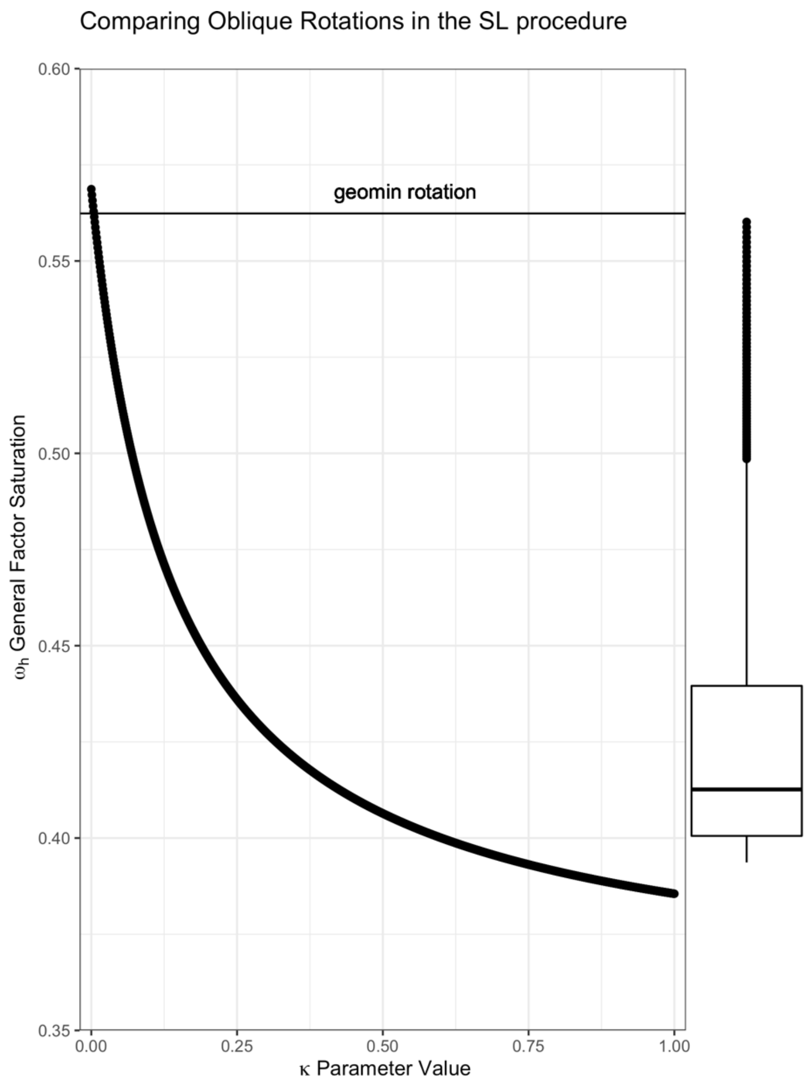
Note: Values in the lower triangle represent the correlations between unit-weighted factor score estimates; values in the upper triangle represent the correlations between Thurstone's regression-based factor score estimates. Values in the matrix diagonal represent the correlation of scores on the same factor by different factor scoring methods.

Figure 1



caption: The diagram on the left-hand side of the figure depicts a (hierarchical or non-hierarchical) bifactor model with a general factor (i.e., the circle labeled 'Gen') and two group factors (i.e., the circles labeled 'Grp_1' and 'Grp_2'). Boxes represent the factor indicators. On the right-hand side of the figure, text boxes contain (non-exhaustive) example applications for modeling the group factors, types factor indicators, and example constructs in which to model a general factor.

Figure 2



Caption: Using data from Hoffman et al. (2010), general factor saturation (ω_h) for each of the 1,001 rotations is plotted against the rotation tuning parameter (κ; Crawford & Ferguson, 1970). The solid horizontal line depicts the general factor saturation obtained from a geomin rotation (ω_h = .56) as a point of reference. General factor saturation ranges from 38.55% to 56.87% of the total sum score variance.