*Original Research Article*

Educational and Psychological Measurement
2024, Vol. 84(1) 5–39

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00131644231155804
journals.sagepub.com/home/epm

# The Impact of Measurement Model Misspecification on Coefficient Omega Estimates of Composite Reliability

**Stephanie M. Bell[1], R. Philip Chalmers[1] and David B. Flora[1]**

## Abstract

Coefficient omega indices are model-based composite reliability estimates that have become increasingly popular. A coefficient omega index estimates how reliably an observed composite score measures a target construct as represented by a factor in a factor-analysis model; as such, the accuracy of omega estimates is likely to depend on correct model specification. The current paper presents a simulation study to investigate the performance of omega-unidimensional (based on the parameters of a one-factor model) and omega-hierarchical (based on a bifactor model) under correct and incorrect model misspecification for high and low reliability composites and different scale lengths. Our results show that coefficient omega estimates are unbiased when calculated from the parameter estimates of a properly specified model. However, omega-unidimensional produced positively biased estimates when the population model was characterized by unmodeled error correlations or multidimensionality, whereas omega-hierarchical was only slightly biased when the population model was either a one-factor model with correlated errors or a higher-order model. These biases were higher when population reliability was lower and increased with scale length. Researchers should carefully evaluate the feasibility of a one-factor model before estimating and reporting omega-unidimensional.

## Keywords

coefficient omega, reliability, measurement model, omega-hierarchical, item factor analysis, factor reliability

[1]York University, Toronto, Ontario, Canada

**Corresponding Author:**
David B. Flora, Department of Psychology, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3.
Email: dflora@yorku.ca

A reliability estimate known as *coefficient omega* represents the proportion of variance in a composite score (calculated by summing or averaging item scores) explained by a latent variable, or factor, that is common to all items comprising the composite (McDonald, 1999; Zinbarg et al., 2006). Omega estimates have become more popular, largely due to a sizable literature advocating for the use of omega for reliability estimation instead of coefficient alpha (e.g., Dunn et al., 2013; Graham, 2006; McNeish, 2018; Watkins, 2017). Although this literature describes the untenability of essential tau-equivalence and independent errors underlying coefficient alpha, another concern is researchers' tendency to equate the classical test theory (CTT) true score with a construct score. As Borsboom (2005) and Borsboom & Mellenbergh (2002) have explained, the expected value definition of the CTT true score does not imply that true scores are determined by a single systematic source of variance (i.e., a single construct or latent variable; also see Bollen, 1989; Ellis, 2021). Consequently, although coefficient alpha may remain viable as an estimate of true score reliability (e.g., Raykov & Marcoulides, 2017; Savalei & Reise, 2019; Sijtsma & Pfadt, 2021), it is often an inadequate estimate of the extent to which an observed score is a reliable estimate of a particular construct score, especially for multidimensional composites (Green & Yang, 2015; Reise et al., 2010). Instead, as a reliability estimate based on the common factor model, a coefficient omega index can directly represent the proportion of observed score variance due to a single factor intended to represent a well-defined construct, over and above nuisance variance sources due to extraneous factors or error covariance. For this reason, McDonald (1999) and Zinbarg et al. (2006) claimed that omega is useful as a "construct validity coefficient" in addition to providing reliability information.

It is extremely common for researchers to attempt to estimate the reliability of a composite's total score under the implied assumption that there is a single construct that all items in the composite have in common (i.e., there is a factor that influences all items; see Flake et al., 2017); a coefficient omega index is meant to do just that. Yet, the term *coefficient omega* does not refer to a single, specific reliability index. Instead, coefficient omega broadly refers to a set of model-based indices which differ according to the measurement model fitted to the items and how the *target construct*—the construct that the composite is intended to measure—is represented as a latent variable within that model. In the current paper, we examine the estimates of different omega parameters that are each intended to represent the reliability of a composite score as a measure of a target construct that influences all items; this target construct may be represented by the only factor in a one-factor model, the general factor in a bifactor model, or the higher-order factor in a higher-order model. The methodological literature cited herein commonly defines omega indices with respect to the parameters of these models; consequently, any investigation of the finite-sample properties of omega estimates must draw on these population model structures. Adapting terms from previous sources (e.g., McDonald, 1999; Zinbarg et al., 2006), Flora (2020) refers to these versions of coefficient omega as $\omega_u$ (i.e., omega-unidimensional), $\omega_H$ (i.e., omega-hierarchical), and $\omega_{ho}$ (omega-higher-order), respectively; we define population parameters for these versions of omega below. Coefficient omega indices not addressed in the current paper include omega-total (representing the proportion of composite variance explained by all factors in the measurement model; Revelle & Zinbarg, 2009) and omega-subscale (also referred to as omega hierarchical subscale, representing the proportion of subscale variance explained by the corresponding specific factor of a bifactor model; Rodriguez et al., 2016).

## Coefficient Omega Definitions

First, define $x_{ij}$ as the observed score on item $j$ for individual test taker $i$. Then, the test total score (or composite score) $X_i$ for individual $i$ is simply the sum of the item scores, $X_i = \sum_{j=1}^{J} x_{ij}$ across all $J$ items comprising the composite.[1] Each version of coefficient omega defined below represents the proportion of variance in $X_i$ that is explained by a single latent variable that influences each of the $J$ items. Furthermore, each version of coefficient omega is a function of the parameters of a factor-analytic measurement model fitted to the $x_{ij}$ observed item scores. Generally, we prefer a confirmatory factor analysis (CFA) approach over exploratory factor analysis (EFA) for estimating coefficient omega, as CFA implies that there is stronger theory or prior research for a given measurement model representing how the individual items are related to the target construct (Flora, 2020; Zinbarg et al., 2006).

### *Omega-Unidimensional*

In its original form, coefficient omega (McDonald, 1999) is a reliability estimate based on the CTT congeneric model (also see Jöreskog, 1971). The congeneric model for the item scores can be expressed as a one-factor model with

$$x_{ij} = \lambda_j f_i + \varepsilon_{ij},$$

where $\lambda_j$ is the factor loading for item $j$, $f_i$ is the unobserved factor score for individual $i$, and $\varepsilon_{ij}$ is the error term. Omega-unidimensional represents the proportion of composite score variance attributable to the single factor. If the one-factor model is identified such that VAR($f$) = 1, then

$$\omega_u = \frac{\left(\sum_{j=1}^{J} \lambda_j\right)^2}{\sigma_X^2},$$

where $\sigma_X^2$ is the variance of the composite score $X$. If the factor loadings are equal across items ($\lambda_1 = \lambda_2 =. . .= \lambda_J$; that is, essential tau-equivalence) and the $\varepsilon_{ij}$ error terms are independent, then $\omega_u$ is equivalent to coefficient alpha (Green & Yang, 2009a).

A CFA approach to the estimation of $\omega_u$ allows the specification of non-zero covariances among the item errors, which in turn impacts the model-implied $\sigma_X^2$ in the denominator of $\omega_u$. In practice, researchers may specify non-zero error covariances to capture method artifacts (e.g., excess covariance among items due to contextual effects or similar wording; see Green & Yang, 2009a). Consequently, the model-implied composite score variance is

$$\sigma_X^2 = \left(\sum_{j=1}^{J}\lambda_j\right)^2 + \sum_{j=1}^{J} VAR(\varepsilon_j) + \sum_{j=1}^{J} COV(\varepsilon_j \varepsilon_{j'}),$$

if VAR($f$) = 1 as above. If the error covariances are fixed to 0, as is often the case, then the model-implied composite score variance reduces to

$$\sigma_X^2 = \left(\sum_{j=1}^{J}\lambda_j\right)^2 + \sum_{j=1}^{J} VAR(\varepsilon_j),$$

which is a common formula for $\sigma_X^2$ presented in articles on coefficient omega. Clearly, however, model misspecification in the form of ignoring true, non-zero error covariances will lead to inaccurate estimates of $\omega_u$ if $\omega_u$ is calculated using the latter formula for model-implied $\sigma_X^2$ (Yang & Green, 2010).

Furthermore, composite unidimensionality is a strong assumption for $\omega_u$. If a one-factor model is not the true data-generating model for the item scores, then fitting a one-factor model to sample data will likely produce $\omega_u$ estimates that are inaccurate representations of composite reliability with respect to the measurement of a factor that influences all items despite the presence of population-level multidimensionality (hence the motivation for $\omega_H$, defined below). In the current study, we investigate the effect of misspecifying a one-factor model for a multidimensional composite, as well as the effect of ignoring true error covariances when estimating $\omega_u$ with a one-factor model.

Kelley and Pornprasertmanit (2016) suggest that $\omega_u$ estimates are robust to model misspecification when $\sigma_X^2$ is instead estimated as the observed variance of $X$, and their simulation found support for this conjecture under minor model misspecification.[2] However, we expect that more substantial misspecification, such as ignoring true error covariances, will produce biased $\omega_u$ estimates even when $\sigma_X^2$ is estimated using the observed variance of $X$ because ignoring error covariance will also have some impact on the factor loading estimates in the numerator of $\omega_u$. Therefore, the current study also compares omega estimates as a function of whether the denominator is calculated using observed composite variance or model-implied variance.

## Omega-Hierarchical

Tests designed to measure a single target construct often have a multidimensional structure. This multidimensionality is often intentional, as when a test is designed to produce subscale scores in addition to a total score. In other situations, the breadth of the construct definition or aspects of item formats (e.g., wording effects) can produce unintended multidimensionality, even if a general target construct that influences all items is still present. In these situations, the one-factor model is unlikely to explain the item-level data adequately, implying that $\omega_u$ is an inappropriate measure of how reliably the total score from a multidimensional test measures the target construct.

A bifactor measurement model is often advocated for representing multidimensionality within composite intended to measure a single, general construct (e.g., Rodriguez et al., 2016; Zinbarg et al., 2006). The bifactor model for item score $x_{ij}$ can be written as

$$x_{ij} = \lambda_{jg}g_i + \lambda_{jk}s_{ki} + \varepsilon_{ij},$$

where $g_i$ is the score for individual $i$ on a general factor $g$ that influences all items in the composite and $s_{ki}$ is the score on the specific factor $k$ that influences item $j$. Whereas the general factor loadings $\lambda_{jg}$ are freely estimated for all $J$ items in the composite, each specific factor (also known as a group factor) influences only a subset of items such that factor loadings $\lambda_{jk}$ for specific factor $k$ are freely estimated for a predetermined subset of items (e.g., items within a proposed subscale) and fixed to 0 for all other items. For model identification, the general and specific factors are orthogonal (Yung et al., 1999) and the variance of each factor is fixed to 1.

A version of coefficient omega known as omega-hierarchical, or $\omega_H$, is a function of the parameters of a bifactor model fitted to the items comprising a composite and represents the proportion of composite score variance that can be attributed to the general factor (Rodriguez et al., 2016; Zinbarg et al., 2006):

$$\omega_H = \frac{\left(\sum_{j=1}^{J} \lambda_{jg}\right)^2}{\sigma_X^2}.$$

Thus, the formula for $\omega_H$ is nearly the same as that for $\omega_u$, except now the numerator is a function of the general factor loadings. Furthermore, the denominator again represents the composite score total variance which can be estimated from either the model-implied total variance or the observed variance of $X$. The bifactor model-implied total variance is a function of the general factor loadings, specific factor loadings, and the item error variances (and potential error covariances).

Given our earlier comment that omega-unidimensional ($\omega_u$) is likely to be an inaccurate measure of reliability with respect to the measurement of a single factor influencing all items in a multidimensional composite or a unidimensional composite characterized by (ignored) error covariance, the current study also investigates the accuracy of omega-hierarchical ($\omega_H$) estimates when the data-generating population model is a bifactor model (correct model specification for $\omega_H$), a one-factor model with population-level error covariances (a misspecified model for $\omega_H$), or a higher-order model (another misspecified model for $\omega_H$); despite differing factor structures, each of these population models is still characterized by having a factor that influences all items. Next, we define omega-higher-order to obtain a correct population omega for the higher-order factor structure against which to compare $\omega_H$ estimates obtained by fitting a misspecified bifactor model to data from a higher-order population model.

## Omega-Higher Order

Although $\omega_H$ is often recommended as a reliability estimate for the measurement of a target construct influencing all items in a multidimensional composite (e.g., Green & Yang, 2015; Reise et al., 2013; Watkins, 2017), the bifactor model underlying $\omega_H$ may not be the correct population model structure for a target construct influencing all items. Alternatively, researchers may hypothesize a higher-order factor structure such that a broad, overarching latent variable (a higher-order factor) causes individual differences in several more conceptually narrow lower-order factors, which in turn directly influence the observed item responses. This model can be expressed with

$$x_{ij} = \lambda_{jk} f_{ki} + \varepsilon_{ij},$$

where $f_{ki}$ is an individual's score on lower-order factor $k$; each lower-order factor influences only a subset of items, such that factor loadings $\lambda_{jk}$ for lower-order factor $k$ are freely estimated for a subset of items (e.g., items within a proposed subscale) and fixed to 0 for all remaining items. Then, instead of allowing the lower-order factors to freely covary with each other, each is directly regressed on a higher-order factor:

$$f_{ki} = \gamma_k h_i + \zeta_k,$$

where $h_i$ is the score on the higher-order factor, $\gamma_k$ is the higher-order factor loading for the $k$th lower-order factor, and $\zeta_k$ is an error term. This higher-order model is identified if there are at least three lower-order factors (Rindskopf & Rose, 1988).

Omega-higher order, or $\omega_{ho}$, is a function of the parameters of a higher-order model fitted to the items comprising a composite and represents the proportion of composite score variance that can be attributed to the higher-order factor. Because the associations between the higher-order factor and the observed item scores are mediated through the lower-order factors, $\omega_{ho}$ is a function of these indirect effects; each indirect effect of the higher-order factor on an item score is the product of the item's lower-order factor loading and the corresponding higher-order factor loading (i.e., $\lambda_{jk} * \gamma_k$; Raykov & Zinbarg, 2011). Consequently,

$$\omega_{ho} = \frac{\left(\sum_{j=1}^{J} \lambda_{jk} \gamma_k\right)^2}{\sigma_X^2},$$

with the numerator of $\omega_{ho}$ equaling the squared total of the indirect effects of the higher-order factor on the observed item scores and the denominator again representing the variance of the composite score.[3]

Although the bifactor model and the higher-order model are formally related (see Yung et al., 1999), the bifactor model's general factor and a higher-order factor have different interpretations: The bifactor general factor has direct effects on the item scores with the effects of the specific factors covaried out, whereas the higher-order factor has indirect effects on the item scores which act entirely through the lower-order factors. Yet, a higher-order factor and a bifactor model's general factor are both latent variables that influence all items in a multidimensional composite and therefore estimates of $\omega_H$ may provide reasonable approximations to $\omega_{ho}$. Another aim of the current study is to assess this possibility.

## The Current Study

The purpose of each version of coefficient omega presented above is to quantify the proportion of composite score variance due to a latent variable that influences all items comprising the composite, with $\omega_u$ being a function of the parameters of a one-factor model, $\omega_H$ based on the parameters of a bifactor model, and $\omega_{ho}$ based on the parameters of a higher-order model. In practice, the true measurement model for a set of items is unknowable, and researchers must rely on a variety of model fit statistics, comparisons of competing models, and previous evidence to support their hypothesized measurement model. Yet, researchers often estimate the reliability of a composite without first testing its dimensionality, which in turn can lead to misleading estimates using either coefficient alpha or $\omega_u$ as researchers may often assume that the composite score is a measure of a target construct that influences all items (see Flake et al., 2017). This practice may be indirectly encouraged by recent resources that facilitate the calculation of $\omega_u$ without the explicit estimation of a factor analysis model (Hancock & An, 2020; Hayes & Coutts, 2020; Kelley, 2022; Pfadt, van den Bergh, Sijtsma, Moshagen, & Wagenmakers, 2022); these implementations return omega estimates which are implicitly based on a one-factor model (estimates of $\omega_u$) with uncorrelated error terms, regardless of whether that model adequately represents the data.

Alternatively, other resources (Gignac, 2014; Green & Yang, 2015; Reise et al., 2013; Rodriguez et al., 2016; Watkins, 2017) advocate estimating the reliability of a multidimensional composite score with respect to the measurement of a general factor common to all items; that is, $\omega_H$, by fitting a bifactor model to item-level scores. Yet, other work has cautioned against the overuse of bifactor models (e.g., Bonifay et al., 2017; Markon, 2019; Reise et al., 2016). Simulations have shown that a bifactor model can produce as good or better fit statistics than the correct model when fit to data from unidimensional, two-factor, and higher-order populations (e.g., Maydeu-Olivares & Coffman, 2006; Morgan et al., 2015; Murray & Johnson, 2013); Bonifay and Cai (2017) found that even when data were generated to follow random patterns, the bifactor model had good fit to a high percentage of samples.

Therefore, it is possible that researchers will report an omega estimate based on an incorrect measurement model and consequently reach an inaccurate conclusion about composite reliability. That is, the true proportion of composite score variance due to a target construct that influences all items may be best defined by a one-factor model (with or without error correlations), a bifactor model, or higher-order model, whereas researchers may be likely to estimate this proportion using an omega estimate calculated from a misspecified model. As yet, the degree to which different coefficient omega statistics differ from the proportion of composite variance due to a target construct when the model is incorrectly specified is not fully known, as previous studies on this issue are limited. Therefore, the main purpose of the current study is to investigate the impact of major model misspecification on the accuracy of $\omega_u$ and $\omega_H$ estimates as measures of the proportion of composite score variance due to a factor influencing all items. In practice, however, it may be that major model misspecification would be detected by common model fit statistics (e.g., root mean square error of approximation [RMSEA], comparative fit index [CFI]), thus leading researchers to revise a hypothesized model (e.g., by freeing error correlation parameters) and then calculate a more accurate omega estimate from the parameter estimates of the revised model. Hence, a secondary purpose of the current study is to investigate the associations between the accuracy of omega estimates and popular model fit statistics.

Previous studies have found that omega estimates are generally unbiased under correct model specification (e.g., Yang & Green, 2010; Zinbarg et al., 2006). Zinbarg et al. (2006) simulated data from a higher-order population model and found that $\omega_H$ produced relatively unbiased estimates of the proportion of variance explained by the higher-order factor, especially with higher values of population $\omega_{ho}$ and longer composites; however, this study did not include $\omega_u$ estimates and used a very low number of replications. Yang and Green (2010) found that ignoring non-zero error covariances led to positively biased estimates of $\omega_u$, although the relative bias was only around 5% with six-item composites and decreased to approximately 2% with a 12-item composite. Yang and Green also simulated data from a bifactor model, but compared $\omega_u$ estimates from a misspecified unidimensional model with a population analog of omega-total (i.e., the proportion of composite variance due to both general and specific factors) rather than $\omega_H$. In the current study, we are instead interested in comparing $\omega_u$ estimates with population $\omega_H$ to determine how well $\omega_u$ estimates the proportion of variance explained by a single factor that is common to all items despite population-level multidimensionality.

To address the issues described above, the current paper presents a Monte Carlo simulation study of the finite sample properties of coefficient omega estimates of the proportion of composite score variance due to a single factor that influences all items as a function of correct and incorrect model specification. Specifically, this study investigated the following research questions:

**Research Question 1 (RQ1):** How well does $\omega_u$ estimate the proportion of variance due to a single factor when non-zero error covariances are ignored or the true model is multidimensional?

**Research Question 2 (RQ2):** How well does $\omega_H$ estimate the proportion of variance due to a single factor when the true model is not a bifactor model?

**Research Question 3 (RQ3):** To what extent is the effect of model misspecification ameliorated if the denominator of omega estimates is the observed composite score variance instead of model-implied variance?
**Research Question 4 (RQ4):** To what extent is the degree of bias related to goodness of fit statistics used to assess model fit?

Our study further considers the effects of the composite length (i.e., number of items), magnitude of population omega, and sample size. We expected that omega estimates calculated from a correctly specified measurement model would be unbiased, but that unmodeled complexity (i.e., incorrectly fitting a unidimensional model to a multidimensional test or incorrectly fixing all error covariances to zero) would introduce considerable bias for $\omega_u$ estimates of the proportion of variance due to a factor common to all items. Furthermore, we expected that $\omega_H$ estimates would provide reasonably accurate estimates of the proportion of variance due to a factor common to all items even when the true model is not a bifactor model (e.g., a higher-order factor model).

## Method

To investigate our research questions about $\omega_u$ and $\omega_H$ estimates, a series of Monte Carlo simulations were run using the SimDesign package in R (Chalmers & Adkins, 2020; R Core Team, 2020); all simulation code is publicly posted at https://osf.io/k7jtz/. Sample data were drawn from multivariate normal distributions with covariance structures consistent with given population CFA models using the mvrnorm function of the MASS package (Venables & Ripley, 2002). Models were estimated using the maximum likelihood estimator in the lavaan package (Rosseel, 2012) and $\omega_u$ and $\omega_H$ were estimated using semTools (Jorgensen et al., 2020). The proportion of total score variance due to the factor common to all items, which we refer to as *population reliability* from here forward, was calculated for each population model and compared with sample omega estimates for 1,000 random samples for each cell of the study design. When non-converged or improper model solutions were obtained for a given cell of the study design, additional replications were drawn until the total number of converged replications with proper solutions equalled 1,000. In total, there were four population factor structures and both a high and low reliability model were generated for both long and short composites. Each factor model was estimated with three sample sizes, creating a total of $(4 \times 2 \times 2 \times 3) = 48$ unique cells.

### *Study Conditions*

Sample data were generated from four population models: a simple one-factor model with no correlated errors, a one-factor model with correlated errors, a bifactor model, and a higher-order model; path diagrams of these models are in Figures 1 through 4. All models were specified such that factor variances equaled 1.0; consequently, the

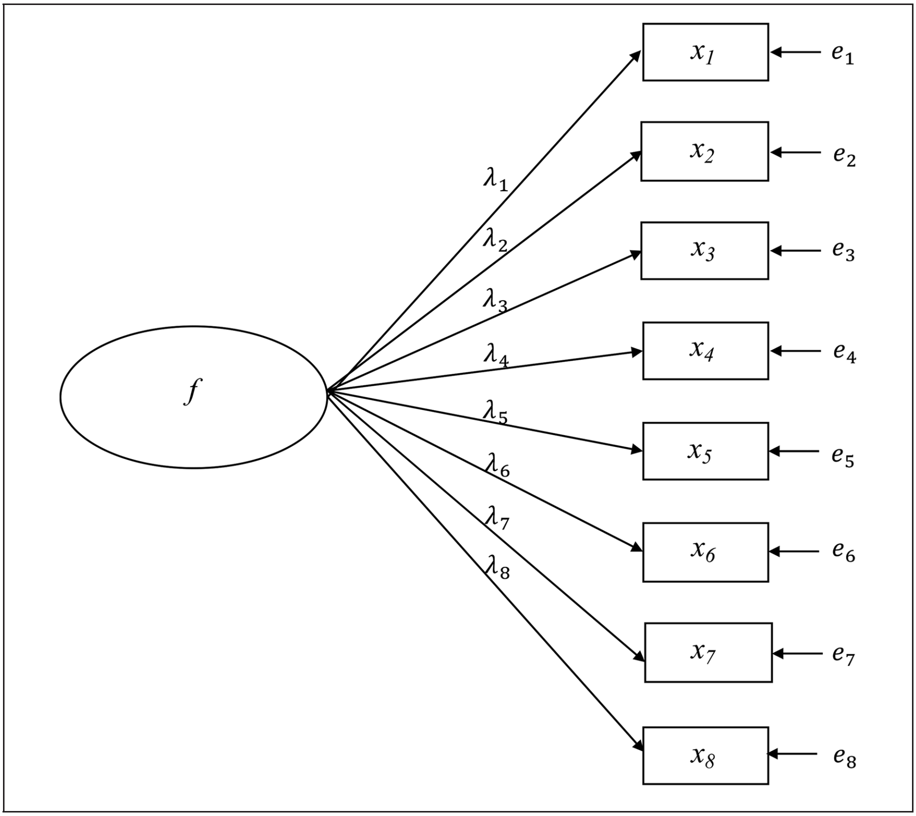


**Figure 1.** Simple One-Factor Population Model.

population-level model-implied covariance structures were in the correlation metric. Samples generated from the simple one-factor model were fit only to the correct model across replications. For samples drawn from all other population models, a simple one-factor model, a correlated errors one-factor model, and a bifactor model were fit to the sample data. Therefore, data from the correlated one-factor and bifactor population models were fit to a correctly specified model as well as two incorrectly specified models, while data from the higher-order populations were fit only to incorrectly specified models.

For each population model structure, there were two conditions of population reliability (i.e., population omega) determined as a function of factor loading parameters. The high-reliability condition set population reliability to .85 while population reliability was .60 in the low-reliability condition. Scale lengths were either short (8 items) or long (16 items) except for the higher-order model, where scale lengths were necessarily longer to ensure enough indicators per factor for model identification. For the higher-order population condition, the short scale was 12 items

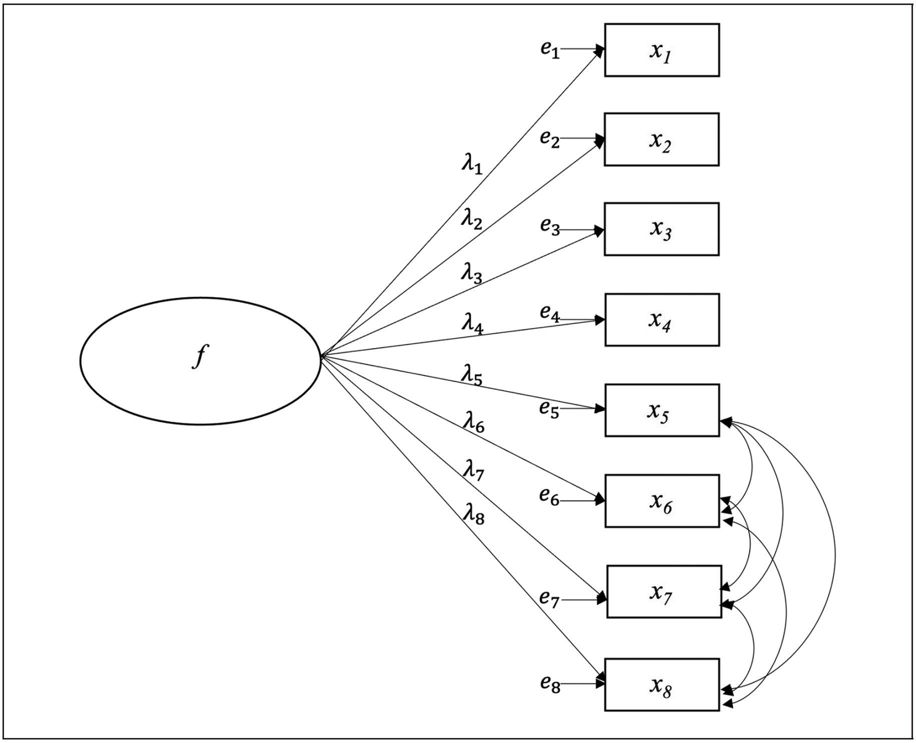


**Figure 2.** One-Factor Population Model With Correlated Errors.

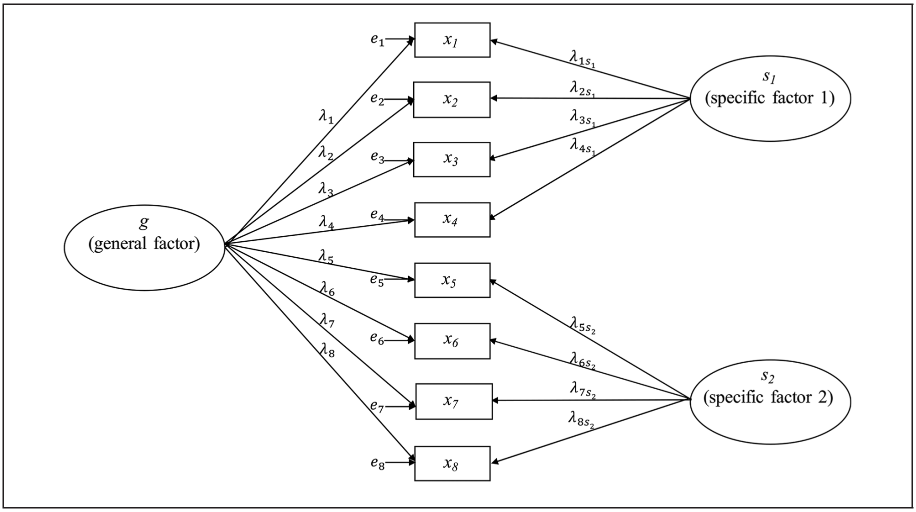


**Figure 3.** Bifactor Population Model.

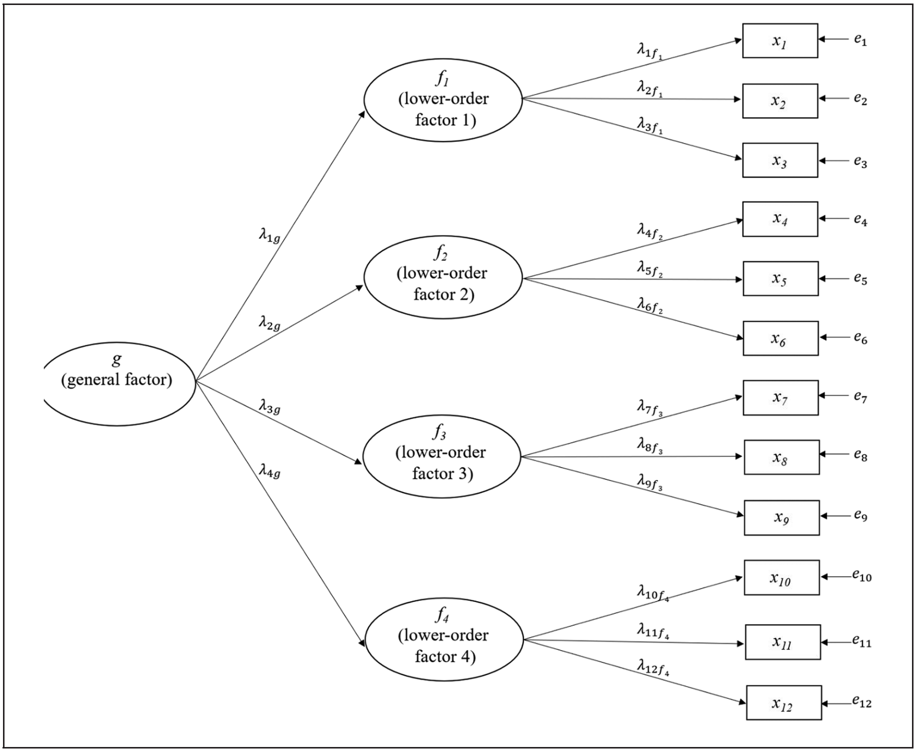


**Figure 4.** Higher-Order Population Model.

(three indicators per lower-order factor) and the long scale was 20 items (five indicators per lower-order factor). We chose these values of population reliability and composite length to be representative of situations commonly encountered in practice, as described by Flake et al. (2017). There were three sample size conditions: $N$ = 100, selected to reflect what is often considered a small sample for the purpose of CFA; $N$ = 250, a medium-sized, commonly observed sample size; and $N$ = 1,000, which is typically considered a large sample for CFA.

Table 1 gives the factor loading parameter values for each model. Factor loadings for the correlated one-factor model ranged from .493 to .837 in the high-reliability conditions and from .493 to .624 in the low-reliability conditions. Half of the item errors for the correlated errors model were allowed to correlate (see Figure 2), as allowing all errors to freely correlate would have produced under-identified models. The error correlations were small to moderate in the high-reliability conditions (approximately .09–.31) and moderate to high in the low-reliability conditions (approximately .19–.52). In a typical bifactor model, every item loads onto both a general factor and a specific factor. However, preliminary simulations showed that bifactor models with two specific factors fitted to data from the one-factor, correlated

**Table 1.** Population Factor Loadings and Error Covariances for Each True Model.

| Model | Factor loading matrix dimension | Population factor loadings | Error covariances |
|---|---|---|---|
| Simple one factor, 8 items, low reliability | 8×1 | .414, .210, .472, .416, .325, .504, .301, .521 | (None) |
| Simple one factor, 8 items, high reliability | 8×1 | .847, .423, .870, .516, .648, .721, .322, .743 | (None) |
| Simple one factor, 16 items, low reliability | 16×1 | .245, .423, .229, .316, .420, .414, .224, .403, .312, .248, .331, .266, .391, .202, .165, .104 | (None) |
| Simple one factor, 16 items, high reliability | 16×1 | .745, .423, .730, .416, .628, .514, .324, .643, .612, .548, .331, .246, .741, .502, .365, .316 | (None) |
| Correlated errors, 8 items, low reliability | 8×1 | .504, .293, .412, .506, .451, .574, .434, .610 | .287, .343, .334, .232 .197, .312 |
| Correlated errors, 8 items, high reliability | 8×1 | .823, .593, .837, .706, .751, .834, .746, .744 | .187, .263, .264, .232, .197, .282 |
| Correlated errors, 16 items, low reliability | 16×1 | .504, .293, .412, .396, .401, .574, .414, .410, .471, .532, .624, .523, .339, .419, .331, .345 | .347, .413, .294, .418, .524, .359, .236, .232, .217, .214, .227, .335, .304, .322, .344, .220, .381, .392, .288, .271, .401, .310, .447, .313, 524, .232, .357, 392 |
| Correlated errors, 16 items, high reliability | 16×1 | .504, .293, .412, .396, .401, .574, .414, .410, .471, .532, .624, .523, .339, .419, .331, .345 | .347, .413, .294, .418, .524, .359, .236, .232, .217, .214, .227, .335, .304, .322, .344, .220, .381, .392, 288, 271, 401, .310, .447, .313, .524, .232, .357, .392 |
| Bifactor, 8 items, low reliability | 8×3 | .462, .340, .420, .659, .314, .501, .608, .410, .426, .414, .279, .338, 0, 0, 0, 0, 0, 0, 0, 0, .314, .448, .213, .417 | (None) |
| Bifactor, 8 items, high reliability | 8×3 | .852, .736, .868, .612, .913, .704, .719, .611, .426, .414, .279, .338, 0, 0, 0, 0, 0, 0, 0, 0, .314, .448, .213, .417 | (None) |

*(continued)*

**Table 1.** (continued)

| Model | Factor loading matrix dimension | Population factor loadings | Error covariances |
|---|---|---|---|
| Bifactor, 16 items, low reliability | 16×3 | .432, .536, .711, .312, .313, .501, .415, .631, .653, .467, .358, .233, .448, .535, .317, .502, .426, .474, .376, .598, .451, .526, .663, .424, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0, .519, .448, .414, .618, .269, .367, .389, .420 | (None) |
| Bifactor, 16 items, high reliability | 16×3 | .852, .636, .868, .612, .513, .704, .719, .831, .653, .774, .358, .721, .448, .835, .612, .718, .426, .374, .276, .238, .351, .248, .319, .314, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, .317, .448, .314, .418, .269, .227, .389, .420 | (None) |
| Higher-order, 12 items, low reliability | 12×4×1 | .60, .72, .64, .62, .68, .73, .83, .54, .64, .44, .75, .66, .59, .62, .69, .61 | (None) |
| Higher-order, 12 items, high reliability | 12×4×1 | .84, .87, .89, .82, 84, .86, .83, .91, .89, .84, .85, .84, .79, .72, .91, .81 | (None) |
| Higher-order, 20 items, low reliability | 20×4×1 | .61, .53, .79, .49, .71, .71, .73, .76, .65, .68, .58, .69, .49, .72, .65, .66, .74, .58, .54, .61, .59, .64, .51, .62 | (None) |
| Higher-order, 20 items, high reliability | 20×4×1s | .84, .87, .89, .82, .84, .86, .83, .91, .86, .84, .85, .84, .76, .81, .68, .87, .74, .80, .54, .88, .79, .72, .89, .82 | (None) |

errors model could not converge consistently. Therefore, only one specific factor capturing these error correlations was included along with the general factor when bifactor models were fitted to data from the one-factor, correlated errors model.

The bifactor population models were specified to include two specific factors pertaining to equal composite halves and a single general factor influencing all items (see Figure 3). The correlated errors model fit to sample data from a bifactor population model allowed all items within each half to correlate with one another, but not with items from the other half of the composite. See Table 1 for population factor loading values across the high- and low-reliability conditions.

Finally, the higher-order population model included a single higher-order factor and four lower-order factors; four is the smallest number of lower-order factors by which a higher-order model can be empirically distinguished from a correlated-factor model (Rindskopf & Rose, 1988). See Table 1 for population factor loading values.

### *Evaluation of Results*

For each replication, $\omega_u$ and $\omega_H$ estimates (i.e., $\hat{\omega}_u$ and $\hat{\omega}_H$) were calculated and compared with population reliability, that is, the omega parameter of the correct, data-generating population model (except $\hat{\omega}_H$ was not calculated for data generated from the one-factor model with independent errors given the gross over-specification of a bifactor model for this condition). Omega estimates were calculated using both model-implied and observed total variance in the denominator. Bias for each estimate within each condition was calculated as the mean difference between the omega estimate and population reliability:

$$\text{bias}(\hat{\omega}) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\omega}_r - \omega),$$

where $R$ is the number of replications, $\hat{\omega}_r$ is the omega estimate for replication $r$, and $\omega$ is the population reliability from the correct data-generating model. We assessed precision using the root mean squared error (RMSE) for omega estimates relative to population reliability with

$$\text{RMSE}(\hat{\omega}) = \sqrt{\frac{1}{R}\sum\nolimits_{r=1}^{R}(\hat{\omega}_r - \omega)^2} = \sqrt{\text{bias}(\hat{\omega})^2 + \text{VAR}(\hat{\omega})}.$$

We also present side-by-side boxplots of omega estimates across conditions to convey the estimate distributions graphically.

## Results

### *Convergence and Proper Solutions*

Table 2 shows the proportions of replications that failed to converge to a proper solution per study condition. In general, estimation of bifactor models was most likely to produce convergence failures, regardless of population model, as bifactor models

**Table 2.** Proportions of Replications That Failed to Converge With Proper Solutions Per Condition.

| | Correlated errors population | | | | Bifactor population | | | | Higher-order population | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample model | Correlated | | Bifactor | | Correlated | | Bifactor | | Correlated | | Bifactor | |
| Test length | 8 items | 16 items | 8 items | 16 items | 8 items | 16 items | 8 items | 16 items | 12 items | 20 items | 12 items | 20 items |
| High reliability | | | | | | | | | | | | |
| $N$ = 100 | .00 | .00 | .002 | .37 | .00 | .06 | .48 | .39 | .002 | .001 | .14 | .10 |
| $N$ = 250 | .00 | .00 | .00 | .18 | .00 | .00 | .20 | .19 | .00 | .00 | .01 | .004 |
| $N$ = 1,000 | .00 | .00 | .00 | .01 | .00 | .00 | .02 | .02 | .00 | .00 | .00 | .00 |
| Low reliability | | | | | | | | | | | | |
| $N$ = 100 | .03 | .00 | .13 | .29 | .05 | .29 | .69 | .27 | .04 | .11 | .48 | .23 |
| $N$ = 250 | .001 | .00 | .02 | .21 | .00 | .09 | .50 | .06 | .00 | .001 | .10 | .12 |
| $N$ = 1,000 | .00 | .00 | .00 | .59 | .00 | .00 | .17 | .00 | .00 | .00 | .00 | .00 |

*Note.* For each cell, additional replications were drawn for each solution that failed to converge or produce a proper solution. Proportions represent the number of improper or non-converged solutions divided by the total number of replications drawn for that respective cell. *High reliability* refers to population reliability = .85; *low reliability* refers to population reliability = .60. All samples fit to a simple one-factor model produced proper solutions regardless of population model. Only three one-factor models produced improper solutions when fit to the correct population model when true reliability was low ($\rho_{xx}$ = .60) and sample size was small ($N$ = 100).

**Table 3.** Mean Bias and RMSE of Omega-Unidimensional Estimates ($\hat{\omega}_u$) for Samples Drawn From a Simple One-Factor Population Model.

| Condition | $\hat{\omega}_\Sigma$ bias | $RMSE_{\hat{\omega}_\Sigma}$ | $\hat{\omega}_S$ bias | $RMSE_{\hat{\omega}_S}$ |
|---|---|---|---|---|
| 8 items | | | | |
| $N$ = 100 | | | | |
| High reliability | .00 | .02 | .00 | .02 |
| Low reliability | .00 | .06 | .00 | .06 |
| $N$ = 250 | | | | |
| High reliability | .00 | .01 | .00 | .01 |
| Low reliability | .00 | .04 | .00 | .04 |
| $N$ = 1,000 | | | | |
| High reliability | .00 | .01 | .00 | .01 |
| Low reliability | .00 | .02 | .00 | .02 |
| 16 items | | | | |
| $N$ = 100 | | | | |
| High reliability | .00 | .02 | .00 | .02 |
| Low reliability | −.01 | .06 | −.01 | .07 |
| $N$ = 250 | | | | |
| High reliability | .00 | .01 | .00 | .01 |
| Low reliability | .00 | .04 | .00 | .04 |
| $N$ = 1,000 | | | | |
| High reliability | .00 | .01 | .00 | .01 |
| Low reliability | .00 | .02 | .00 | .02 |

*Note.* $\hat{\omega}_\Sigma$ represents omega estimates calculated using the model-implied total variance as the equation denominator; $\hat{\omega}_S$ represents the estimates calculated using the observed total variance. *High reliability* refers to population reliability = .85; *low reliability* refers to population reliability = .60.

fitted to data from a bifactor population had the highest error rates. Non-convergence was mitigated by sample size, such that an increase in sample size produced fewer errors for all conditions. The fewest errors occurred when estimating one-factor models with no correlated errors, where only three replications failed to produce proper solutions, all from samples of $N = 100$.

## Simple One-Factor Population Model

Table 3 shows the mean bias and RMSE of $\hat{\omega}_u$ obtained from correctly specified one-factor models with no correlated errors while the distributions of bias for each condition within each sample size are shown in Figure 5. The $\hat{\omega}_u$ estimates were unbiased on average for all conditions, with the highest bias observed in the $N = 100$, 16-item, low reliability condition, where $\hat{\omega}_u$ underestimated reliability by only approximately .01 on average. The RMSE of $\hat{\omega}_u$ was most strongly related to sample size and population reliability, such that RMSE decreased as reliability and sample size increased. Overall, given a reasonable sample size, $\hat{\omega}_u$ produced good reliability estimates in the ideal circumstance of a correctly specified one-factor model with independent errors.

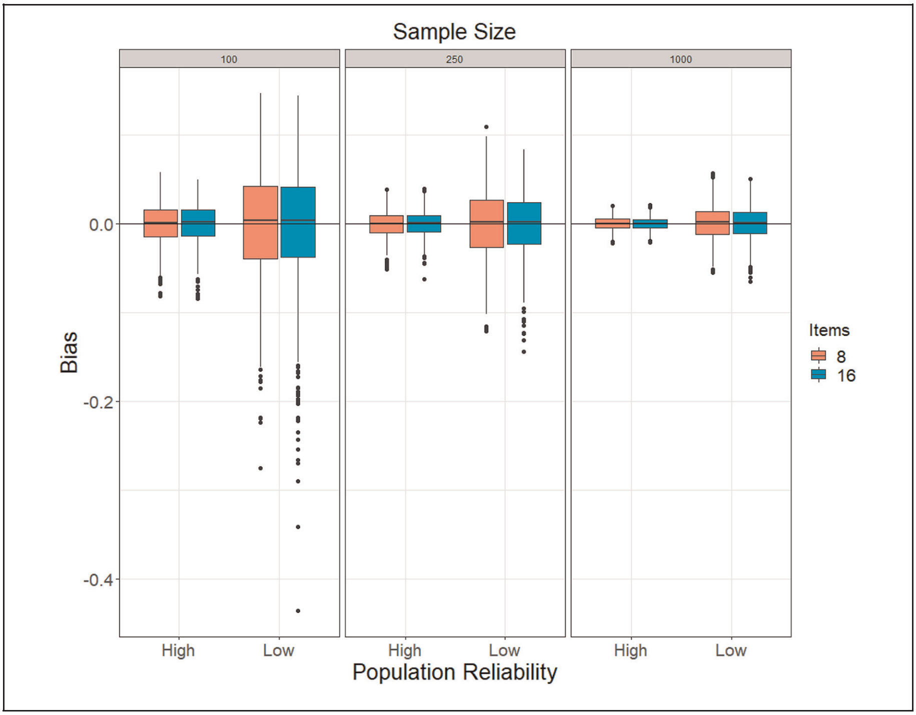


**Figure 5.** Boxplots of Bias of $\omega_u$ Estimates (Using the Model-Implied Total Variance as Its Denominator) Obtained in the Population One-Factor Model (No Correlated Errors) Conditions.

## *One-Factor Population Model With Correlated Errors*

The mean bias and RSME for omega estimates from a one-factor population model with correlated errors are shown in Table 4, and the distribution of bias with $N = 250$ is shown in Figure 6 (similar patterns were observed across sample sizes; see online supplement for figures with $N = 100$ and $N = 1{,}000$). Estimating the correct model yielded unbiased $\hat{\omega}_u$ estimates on average, regardless of condition. Fitting a bifactor model to the sample data resulted in $\hat{\omega}_H$ estimates showing similar performance with $\hat{\omega}_u$ (calculated with correctly specified error correlations). Only the low reliability, long test condition showed a small bias, which was approximately .01. For both $\hat{\omega}_u$ and $\hat{\omega}_H$, RMSE decreased with higher sample sizes and higher population reliability; RMSE ranged from .02 (for $\hat{\omega}_u$) or .03 (for $\hat{\omega}_H$) to .09 in the low-reliability conditions, and from .01 to .03 when population reliability was high. Incorrect specification of a fitted model as bifactor, therefore, produced slightly worse estimates than the correct model in some conditions, but the overall difference was close to negligible. In general, $\hat{\omega}_H$ produced reasonable estimates of population reliability when estimated with data from a one-factor population with correlated errors.

**Table 4.** Mean Bias and RMSE of Omega Estimates for Samples Drawn From a One-Factor Population Model With Correlated Errors.

| | $\hat{\omega}_u$ estimates (misspecified one-factor model with independent errors) | | $\hat{\omega}_u$ estimates (correctly specified one-factor model with correlated errors) | | $\hat{\omega}_H$ estimates (misspecified bifactor model) | |
|---|---|---|---|---|---|---|
| Condition | $\hat{\omega}_\Sigma$ bias (RMSE) | $\hat{\omega}_S$ bias (RMSE) | $\hat{\omega}_\Sigma$ bias (RMSE) | $\hat{\omega}_S$ bias (RMSE) | $\hat{\omega}_\Sigma$ bias (RMSE) | $\hat{\omega}_S$ bias (RMSE) |
| 8 items | | | | | | |
| *N* = 100 | | | | | | |
| High reliability | .07 (.08) | .06 (.06) | **.00 (.03)** | **.00 (.03)** | .00 (.03) | .00 (.03) |
| Low reliability | .17 (.17) | .15 (.16) | **.00 (.09)** | **.00 (.09)** | .00 (.09) | .00 (.09) |
| *N* = 250 | | | | | | |
| High reliability | .08 (.08) | .06 (.06) | **.00 (.02)** | **.00 (.02)** | .00 (.02) | .00 (.02) |
| Low reliability | .17 (.17) | .15 (.15) | **.00 (.05)** | **.00 (.05)** | .00 (.06) | .00 (.06) |
| *N* = 1,000 | | | | | | |
| High reliability | .08 (.08) | .06 (.06) | **.00 (.01)** | **.00 (.01)** | .00 (.01) | .00 (.01) |
| Low reliability | .17 (.17) | .15 (.15) | **.00 (.03)** | **.00 (.03)** | .00 (.03) | .00 (.03) |
| 16 items | | | | | | |
| *N* = 100 | | | | | | |
| High reliability | .09 (.09) | .07 (.07) | **.00 (.03)** | **.00 (.03)** | .00 (.03) | .00 (.03) |
| Low reliability | .25 (.25) | .22 (.22) | **.00 (.08)** | **.00 (.08)** | .01 (.09) | .01 (.09) |
| *N* = 250 | | | | | | |
| High reliability | .09 (.09) | .07 (.07) | **.00 (.02)** | **.00 (.02)** | .00 (.02) | .00 (.02) |
| Low reliability | .25 (.25) | .22 (.22) | **.00 (.05)** | **.00 (.05)** | .02 (.06) | .01 (.06) |
| *N* = 1,000 | | | | | | |
| High reliability | .09 (.09) | .07 (.07) | **.00 (.01)** | **.00 (.01)** | .00 (.01) | .00 (.01) |
| Low reliability | .25 (.25) | .22 (.22) | **.00 (.02)** | **.00 (.02)** | .01 (.03) | .01 (.03) |

*Note.* Bold values represent results for the correctly specified model. $\hat{\omega}_\Sigma$ represents coefficient omega calculated using the model-implied total variance as the equation denominator. $\hat{\omega}_S$ represents coefficient omega calculated using the observed total variance as the equation denominator. *High reliability* refers to population reliability = .85; *low reliability* refers to population reliability = .60. RMSE = root mean squared error.

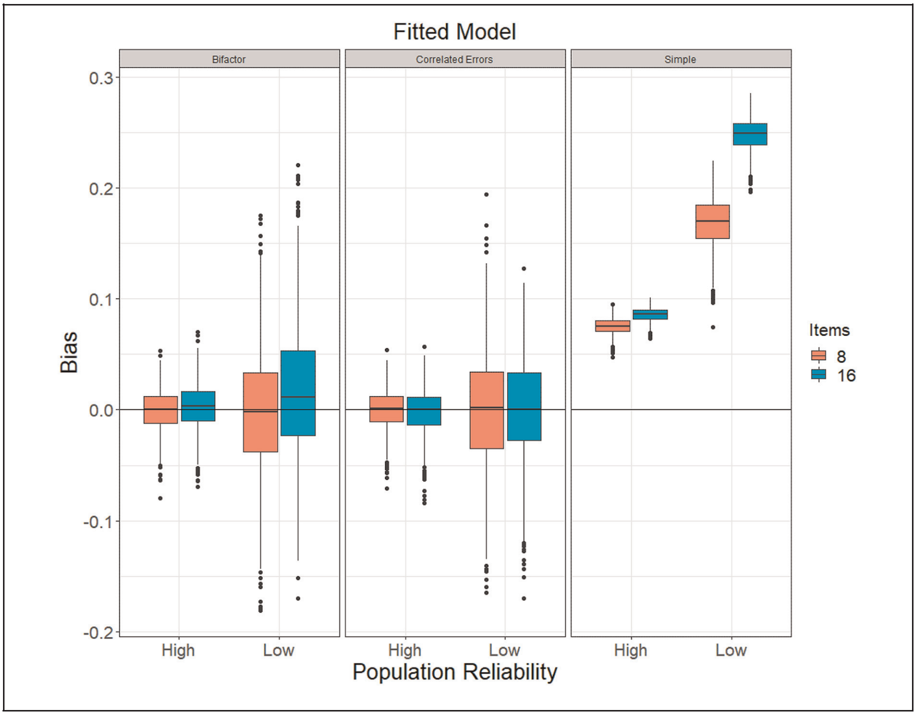


**Figure 6.** Boxplots of Bias of $\omega$ Estimates (Using the Model-Implied Total Variance as its Denominator) by Fitted Model Specification Obtained With Samples of N = 250 Drawn From the Population One-Factor Model, Correlated Error Conditions. Middle Panel Corresponds to the Correct Model Specification Condition. "Simple" Refers to a One-Factor Model With No Error Correlations.

The misspecified, simple one-factor model produced highly biased $\hat{\omega}_u$ estimates, on average. This bias was lowest in the 8-item condition with high population reliability, between .06 and .09. Bias of $\hat{\omega}_u$ was much higher in low-reliability conditions, ranging from .15 to .17 in the 8-item condition and from .22 to .25 in the 16-item condition. Bias was slightly lower when $\hat{\omega}_u$ was calculated using the observed total variance denominator, but this improvement was not enough to produce a reasonably unbiased estimate. Overall, $\hat{\omega}_u$ showed very poor performance when correlated errors were not correctly specified.

## *Bifactor Population Model*

Table 5 shows the mean bias and RMSE of omega estimates from a bifactor population model and Figure 7 shows the distributions of bias with $N = 250$ (see online supplement for figures with $N$ = 100 and $N$ = 1,000). A correctly specified bifactor

**Table 5.** Mean Bias and RMSE of Omega Estimates for Samples Drawn From a Bifactor Population Model.

| | $\hat{\omega}_u$ estimates (misspecified one-factor model with independent errors) | | $\hat{\omega}_u$ estimates (misspecified one-factor model with correlated errors) | | $\hat{\omega}_H$ estimates (correctly specified bifactor model) | |
|---|---|---|---|---|---|---|
| Condition | $\hat{\omega}_\Sigma$ bias (RMSE) | $\hat{\omega}_S$ bias (RMSE) | $\hat{\omega}_\Sigma$ bias (RMSE) | $\hat{\omega}_S$ bias (RMSE) | $\hat{\omega}_\Sigma$ bias (RMSE) | $\hat{\omega}_S$ bias (RMSE) |
| 8 items | | | | | | |
| *N* = 100 | | | | | | |
| High reliability | .08 (.08) | .08 (.08) | .00 (.03) | .00 (.03) | **.00 (.03)** | **.00 (.03)** |
| Low reliability | .15 (.15) | .14 (.15) | .01 (.07) | .01 (.07) | **.00 (.09)** | **.00 (.09)** |
| *N* = 250 | | | | | | |
| High reliability | .08(.08) | .08 (.08) | .00 (.02) | .00 (.02) | **.00 (.02)** | **.00 (.02)** |
| Low reliability | .15 (.15) | .15 (.15) | .01 (.05) | .01 (.05) | **.00 (.05)** | **.00 (.05)** |
| *N* = 1,000 | | | | | | |
| High reliability | .08 (.08) | .08 (.08) | .00 (.01) | .00 (.01) | **.00 (.01)** | **.00 (.01)** |
| Low reliability | .15 (.15) | .15 (.15) | .01 (.02) | .01 (.02) | **.01 (.03)** | **.01 (.03)** |
| 16 items | | | | | | |
| *N* = 100 | | | | | | |
| High reliability | .09 (.09) | .09 (.09) | .02 (.03) | .02 (.03) | **.00 (.03)** | **.00 (.03)** |
| Low reliability | .27 (.28) | .25 (.25) | .12 (.12) | .12 (.13) | **.00 (.09)** | **.00 (.09)** |
| *N* = 250 | | | | | | |
| High reliability | .09 (.10) | .09 (.09) | .02 (.02) | .02 (.03) | **.00 (.02)** | **.00 (.02)** |
| Low reliability | .28 (.28) | .26 (.26) | .11 (.12) | .12 (.12) | **.01 (.05)** | **.01 (.05)** |
| *N* = 1,000 | | | | | | |
| High reliability | .10 (.10) | .09 (.09) | .02 (.02) | .02 (.02) | **.00 (.01)** | **.00 (.01)** |
| Low reliability | .28 (.28) | .26 (.26) | .11 (.11) | .12 (.12) | **.00 (.02)** | **.00 (.02)** |

*Note*. Bold values represent results for the correctly specified model. $\hat{\omega}_\Sigma$ represents coefficient omega calculated using the model-implied total variance as the equation denominator. $\hat{\omega}_S$ represents coefficient omega calculated using the observed total variance as the equation denominator. *High reliability* refers to population reliability = .85; *low reliability* refers to population reliability = .60. RMSE = root mean squared error.

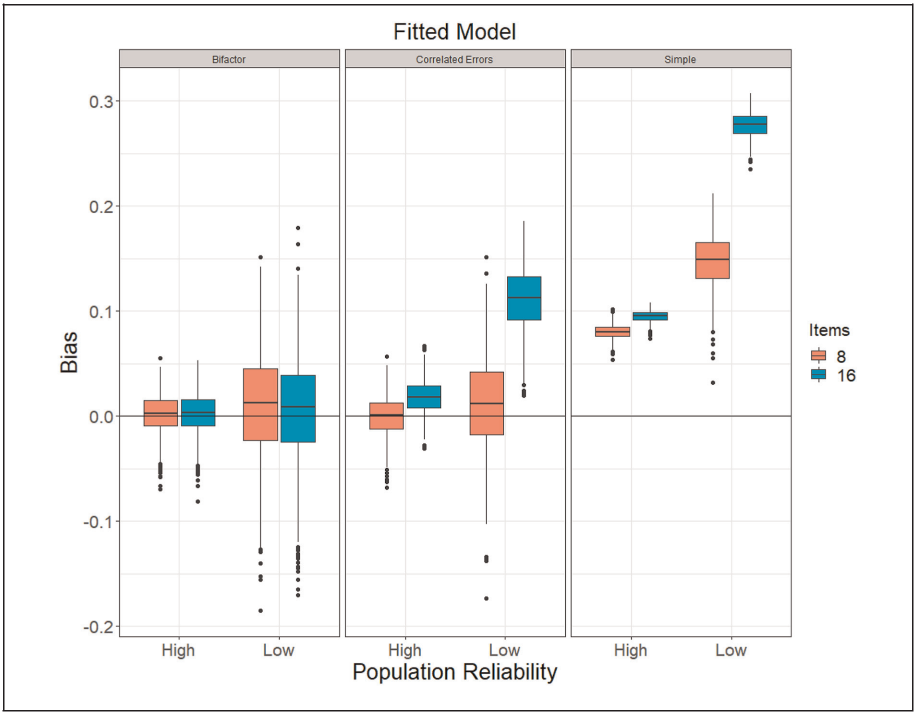


**Figure 7.** Boxplots of Bias of $\omega$ Estimates (Using the Model-Implied Total Variance as Its Denominator) by Fitted Model Specification Obtained With Samples of N = 250 Drawn From the Population Bifactor Model. Left Panel Corresponds to the Correct Model Specification Condition. "Simple" Refers to a One-Factor Model With No Error Correlations.

model produced relatively unbiased $\hat{\omega}_H$ estimates on average. The RMSE of $\hat{\omega}_H$ decreased as population reliability and sample size increased. Overall, $\hat{\omega}_H$ provided good reliability estimates when based on a correctly specified bifactor model.

When $\hat{\omega}_u$ was calculated from one-factor models with specified correlated errors, the estimates were relatively unbiased, on average, in the high reliability, 8-item condition and had a mean bias of .01 in the low reliability, 8-item condition. In the 16-item condition, $\hat{\omega}_u$ had a mean bias of .02 with high reliability; however, mean bias fell between .11 and .12 in the low-reliability, 16-item condition, regardless of sample size. Therefore, $\hat{\omega}_u$ provided good estimates of population reliability only for certain situations when calculated from a one-factor, correlated errors model fitted to data from a bifactor population model.

When $\hat{\omega}_u$ was calculated from simple one-factor models with no free error correlations, mean bias fell between .08 and .10 in the high-reliability condition, with lower bias in the 8-item condition. In the low-reliability condition, mean bias was approximately .15 in the 8-item condition and between .25 and .28 in the 16-item condition.

Therefore, $\hat{\omega}_u$ provided poor estimates of population reliability when calculated from a simple one-factor model fitted to data from a bifactor population model.

### Higher-Order Population Model

The bias and RMSE of omega estimates obtained from the higher-order population model are in Table 6, and distributions of bias are shown in Figure 8 (see online supplement for figures with $N = 100$ and $N = 1,000$). When $\hat{\omega}_u$ was calculated from one-factor models with no free error correlations, mean bias ranged from .07 to .23 across conditions, worsened by both longer scale length and low population reliability. However, specifying correlated errors improved the performance of $\hat{\omega}_u$ such that mean bias ranged from $-.04$ to 0, with the only condition showing an average bias greater than 0 being the low reliability, 16-item condition with $N = 100$. Similarly, $\hat{\omega}_H$ estimates were consistently unbiased except in the 16-item, low reliability condition with $N = 100$, where the mean bias of $\hat{\omega}_H$ was $-.04$. Thus, overall, $\hat{\omega}_H$ produced accurate estimates of population reliability when calculated from a bifactor model fitted to data drawn from a higher-order model, and $\hat{\omega}_u$ produced accurate estimates when calculated from a one-factor model with correlated errors fitted to data drawn from a higher-order model.

### Denominators of Coefficient Omega

For all conditions, omega estimates were calculated using both the model-implied total variance of the composite sum score and the observed sum score variance, as explained earlier. In general, there were only small differences in omega estimates between the two calculations. When models were correctly specified, mean bias and RMSE were nearly identical; when omega estimates were obtained from misspecified models, using the observed variance offered a very slight advantage over the model-implied variance, but differences in mean bias did not exceed .03 in any cell of the study and RMSE differences were negligible.

### Relationship With Model Fit

We also investigated whether the accuracy of omega estimates corresponded to model fit using three major model fit statistics (RMSEA, CFI, and Tucker–Lewis index [TLI]). Because bias could be positive or negative, lower values did not necessarily indicate lower bias. Therefore, we assessed the relation between the absolute value of bias and model fit, but not direction of bias.

Table 7 shows rank-order correlations between descriptive model fit statistics and absolute values of bias of omega estimates within each population model. In general, results for the CFI and TLI were such that better model fit (higher fit index values) was associated with lower bias. Correlations with CFI and TLI were relatively weak ($-.14$ or weaker) when the fitted model was correctly specified, and strongest when

**Table 6.** Mean Bias and RMSE of Omega Estimates for Samples Drawn From a Higher-Order Factor Population Model.

| | $\hat{\omega}_u$ estimates (one-factor model; independent errors) | | $\hat{\omega}_u$ estimates (one-factor model; correlated errors) | | $\hat{\omega}_H$ estimates (bifactor model) | |
|---|---|---|---|---|---|---|
| Condition | $\hat{\omega}_\Sigma$ bias (RMSE) | $\hat{\omega}_S$ bias (RMSE) | $\hat{\omega}_\Sigma$ bias (RMSE) | $\hat{\omega}_S$ bias (RMSE) | $\hat{\omega}_\Sigma$ bias (RMSE) | $\hat{\omega}_S$ bias (RMSE) |
| 12 items | | | | | | |
| $N$ = 100 | | | | | | |
| High reliability | .07 (.08) | .07 (.07) | .00 (.03) | .00 (.03) | .00 (.03) | .00 (.03) |
| Low reliability | .16 (.16) | .15 (.16) | .00 (.08) | .01 (.08) | .00 (.07) | −.01 (.08) |
| $N$ = 250 | | | | | | |
| High reliability | .08(.08) | .07 (.07) | .00 (.02) | .00 (.02) | .00 (.02) | .00 (.02) |
| Low reliability | .16 (.16) | .16 (.16) | .00 (.04) | .00 (.04) | .00 (.04) | .00 (.04) |
| $N$ = 1,000 | | | | | | |
| High reliability | .08 (.08) | .07 (.07) | .00 (.01) | .00 (.01) | .00 (.01) | .00 (.01) |
| Low reliability | .16 (.17) | .16 (.16) | .00 (.02) | .00 (.02) | .00 (.02) | .00 (.02) |
| 20 items | | | | | | |
| $N$ = 100 | | | | | | |
| High reliability | .10 (.10) | .10 (.10) | .00 (.03) | .00 (.03) | .00 (.03) | .00 (.03) |
| Low reliability | .22 (.23) | .19 (.20) | −.02 (.10) | −.04 (.11) | −.02 (.09) | −.04 (.10) |
| $N$ = 250 | | | | | | |
| High reliability | .10 (.10) | .10 (.10) | .00 (.02) | .00 (.02) | .00 (.02) | .00 (.02) |
| Low reliability | .23 (.23) | .20 (.20) | .00 (.04) | .00 (.05) | .00 (.04) | .00 (.05) |
| $N$ = 1,000 | | | | | | |
| High reliability | .10 (.10) | .10 (.10) | .00 (.01) | .00 (.01) | .00 (.01) | .00 (.01) |
| Low reliability | .23 (.23) | .20 (.20) | .00 (.02) | .00 (.02) | .00 (.02) | .00 (.02) |

*Note*: $\hat{\omega}_\Sigma$ represents coefficient omega calculated using the model-implied total variance as the equation denominator. $\hat{\omega}_S$ represents coefficient omega calculated using the observed total variance as the equation denominator. *High reliability* refers to population reliability = .85; *low reliability* refers to population reliability = .60. RMSE = root mean squared error.

**Table 7.** Spearman Correlations Between Bias and Model Fit Indices.

| Population model | Fitted model | RMSEA | CFI | TLI |
|---|---|---|---|---|
| One-factor, independent errors | One-factor, independent errors | .12 | −.07 | −.06 |
| One-factor, correlated errors | One-factor, independent errors | −.60 | −.25 | −.25 |
| | One-factor, correlated errors | .11 | −.14 | −.04 |
| | Bifactor | −.10 | .03 | .03 |
| Bifactor | One-factor, independent errors | −.46 | −.50 | −.48 |
| | One-factor, correlated errors | .34 | −.49 | −.43 |
| | Bifactor | .08 | −.10 | .02 |
| Higher-order | One-factor, independent errors | −.90 | −.74 | −.68 |
| | One-factor, correlated errors | .14 | −.19 | −.11 |
| | Bifactor | .13 | −.18 | −.10 |

*Note.* Tabled values are Spearman correlations between each fit statistic and the absolute value of the bias of omega estimates calculated using the model-implied total variance denominator. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis index.

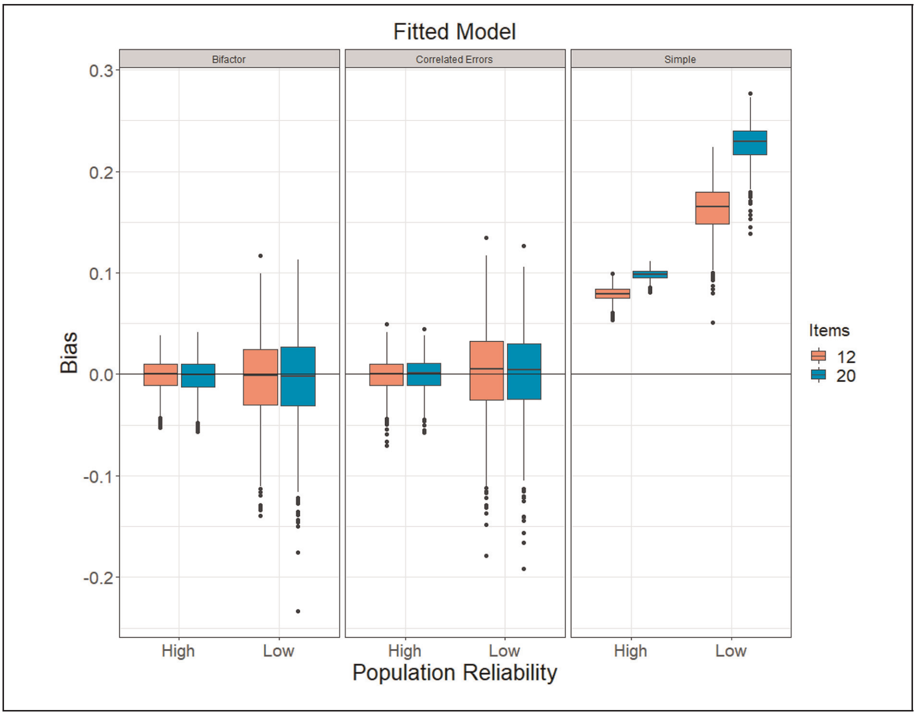


**Figure 8.** Boxplots of Bias of $\omega$ Estimates (Using the Model-Implied Total Variance as Its Denominator) by Fitted Model Specification Obtained With Samples of N = 250 Drawn From the Population Higher-Order Model. "Simple" Refers to a One-Factor Model With No Error Correlations.

simple one-factor models were incorrectly fitted to data drawn from more complex population model structures (from −.25 to −.74). Regarding RMSEA, the correlations generally indicated that better fit (lower values of RMSEA) was weakly associated with lower bias, except for strong, negative correlations between RMSEA and bias occurring when simple one-factor models were incorrectly fitted to data from other population structures. Scatterplots of RMSEA by bias revealed strong, non-monotonic patterns, indicating that these negative correlations occurred because of strong negative associations with RMSEA values ranging from approximately .12 and greater, whereas when RMSEA was in a range generally indicative of good to marginal fit (RMSEA < approximately .10), the association between RMSEA and bias was rather flat.

## Discussion

Coefficient omega has become a popular composite reliability index, as multiple authors have suggested that omega estimates should be broadly preferred over coefficient alpha. Yet, *coefficient omega* does not refer to a single reliability estimate; instead, alternative versions of omega differ depending on the factor-analytic measurement model fitted to item scores (see Flora, 2020). Specifically, the current paper focuses on *omega-unidimensional*, $\omega_u$, which represents the proportion of total score variance due to the single factor of a one-factor model, and *omega-hierarchical*, $\omega_H$, which represents the proportion of total score variance due to the general factor of a bifactor model. Both $\omega_u$ and $\omega_H$ measure the proportion of composite score variance due to a single factor that influences all items; the primary aim of the current study was to determine whether $\omega_u$ and $\omega_H$ can provide unbiased estimates of this proportion when the underlying measurement model is misspecified. Our results are consistent with previous studies, which indicate that omega estimates are generally accurate under correct model specification and that misspecification leads to increased bias as population reliability decreases (e.g., Yang & Green, 2010; Zinbarg et al., 2006).

### *Performance of* $\omega_u$ *Estimates*

Estimates of $\omega_u$ calculated from a correctly specified one-factor model were unbiased on average, regardless of composite length (8 vs. 16 items) and population reliability (.60 vs .85), and showed good precision given an adequately large sample size ($N \geq 250$). However, $\hat{\omega}_u$ tended to produce badly positively biased estimates of the proportion of variance due to a single factor common to all items (a) when the population model consisted of error correlations that were not specified as free parameters in the fitted one-factor model and (b) when the population model was multidimensional (i.e., either a bifactor or higher-order factor structure). These results were not surprising, but given that there are several prominent resources that facilitate calculation of $\hat{\omega}_u$ without first evaluating the suitability of a simple one-factor model for the item-response variables (e.g., Hayes & Coutts, 2020; Kelley, 2022), it is important for the psychometric literature to show that this is a dangerous practice. Our results then showed that $\hat{\omega}_u$ becomes unbiased when error covariances in a one-factor model are correctly specified.

In our one-factor, correlated errors population condition, the error correlations were rather large, having been chosen to represent a situation where item wording effects, such as those obtained with negatively valanced item stems (i.e., items that would typically be reverse scored), are particularly strong. In many applied situations, however, error correlations among item scores are much weaker (e.g., .10 or less). When population error correlations are in this smaller range, the bias in $\hat{\omega}_u$ induced by estimating a one-factor model with independent errors should be smaller than that obtained in the current study; yet, we expect that ignoring small, but non-zero error correlations would still lead to some degree of potentially problematic positive bias for $\hat{\omega}_u$.

Our results support Yang and Green's (2010) observation that failure to specify true correlated errors for a one-factor model will result in substantial bias for $\hat{\omega}_u$. Yang and Green (2010) additionally reported that their omega estimates had lower bias with longer scales under model misspecification. This result differs from our finding that scale length exacerbated bias; however, this difference can be explained by the degree of misspecification. In the current study, bifactor and higher-order population models included two specific and four lower-order factors, respectively, and the unidimensional model with correlated errors included error correlations for half of the items, whereas Yang and Green's (2010) results were based on population models which included only zero to two correlated errors or a bifactor model with only one specific factor. In the current study, more items, therefore, produced more complexity, overriding any stabilizing effect of adding items. The effect of scale length then appears to depend on the degree of misspecification.

In conclusion, researchers should report $\hat{\omega}_u$ as an estimate of composite reliability only when their sample data and previous research provide strong support for the unidimensionality of the composite, that is, when a one-factor model is supported. Furthermore, if there are substantial error covariances among the items over and above the common factor, it is important to account for those terms when calculating $\hat{\omega}_u$ (as demonstrated by Flora, 2020).

## *Performance of* $\omega_H$ *Estimates*

In addition, we assessed how well estimates of $\omega_H$, which is based on the parameters of a bifactor model, represent the proportion of composite variance due to a single factor common to all items under conditions of correct and incorrect model specification. Previous simulations have indicated that bifactor models may be overused and can produce good model fit statistics due to overfitting (e.g., Bonifay & Cai, 2017; Morgan et al., 2015). Yet, our results showed that while $\hat{\omega}_H$ provides unbiased estimates when the bifactor model is correctly specified, $\hat{\omega}_H$ also tends to provide a relatively unbiased estimate of the proportion of composite variance due to a factor common to all items when the population model is either a one-factor model with correlated errors or a higher-order factor model. This finding confirms results in Zinbarg et al. (2006) showing that $\hat{\omega}_H$ produced relatively unbiased estimates of the proportion of variance explained by a higher-order factor; we extended the findings of Zinbarg et al. by including a one-factor, correlated errors population model.

The ability of $\hat{\omega}_H$ to provide minimally biased estimates of $\omega_{ho}$ (the proportion of composite variance due to a population higher-order factor) is especially noteworthy given the conceptual and empirical difficulty of distinguishing between a bifactor model and a higher-order model. Yung et al. (1999) showed that the higher-order model is formally nested within a bifactor model and that a bifactor model can be made equivalent to a higher-order model using a set of proportionality constraints on the factor loadings (based on the transformation proposed by Schmid & Leiman, 1957). But Yung et al. also emphasized that the interpretation of a higher-order factor as a ''superordination'' factor (i.e., a factor with indirect effects on items) is distinct from the ''breadth'' conceptualization of the general factor in a bifactor model (i.e., a factor with direct effects on all items). Nonetheless, both conceptualizations involve a factor that influences all items in a composite (either directly or indirectly) and our results show that $\hat{\omega}_H$ provides reasonable estimates of the proportion of composite variance due to this factor, even when the factor is incorrectly specified as a ''breadth'' factor instead of a ''superordination'' factor.

*Observed Versus Model-Implied Variance.* In that each version of coefficient omega considered herein is a measure of the proportion of composite score variance explained, another research question for the current study was whether there is any advantage to calculating omega estimates as a function of observed composite variance instead of the model-implied composite variance. Our results provided only modest support to the claim by Kelley and Pornprasertmanit (2016) that use of observed total variance is robust to misspecification in that estimates calculated using observed variance were only slightly less biased than estimates based on model-implied variance in the conditions with the most severe model misspecification; yet, Kelley and Pornprasertmanit only studied minor misspecification in their simulations. Bentler (2009) suggested that use of model-implied composite variance would lead to more efficient omega estimates, but our results do not support this proposal, instead showing that omega estimates based on the observed composite variance have comparable or slightly less variation (as shown by our RMSE values) than estimates computed with the model-implied variance.

*Associations With Model Fit.* Finally, we also examined the associations between common descriptive model fit statistics—RMSEA, CFI, and TLI—and the bias of omega statistics as estimates of the proportion of composite variance due to a factor influencing all items. If the bias of omega estimates is highly correlated with model fit, then in practice, a researcher may detect model misspecification by observing a high RMSEA or low CFI, revise the hypothesized model (e.g., by freeing error correlation parameters), and then calculate an appropriate omega estimate from the parameter estimates of the revised model and could thereby avoid calculating a highly biased omega estimate altogether. We did find modest correlations between model fit statistics and bias of omega estimates, suggesting some protection against model misspecification, especially when simple one-factor models were fitted to data drawn from more complex models. But there are situations in which misspecified models still fit sample data well, which could lead to researchers obtaining an omega estimate that overestimates how precisely a composite measures a factor common to all items. In addition to model fit statistics, researchers should employ theory, results of CFAs from previous studies, and model comparison (using statistics such as the Bayesian Information Criterion, BIC) to establish the optimal measurement model for a composite prior to calculating an omega estimate.

## Implications for Applied Research

The current findings have important implications for researchers wishing to estimate composite reliability using coefficient omega. Here, we remind the reader that applied researchers are often interested in determining how reliably a composite measures a target construct, which can be represented by a factor in a factor-analysis model and does not necessarily correspond to the CTT true score (Borsboom, 2005). In the current study, we investigated how well different forms of coefficient omega estimate the proportion of total composite variance that is due to a factor that influences all items in a composite as a function of model misspecification. In other words, we investigated how well omega estimates the reliability of a composite as a measure of a common factor. The high bias of $\hat{\omega}_u$ with data drawn from complex population factor structures (i.e., models with error correlations, bifactor models, and higher-order models) emphasizes the importance of investigating the fit and feasibility of a simple one-factor model *before* calculating $\hat{\omega}_u$. If a one-factor model is rejected, then researchers may respecify the model to free error correlation parameters (if justifiable based on theoretical considerations of item content) and consequently obtain an improved $\hat{\omega}_u$. Alternatively, if a composite is multidimensional, then a bifactor model may be estimated to obtain $\hat{\omega}_H$ as an estimate of composite reliability for the measurement of a general factor; $\hat{\omega}_H$ should still provide a reasonably accurate estimate if a higher-order model is the correct population model rather than a bifactor model.

Regarding software implementation, we used the reliability function of the semTools package in R to obtain all omega estimates in the current study. This function automates the calculation of omega from the estimates from a CFA model previously fitted using the lavaan package, and as such, the user is required to estimate a measurement model for the composite prior to calling the reliability function, which will subsequently return omega values giving the estimated proportion of composite variance explained by each factor in the CFA model. For this reason, we recommend the general use of semTools::reliability as demonstrated in the Flora (2020) tutorial, as users are forced to consider an appropriate CFA model before obtaining omega estimates. The ci.reliability function of the MBESS package (Kelley, 2022) calculates $\hat{\omega}_u$ without providing any indication of the adequacy of a simple one-factor model for the composite's items, whereas the strel function of the Bayesrel package (Pfadt, van den Bergh, Sijtsma, Moshagen, & Wagenmakers, 2022) uses a Bayesian approach to calculate $\hat{\omega}_u$ while the Bayesrel::omega_fit function provides a set of model-fit statistics for the one-factor model (see Pfadt, van den Bergh, Sijtsma, & Wagenmakers,, 2022 for implementation with JASP software). At this time, neither MBESS::ci.reliability nor Bayesrel::strel can account for error correlation parameters. Finally, Hayes and Coutts (2020) present a macro for SAS and SPSS that calculates $\hat{\omega}_u$ without providing any indication of the adequacy of a simple one-factor model.

A more encouraging result from our study is that $\hat{\omega}_H$ provides relatively accurate estimates of the proportion of composite variance due to a factor influencing all items, even when the true factor structure is characterized by a higher-order model or a one-factor model with strong error correlations rather than a bifactor model. Thus, despite the tendency of a bifactor model to have good fit to data generated from different population models, use of $\hat{\omega}_H$ as a composite reliability estimate may still be generally viable in situations where a composite is characterized by multidimensionality, yet researchers still are interested in the composite as a measure of a general construct underlying all items. Flora (2020) explains how the semTools::reliability function will return $\hat{\omega}_H$ as the omega estimate for the general factor based on a bifactor CFA model fitted with lavaan.

Alternatively, output of the omega function of the popular R package psych (Revelle, 2022) provides a statistic labeled ''omega hierarchical'' which is calculated by (a) fitting a EFA model (with three factors by default) to the item-level data, (b) rotating the factors obliquely and determining higher-order factor loadings from the inter-factor correlation matrix, (c) applying the Schmid-Leiman transformation to obtain an exploratory bifactor model, and finally (d) calculating an omega index as a function of the resulting general factor loadings. Thus, psych::omega is an attractive option for researchers who determine that a simple unidimensional model is not appropriate for their composite but are not able to specify a theoretically informed CFA model. Nevertheless, we caution that the omega-hierarchical statistic provided by psych::omega does not correspond to the $\hat{\omega}_H$ statistic studied herein because it is based on (a) the estimates of an EFA model (i.e., a completely free factor pattern and consequent rotational indeterminacy) rather than a more restricted CFA model and (b) a restricted bifactor structure that is equivalent to a higher-order structure due to use of the Schmid-Leiman transformation. Cho (2022) reports promising results for omega estimates based on EFA parameter estimates.

## *Limitations and Directions for Future Research*

This study investigated the accuracy of omega estimates under a range of conditions; however, as with all simulation studies, there are omitted conditions which should be addressed by future studies. Although correct versus incorrect model specification was the primary independent variable of our design, this manipulation was constrained to selection of an incorrect model type (i.e., one-factor models with and without error correlations, bifactor model, or higher-order model). The effect of more minor misspecification—for example, failing to model cross-loadings or small correlated errors within a bifactor model—was not addressed. Further studies should verify and expand on these findings by examining different types and degrees of misspecification. In addition, although the current study involved generating data from a higher-order population model to compare $\hat{\omega}_H$ estimates with population $\omega_{ho}$, we did not also calculate $\hat{\omega}_{ho}$ estimates; thus, future research should investigate finite-sample properties of $\hat{\omega}_{ho}$ under both a correctly specified higher-order model and misspecified models.

Next, we generated data from the multivariate normal distributions implied by our population model specifications. This procedure allowed us to examine the effects of model misspecification (as well as scale length and strength of population omega) without the confound of categorical versus continuous responses. In practice, many composites are comprised of binary or ordered, categorical item response variables which are best modeled with a categorical variable methodology, such as by fitting CFA models to polychoric correlations. In these situations, estimates of coefficient omega should be adapted to enable the CFA model's estimates to be properly scaled into the composite's observed total score metric (see Flora, 2020; Green & Yang, 2009b; Yang & Green, 2010). Therefore, important expansions on the current findings will be to investigate the performance of omega estimates for both continuous and categorical item–response variables under additional conditions of misspecification.

## Conclusion

Numerous authors have advocated for the regular use of coefficient omega to estimate composite reliability in place of more traditional indices such as coefficient alpha, despite that there is relatively little simulation evidence about the finite sample properties of different forms of coefficient omega. Many of these recommendations focus on omega-unidimensional and pay little heed to its dependence on the adequacy of a one-factor model for item-level data; similarly, omega-hierarchical is based on the parameters of a bifactor model, but it has been shown that misspecified bifactor models often fit item-level data well. The current paper presents a simulation study indicating that omega-unidimensional and omega-hierarchical provide unbiased estimates of the reliability of a composite with respect to the measurement of a target construct when the fitted factor model is correctly specified. However, under misspecification, omega-unidimensional provided strongly biased estimates of the reliability of a composite as a measure of a factor common to all items and should therefore only be used in the case of a single-factor congeneric model, with care taken to account for potential error correlations among items. Omega-hierarchical, however, provided relatively unbiased estimates of composite reliability with respect to a general factor influencing all items, even when the population model was a higher-order model or a one-factor model with correlated errors. Researchers who wish to use estimate reliability using omega should therefore take care to ensure adequate sample size ($N \geq 250$) and carefully factor analyze their results to select the best model, based not only on fit indices but also on theory and previous evidence.

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

## ORCID iDs

R. Philip Chalmers https://orcid.org/0000-0001-5332-2810
David B. Flora https://orcid.org/0000-0001-7472-0914

## Supplemental Material

Supplemental material for this article is available online at https://osf.io/k7jtz/.

## Notes

1. The results presented herein generalize if $X_i$ is instead defined as the unweighted mean of the item scores.
2. When is $\sigma_X^2$ estimated using the observed composite variance, Kelley and Pornprasertmanit (2016) refer to the resulting statistic based on the one-factor model as *hierarchical omega*, which is not to be confused with $\omega_H$, the omega-hierarchical parameter, which is based on the parameters of a bifactor measurement model, as defined later.
3. Raykov and Zinbarg (2011) refer to this version of omega, ''the proportion of total scale score variance that is accounted for by the general factor, which could be viewed as common to all components,'' as omega-hierarchical, despite that their presentation is based on a higher-order model instead of a bifactor model.

## References

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74*(1), 137–143. https://doi.org/10.1007/S11336-008-9100-1

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.

Bonifay, W., & Cai, L. (2017). On the complexity of item response theory model. *Multivariate Behavioral Research*, *52*(4), 465–484. https://doi.org/10.1080/00273171.2017.1309262

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, *5*(1), 184–186. https://doi.org/10.1177/2167702616657069

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge University Press.

Borsboom, D., & Mellenbergh, G. J. (2002). True scores, latent variables, and constructs: A comment on Schmidt and Hunter. *Intelligence*, *30*(6), 505–514. https://doi.org/10.1016/S0160-2896(02)00082-X

Chalmers, R. P., & Adkins, M. C. (2020). Writing effective and reliable Monte Carlo simulations with the SimDesign package. *The Quantitative Methods for Psychology*, *16*(4), 248–280. https://doi.org/10.20982/tqmp.16.4.p248

Cho, E. (2022). Reliability and omega hierarchical in multidimensional data: A comparison of various estimators. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000525

Dunn, T. J., Baguley, T., & Brunsden, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Ellis, J. L. (2021). A test can have multiple reliabilities. *Psychometrika*, *86*(4), 869–876. https://doi.org/10.1007/s11336-021-09800-2

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484–501. https://doi.org/10.1177/2515245920951747

Gignac, G. E. (2014). On the inappropriateness of using items to calculate total scale score reliability via coefficient alpha for multidimensional scales. *European Journal of Psychological Assessment*, *30*(2), 130–139. https://doi.org/10.1027/1015-5759/a000181

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement*, *66*(6), 930–944. https://doi.org/10.1177/0013164406288165

Green, S. B., & Yang, Y. (2009a). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, *74*(1), 121–135. https://doi.org/10.1007/S11336-008-9098-4

Green, S. B., & Yang, Y. (2009b). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, *74*(1), 155–167. https://doi.org/10.1007/S11336-008-9099-3

Green, S. B., & Yang, Y. (2015). Evaluation of dimensionality in the assessment of internal consistency reliability: Coefficient alpha and omega coefficients. *Educational Measurement: Issues and Practice*, *34*(4), 14–20. https://doi.org/10.1111/emip.12100

Hancock, G. R., & An, J. (2020). A closed-form alternative for estimating $\omega$ reliability under unidimensionality. *Measurement: Interdisciplinary Research and Perspectives*, *18*, 1–14. https://doi.org/10.1080/15366367.2019.1656049

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2020). *semTools: Useful tools for structural equation modeling* [R package]. http://CRAN.R-project.org/package=semTools

Kelley, K. (2022). *MBESS: The MBESS R package* [R package version 4.8.0]. https://CRAN.R-project.org/package=MBESS

Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods*, *21*(1), 69–92. https://doi.org/10.1037/a0040086

Markon, K. E. (2019). Bifactor and hierarchical models: Specification, inference, and interpretation. *Annual Review of Clinical Psychology*, *15*, 51–69. https://doi.org/10.1146/annurev-clinpsy-050718-095522

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept factor analysis. *Psychological Methods*, *11*(4), 344–362. https://doi.org/10.1037/1082-989X.11.4.344

McDonald, R. P. (1999). *Test theory: A unified approach*. Lawrence Erlbaum.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, *3*(1), 2–20. https://doi.org/10.3390/jintelligence3010002

Murray, A. J., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Journal of Intelligence*, *41*(5), 407–422. https://doi.org/10.1016/j.intell.201306.004

Pfadt, J. M., van den Bergh, D., Sijtsma, K., Moshagen, M., & Wagenmakers, E. J. (2022). Bayesian estimation of single-test reliability coefficients. *Multivariate Behavioral Research*, *57*(4), 620–641. https://doi.org/10.1080/00273171.2021.1891855

Pfadt, J. M., van den Bergh, D., Sijtsma, K., & Wagenmakers, E. J. (2022). A tutorial on Bayesian single-test reliability analysis with JASP. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-021-01778-0

Raykov, T., & Marcoulides, G. A. (2017). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement*, *79*(1), 200–210. https://doi.org/10.1177/0013164417725127

Raykov, T., & Zinbarg, R. E. (2011). Proportion of general factor variance in a hierarchical multiple-component measuring instrument: A note on a confidence interval estimation procedure. *British Journal of Mathematical and Statistical Psychology*, *64*(2), 193–207. https://doi.org/0.1348/000711009X479714

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Reise, S. P., Kim, D. S., Mansolf, M., & Widaman, K. F. (2016). Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg Self-Esteem Scale. *Multivariate Behavioral Research*, *51*(6), 818–838. https://doi.org/10.1080/00273171.2016.1243461

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*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Revelle, W. (2022). *psych: Procedures for personality and psychological research*. Northwestern University. https://cran.r-project.org/web/packages/psych/index.html

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145–154. https://doi.org/10.1007/S11336-008-9102-Z

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

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Rosseel, Y. (2012). lavaan: An R Package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Savalei, V., & Reise, S. P. (2019). Don't forget the model in your model-based reliability coefficients: A reply to McNeish (2018). *Collabra: Psychology*, *5*(1), Article 36. https://doi.org/10.1525/collabra.247

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*(1), 83–90. https://doi.org/10.1007/BF02289209

Sijtsma, K., & Pfadt, J. M. (2021). Part II: On the use, the misuse, and the very limited usefulness of Cronbach's alpha: Discussing lower bounds and correlated errors. *Psychometrika*, *86*(4), 843–860. https://doi.org/10.1007/s11336-021-09789-8

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.

Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. *The Clinical Neuropsychologist*, *31*(6–7), 1113–1126. https://doi.org/10.1080/13854046.2017.1317364

Yang, Y., & Green, S. B. (2010). A note on structural equation modeling estimates of reliability. *Structural Equation Modeling*, *17*(1), 66–81. https://doi.org/10.1080/10705510903438963

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

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ωh. *Applied Psychological Measurement*, *30*(2), 121–144. https://doi.org/10.1177/0146621605278814