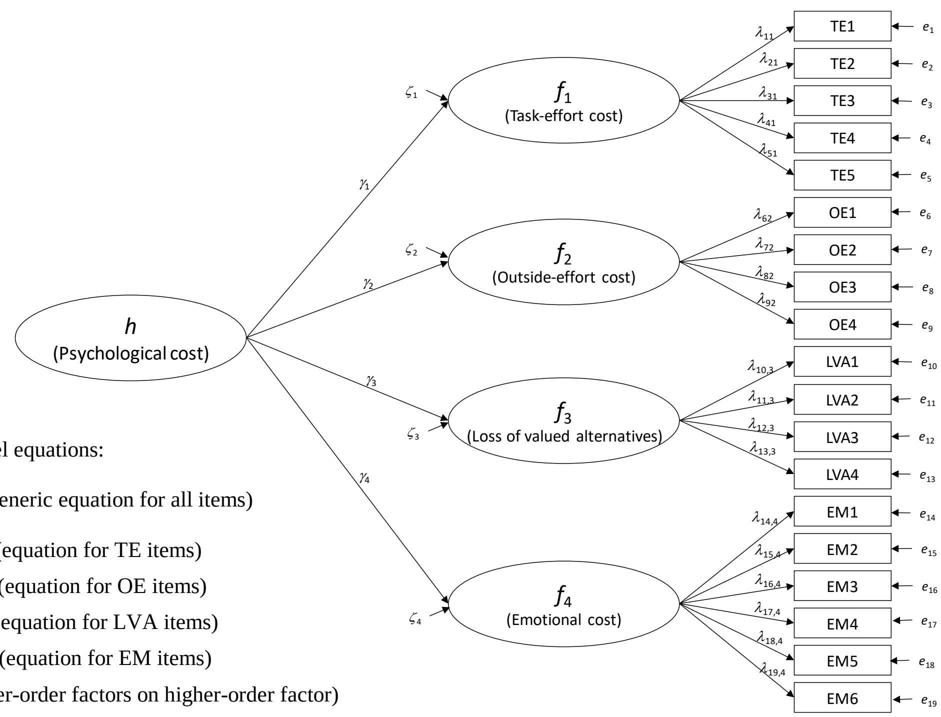
Supplemental material for “Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates”

Higher-order models. As an example of a higher-order model as explained in the main text of the article, Flake et al. (2015) developed a 19-item test to measure a broad construct termed *psychological cost*. Flake et al. designed this test to conform to a higher-order structure in which a cost higher-order factor influences four lower-order factors termed *task-effort cost*, *outside-effort cost*, *loss of valued alternatives*, and *emotional cost*; a path diagram of this model is in the figure below:



Model equations:

$$x_j = \left(\sum_{k=1}^{K} \lambda_{jk} f_k \right) + e_j \text{ (generic equation for all items)}$$

$$TE_j = \lambda_{j1} f_1 + e_j \text{ (equation for TE items)}$$

$$OE_j = \lambda_{j2} f_2 + e_j \text{ (equation for OE items)}$$

$$LVA_j = \lambda_{j3} f_3 + e_j \text{ (equation for LVA items)}$$

$$EM_j = \lambda_{j4} f_4 + e_j \text{ (equation for EM items)}$$

$$f_k = \gamma_k h + \zeta_k \text{ (regression of lower-order factors on higher-order factor)}$$

In this higher-order model, each of the four lower-order factors, f_1 to f_4, directly influences a subset of the 19 items (as depicted by arrows from the four lower-order factors to the items), while the psychological cost higher-order factor influences the lower-order factors (as depicted by arrows from *h* to each of the four lower-order factors). This model is typically represented using two equations (see the figure): In the first equation, the observed item-response variables are regressed on the lower-order factors, with the λ_{jk} factor loadings representing the strength of the association between item j and its respective lower-order factor. Next, instead of allowing the lower-order factors to freely covary with one another, the associations among the lower-order factors are explained by the higher-order factor: Each lower-order factor is regressed on the higher-order factor, with a higher-order factor loading γ_k representing the linear association between the kth lower-order factor and the higher-order factor.[1]

Omega-higher order. When item scores arise from a higher-order model, a reliability measure we term *omega-higher order*, or ω_{ho}, represents the proportion of total score variance that is due to the higher-order factor; parameter estimates from a higher-order model are used to calculate ω_{ho}. The associations between the higher-order factor and the observed item scores are mediated through the lower-order factor. Thus, each indirect effect of the higher-order factor on an item is the product of the item's lower-order factor loading and the corresponding higher-order factor loading (Raykov & Zinbarg, 2011); that is, for item x_j loading on the kth lower-order factor, the indirect effect of the higher-order on x_j equals (λ_{jk}*γ_k). Thus, to represent reliability of a total score for measuring a construct represented by the higher order factor, ω_{ho} is a function of these indirect effects, as shown in the equation below:

$$\omega_{ho} = \frac{\left(\sum_{j=1}^{J} \hat{\lambda}_{jk}\hat{\gamma}_k\right)^2}{\hat{\sigma}_X^2}$$

[1] There must be at least three lower-order factors for a higher-order model to be identified (i.e., a model with only two lower-order factors cannot be uniquely estimated) and there must be at least four lower-order factors for a higher-order model to be empirically distinguishable from a model with no higher-order factor (i.e., a model with three correlated factors is equivalent to a higher-order model with three lower-order factors; Rindskopf & Rose, 1988).

In this formula, the numerator represents the accumulation of the indirect effects across all items, and, as with other forms of coefficient omega for the reliability of a total score *X*, the denominator again represents the estimated variance of the total score *X*.

Example calculation of ω_{ho} in R. To demonstrate the estimation of ω_{ho} using R, we use data from Flake et al. (2017), who administered the 19-item psychological cost scale (PCS) of Flake at al. (2015) to *N* = 154 students in an introductory statistics course. The complete R syntax and output for this example are in the .rmd markdown file (and resulting .pdf) accompanying the article. With these data, α = .96 for the PCS items, but this is likely a misleading reliability estimate because of multidimensionality. We fitted the higher-order model depicted in the figure above to the item-response data using ML, thus treating the item scores as continuous variables because their responses were given using a six-point scale.

Syntax to specify this model for `lavaan` is

```
> homod <- 'TE =~ TE1 + TE2 + TE3 + TE4 + TE5
          OE =~ OE1 + OE2 + OE3 + OE4
          LV =~ LVA1 + LVA2 + LVA3 + LVA4
          EM =~ EM1 + EM2 + EM3 + EM4 + EM5 + EM6
          cost =~ TE + OE + LV + EM'
```

where `TE` (*task-effort cost*), `OE` (*outside-effort cost*), `LV` (*loss of valued alternatives*), and `EM` (*emotional cost*) are four lower-order factors, each measured by the item-level variables listed on the righthand side of their respective `=~` operators. The final line of the syntax specifies that the higher-order factor `cost` is measured by the four lower-order factors. With this higher-order factor included, `lavaan` will restrict the (residual) covariances among the lower-order factors to 0 by default. The model is estimated with

```
> fitHo <- cfa(homod, data=pcs, std.lv=T, estimator='MLM')
```

where the `MLM` estimator requests ML with robust model fit statistics. The results summary indicates that this higher-order model fits the data well with robust model fit statistics CFI = .97, TLI = .97, and RMSEA = .06. Thus, it is reasonable to calculate ω_{ho} to estimate how reliably the 19 items measure the higher-order psychological cost factor.

The `semTools::reliability` function used earlier does not return a reliability estimate based on a higher-order factor; instead, the `reliabilityL2` function of the `semTools` package can calculate ω_{ho}. For the current example, ω_{ho} is obtained with

```
> reliabilityL2(fitHo, 'cost')
```

where the first argument is the name of the higher-order model created above (i.e., `fitHo`) and the second argument is the name of the higher-order factor. The resulting output is

```
       omegaL1        omegaL2 partialOmegaL1
     0.9088177      0.9307391      0.9734520
```

The estimate listed under `omegaL1` represents ω_{ho}, the proportion of PCS total score variance due to the higher-order factor, that is, the overarching psychological cost construct. Therefore, in contrast with $\alpha = .96$ for this psychological-cost scale, when we estimate an omega coefficient that accounts for the scale's multidimensionality as per a well-fitting and theoretically informed higher-order model, we obtain $\omega_{ho} = .91$ as an estimate of the proportion of the scale's total score variance that is attributable to an over-arching psychological-cost construct.

Subscale reliability. Extending the psychological cost example, the higher-order factor structure also suggests that the PCS can be partitioned into four subscales to measure more narrow constructs represented by the four lower-order factors. We can therefore estimate how reliably these subscale scores measure the lower-order factors by applying `semTools::reliability` to the fitted higher-order model object,

```
> reliability(fitHo)
```

producing the following output:

```
              TE        OE        LV        EM
alpha  0.9250420 0.8992820 0.9052459 0.9405882
omega  0.9260207 0.9000550 0.9077522 0.9415490
omega2 0.9260207 0.9000550 0.9077522 0.9415490
```

```
omega3 0.9256736 0.9014417 0.9125254 0.9404925
avevar 0.7155338 0.6925128 0.7111717 0.7299180
```

This output therefore gives ω estimates for subscale scores calculated based on each of the lower-order factors, respectively listed under the `TE`, `OE`, `LV`, and `EM` columns, and as such they are essentially individual ω_u estimates calculated for each subscale separately. For example, ω = .93 for the `TE` factor is a reliability estimate for a total score for the five *task-effort* items only, that is, the task-effort subscale. In other words, 93% of the variance of a total score *calculated using only the five task-effort items* is explained by the `TE` lower-order factor. However, careful interpretation of these ω estimates involves also considering $\omega_{h\text{-}ss}$ estimates for the subscales, as described next.

Omega-hierarchical-subscale. To supplement the PCS example, one might ask how well a given subscale reliably measures a narrower construct that is *independent* from the broader higher-order construct that also influences the other subscales. For instance, although the higher-order model might represent the theoretically correct dimensional structure for the PCS, a bifactor model specification can also provide useful information regarding the reliability of the subscale scores. Because the higher-order model is nested within a bifactor model (Yung et al., 1999), it is reasonable to fit a bifactor model to any test that produces data with a higher-order structure (Rodriguez et al., 2016). Consequently, we can obtain ω_h for the PCS in addition to estimates that have been termed *omega-hierarchical-subscale* (Rodriguez et al.), which we abbreviate $\omega_{h\text{-}ss}$, to assess how well a given subscale score represents reliable variance over and above variance due to the general factor. This interpretation arises because specific factors in the bifactor model are mathematically and conceptually distinct from lower-order factors in the higher-order model: In the bifactor model, specific factors capture *residual* covariance within item sets beyond covariance captured by the general factor because each item is regressed on both the general and a specific factor, but in the higher-order model, each item is directly regressed on only its corresponding lower-order factor. Thus, $\omega_{h\text{-}ss}$ is the reliability of a subscale score for the measurement of a construct represented by the specific factor which is independent from the general factor; in other words, $\omega_{h\text{-}ss}$ represents the proportion of variance in a subscale that is due to the corresponding specific factor, over and above the influence of the general factor. If only a small proportion of the observed variance of a subscale score is due to the corresponding specific factor, then the subscale might simply be thought of as a short-form measure of the construct represented by general factor instead of a measure of a construct which represents content distinct from the general factor, as we show below for the task-effort subscale of the PCS. See DeMars (2013) for further discussion of subscale score interpretation based on bifactor models.

The formula for $\omega_{h\text{-}ss}$ is like that for ω_h, except now the numerator is a function of the specific factor loadings only for the items associated with the subscale and the denominator represents the variance of a total score only for those items (rather than the total score of the entire test):

$$\omega_{h-ss} = \frac{\left(\sum_{m=1}^{M} \hat{\lambda}_{ms}\right)^2}{\hat{\sigma}^2_{SS_k}}.$$

In some situations, the specific factors represent methodological artifacts (i.e., item-wording effects) and the resulting $\omega_{h\text{-}ss}$ estimates are not particularly meaningful substantively. But when tests may produce substantively meaningful subscales, $\omega_{h\text{-}ss}$ estimates are more useful; the PCS examples is one such situation.

Returning to the bifactor model for the PCS, the command

```
> reliability(fitBf)
```

produces the following output:

```
              gen         s1        s2        s3        s4
alpha  0.9638781 0.92504205 0.8992820 0.9052459 0.9405882
omega  0.9741033 0.56377307 0.7884791 0.6766430 0.7816839
omega2 0.9094893 0.09237594 0.3666293 0.1880759 0.2054075
omega3 0.9077636 0.09240479 0.3666634 0.1878380 0.2053012
avevar        NA         NA        NA        NA        NA
```

which is reproduced from the section on omega-hierarchical in the main article. Here, we highlight that in addition to the ω_h estimates listed under the `gen` column, there are also omega estimates corresponding to the four specific factors, `s1` through `s4`. Specifically, the values on the `omega2` and `omega3` rows under the specific factor columns correspond to $\omega_{h\text{-}ss}$ estimates (as before, the `omega` row gives estimates which are not reliability estimates for the subscale or total scores). Notice that these $\omega_{h\text{-}ss}$ values are much lower than those given earlier for the reliability due to the lower-order factors of the higher order model: Whereas we obtained ω = .93 for a subscale based on the five *task-effort* items, we now have $\omega_{h\text{-}ss}$ = .03 for the `s1` factor. Taken together, these values suggest that most of the reliable variance of the task-effort subscale is attributable to the general psychological cost factor instead of a construct representing aspects of task effort that are independent from a general psychological cost construct.

Exploratory Omega Estimates in R

The `omega` function from the `psych` package (i.e., `psych::omega`; version 1.8.12; Revelle, 2018) can be used to obtain omega estimates based on an EFA approach. Adapting an EFA approach to understanding the dimensional structure of a scale is advisable in the earlier stages of scale development when no hypothesized multidimensional CFA model can adequately explain the item-level data.

Specifically, `psych::omega` estimates ω_h by first estimating a higher-order exploratory factor structure. In an exploratory higher-order model, every item has a non-zero factor loading on *each* lower-order factor, whereas in a higher-order CFA model, each item has a non-zero factor loading on only one lower-order factor while its loadings on the other lower-order factors are fixed to zero a priori based on strong hypotheses about which items are directly influenced by which factors. In EFA, *factor rotation* is then used to aid interpretation by making some loadings large and others small. After `psych::omega` estimates this exploratory higher-order model, it is translated into an exploratory bifactor model using the Schmid-Leiman (1957) transformation. In contrast with the CFA bifactor model presented above, this transformation produces a *restricted* bifactor model that is equivalent to the higher-order model (Yung et al., 1999) and may not produce an "appropriate" bifactor model as described by Jennrich and Bentler (2011). For this reason, the `Omega Hierarchical` result output by `psych::omega` may be better understood as an EFA-based estimate of ω_{ho} rather than ω_h.

To demonstrate with the PCS data, the reliability of a total score for the measurement of a general *cost* construct can be estimated with

```
> library(psych)
> omega(pcs, nfactors = 4)
```

where `pcs` is the data matrix (or data frame) with the 19 item-response variables (and no other variables[2]) and `nfactors = 4` indicates the number of specific factors (or lower-order factors) for the EFA model; this model will have five factors including the general factor, but this `nfactors` option only counts the specific factors (although a unidimensional model can be obtained with `nfactors = 1`).[3] It is also possible to include the option `poly=TRUE` to fit the model to polychoric correlations; here, the polychoric option is not invoked because the PCS

---

[2] If any other variable (such as a participant ID variable, a covariate such as gender, or items from another scale) is included in the data set for the `omega` function, that variable or variables will be incorrectly included in the factor analysis.

[3] Here, the number of specific factors was specified based on expectations from prior research (i.e., Flake et al., 2015). But often, a key aspect of EFA is determining the optimal number of factors when there is very little guidance from previous research; a range of methods is available for this purpose (see Fabrigar & Wegener, 2012). If the number of specific factors is not specified, `psych::omega` will estimate a model with three specific factors by default (because a higher-order model is not identified when there are only two lower-order factors).

item responses have a six-point scale. The output from this function call is somewhat voluminous (see Rodriguez et al., 2016, for more explanation), and so below we paste only the critical result:

```
Omega Hierarchical:    0.85
```

indicating that ω_h = .85 for the PCS as calculated using the parameter estimates of this EFA model. This result differs from ω_{ho} = .91 obtained using `semTools::reliabilityL2`; the difference is partly attributable to the distinction between the exploratory higher-order model and the CFA model.

The `psych::omega` function can accommodate the ordinal nature of common item-response formats by calculating omega estimates from an EFA model fitted to polychoric correlations. However, unlike `semTools::reliability`, the omega estimates produced by `psych::omega` lead to reliability estimates for a latent total score X^* instead of the observed total score X, as described in the discussion of ordinal alpha in the main text of the article. Consequently, estimates produced by `psych::omega` for categorical items may be biased relative to reliability estimates which are in the observed total score scale as per Green and Yang (2009b).

References cited above but not in the main article:

DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing, 13*(4), 354-378.

Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.