**Supplemental Materials**

**Method Biases in Single-Source Personality Assessments**

**by R. R. McCrae, 2018, *Psychological Assessment***

**http://dx.doi.org/10.1037/pas0000566**

**Appendix: Modeling Correlations**

*Simulating Scales*

 Psychometric models of the properties of a scale typically distinguish different components that contribute to the observed score. The most common model assumes that the observed score is composed of true score (what the scale is intended to assess) plus random error of measurement. These components are independent (uncorrelated), and are usually assumed to be normally distributed. Computationally, one can create a simulated scale by summing two standard normal random variables, one representing the true score and one representing error. These two components may be combined in any proportion; the higher the ratio of true score to error, the better the scale.

 Although the weights used are arbitrary, there are computational advantages to choosing weights such that the simulated observed score is itself a standard normal variable. That can be accomplished by dividing the initial weights by the square root of the sum of their squares. For example, if one wishes to simulate a scale in which there is twice as much true score (weight = 2) as error (weight = 1), the true score random variable should be multiplied by

2 / SQRT(22 + 12) = 2 / 2.236 = 0.894

and the error variable should be multiplied by

1 / SQRT(22 + 12) = 1 / 2.236 = 0.447.

I will refer to these rescaled weights as *coefficients*, and mark them with prime symbols. When a standard normal variable is multiplied by a coefficient, the standard deviation of the resulting variable is equal to the coefficient itself (because each normal score, *SD* = 1, has been rescaled by that amount), and the variance of the weighted variable is the square of the coefficient. When two components (say, true score and error) are summed, the variance of their sum is given by the familiar formula:

VAR(A + B) = VAR(A) + VAR(B) +2COV(A, B).

Because true score and error are independent, the covariance term in the above expression is zero, and the variance of the sum is the sum of the squares of the two coefficients, which is 1.0

(e.g., .8942 + .4472 = .8 + .2 = 1.0). The standard deviation of the simulated observed score is therefore 1.0, making it a standard normal variable.

 This method can be applied to any number of independent components. For example, one might wish to model a scale as containing true score, method bias, and error; or as general trait, facet-specific variance, and error. If coefficients are generated by dividing weights by the square root of the sum of the squared weights, the combined score will be a standard normal variable.

*Correlating Scales*

 These simulated variables are of interest because they allow inferences about the components of variance in a scale when data on the correlation between scales is known. Consider two different measures of a construct—say, anxiety. Both are assumed to be composed solely of true score and random error, and the observed correlation between them is .70. What is the proportion of true score in each scale?

 We can model each scale as a combination of anxiety (*A*) and error (ε), such that

Anxiety1 = a'*A* + b'ε1

Anxiety2 = c'*A* + d'ε­2.

The error terms for the two anxiety scales are themselves uncorrelated, so they require distinguishing subscripts. Because a', b', c', and d' are coefficients, a'2 + b'2 = c'2 + d'2 = 1.0, and both scales are thus in the form of standard normal variables. By the definition of a Pearson correlation, *r*ANX, the correlation between the two anxiety scales, is the mean cross-product of standard scores, that is,

{Σ [(a'*A* + b'ε1)\*( c'*A* + d'ε2)]} / *N* =

[a'c' (Σ*A*2) + a'd' (Σ*A*ε2) + b'c' (Σ*A*ε1) + b'd' (Σε1ε2)] / N.

Note that the sum of squares of a standard normal variable (such as *A*) = *N*, and that the sum of the cross-products of two uncorrelated standard normal variables (such as *A* and ε2) = 0. Thus, the equation reduces to

[a'c' (*N*) + a'd' (0) + b'c' (0) + b'd' (0)] / *N* =

[a'c' (*N*)] / *N* = a'c' = *r*ANX

 We are given that *r*ANX = .70, so a'c' must = .70. If this is the only information we have, we cannot determine the actual values of a' or c', although we can state that they will both lie within the range of .7 to 1.0, and thus account for 49% to 100% of the variance in each observed scale. If we also know the retest reliability of Anxiety1 (say, *r*tt = .80), and if we assume the proportion of true score variance is constant across occasions, then we could use the same argument given above to show that a'\*a' = .80, and a' = .894. Then c' = .70/.894 = .783, and the predicted retest reliability of Anxiety2 ­ = c'\*c' = .61. Thus, the model yields a testable hypothesis.

 This method can be extended to more complex models that include method variance, specific variance, and so on. Because all the components of the models are assumed to be independent, all their cross-products contribute nothing to the observed correlation. In words, *the correlation between two variables is equal to the sum of the products of the corresponding coefficients of* shared *components of variance*, recalling that coefficients are rescaled weights. If, for example, both scales are self-reports, they might share method variance as well as trait variance. Multiply the two method coefficients, multiply the two trait coefficients, and sum to get the correlation between the two scales.

 The usefulness of these methods of modeling correlations depends entirely on the accuracy of the assumptions built into the models. Assumptions about normal distributions are probably reasonable in most cases, given the complex sources of variance in psychological variables. But specifying the correct components is crucial, for example, omitting method variance, will likely exaggerate the importance of true score variance. In the analyses in this article, I make a number of simplifying assumptions (e.g., that the coefficients for different components are equal in different observers and different traits); in consequence, the results must be viewed as rough estimates of the quantities of interest.