g-Disattenuation: Using Omega Coefficients to Estimate Effect Sizes for the Underlying General Factors

Marco Del Giudice

University of Trieste

Preprint v. 1 [September 2024]

Marco Del Giudice, Department of Life Sciences, University of Trieste.

Address correspondence to Department of Life Sciences, University of Trieste. Via Weiss 2, Trieste, Italy; email: marco.delgiudice@units.it

Abstract

Researchers are increasingly replacing ot supplementing Cronbach's alpha with model-based coefficients of the omega family, including *omega-total* (the proportion of observed score variance explained by all the common factors of a test) and *omega-general*/*omega-hierarchical* (the proportion of observed score variance solely explained by the general factor). Under the appropriate assumptions, values of omega-general/omega-hierarchical can be used in the disattenuation formula to estimate effect sizes for the underlying general factor(s), net of the contribution of secondary lower-order/group factors, without the need for complex modeling or access to the raw data. In this paper I present the rationale for *g-disattenuation* and discuss the assumptions that have to be met to calculate correlations, standardized mean differences, and multivariate differences at the level of general factors. I illustrate these procedures with a worked-out empirical example from a large personality dataset.

Keywords: disattenuation; effect sizes; latent variables; omega; reliability.

Reliability and the Disattenuation of Effect Sizes

When variables are contaminated by random measurement error, standardized effect sizes such as the correlation coefficient and Cohen's *d* become deflated or *attenuated*. As first noted by Spearman (1904), this bias toward zero can be corrected if one knows the variables' reliability—that is, the proportion of the observed score variance $\mathrm{Var}(X)$ made up by "true score" variance $\mathrm{Var}(T)$ rather than measurement error variance $\mathrm{Var}(E)$:

$$r_{XX} = \frac{\mathrm{Var}(T)}{\mathrm{Var}(X)} = \frac{\mathrm{Var}(T)}{\mathrm{Var}(T)+\mathrm{Var}(E)}\,. \tag{1}$$

Estimates of reliability can be used to *disattenuate* the observed effect sizes. If r_{obs} is the observed correlation between *X* and *Y*, the disattenuated (or "corrected") correlation r_{c} is given by:

$$r_{\mathrm{c}} = \frac{r_{\mathrm{obs}}}{\sqrt{r_{XX}}\sqrt{r_{YY}}}\,. \tag{2}$$

The disattenuated r_{c} estimates the correlation between the true scores of the two variables $r(T_X,\, T_Y)$, which is the correlation that would obtain if *X* and *Y* had been measured without error. For the standardized mean difference between two groups (as in Cohen's *d*), the disattenuation formula is:

$$d_{\mathrm{c}} = \frac{d_{\mathrm{obs}}}{\sqrt{r_{XX}}}\,. \tag{3}$$

The reliability of a test can be estimated in several different ways (Revelle & Condon, 2018). Historically, the most common approach in psychometrics has been to rely on internal consistency indices, and specifically on Cronbach's alpha. The α coefficient provides a lower-bound estimate of reliability, approaching the true reliability as the test gets closer to *tau-equivalence* (i.e., constant true score variance across items). However, tau-equivalence is rarely satisfied in practice, meaning that α tends to underestimate the reliability of tests (see Cortina, 1993; Dunn et al., 2014; McNeish, 2018).

The Family of Omega Coefficients

More recently, researchers have started to replace or supplement α with the model-based coefficients of the *omega* family (McDonald, 1999; McNeish, 2018; Zinbarg et al., 2005). These indices relax the assumption of tau-equivalence, allowing the true score variance to vary across items (Dunn et al., 2014; Flora, 2020; McNeish, 2018). Even more importantly, they explicitly take into account the latent structure of a test, and the presence of additional factors beyond the general factor that usually describes the main construct of interest.[1] Additional factors may arise because the items measure other narrower constructs alongside the broader main construct (e.g., the items of an extroversion scale may also measure specific aspects of extraversion such as assertiveness and enthusiasm), but also as a consequence of methodological artifacts (e.g., positively and negatively worded items may load on different additional factors).

The factorial complexity of a test can be represented with a *hierarchical* model, in which a number of lower-order factors load on a higher-order general factor. Alternatively, one can specify a *bifactor* model consisting of a general factor and a number of orthogonal "group factors" (i.e., factors that only affect a specific group of items; see Reise et al., 2010). While the distinction between hierarchical and bifactor models can be theoretically meaningful, it is peripheral to the purpose of this paper and I will not discuss it in any detail. A hierarchical factor solution can be turned into an equivalent bifactor solution through the Schmid-Leiman transformation (see Wolf & Preising, 2005). The Schmid-Leiman transformation is useful

[1] The dominant approach in psychometrics is to focus on latent constructs, and particularly general factors, as the proper target of measurement. In this paper I adopt the standard perspective; however, it is important to note that there are alternative conceptions of measurement in which latent constructs are strongly de-emphasized, and the goal is to build heterogeneous item composites that perform well in real-world prediction (see Revelle, 2024a).

because yields direct estimates of the items' loadings on both the general factor and the orthogonalized group factors.[2] In this paper, I refer to group factors in bifactor models and lower-order factors in hierarchical models with the generic term *secondary factors*, to distinguish them from the general factor of a test.

The *omega-total* coefficient (ω_{t}) is a reliability index analogous to α; it quantifies the proportion of variance in observed scores explained by all the sources of common variance among items, which comprise both the general and group factors. The ω_{t} coefficient can be calculated from a bifactor solution as:

$$\omega_{\mathrm{t}} = \frac{\left(\sum_{i=1}^{k} \lambda_{\mathrm{g}i}\right)^2 + \sum_{f=1}^{n}\left(\sum_{i=1}^{k} \lambda_{fi}\right)^2}{\mathrm{Var}(X)}, \quad (4)$$

where *k* is the number of items, *n* is the number of group factors, $\lambda_{\mathrm{g}i}$ and λ_{fi} are loadings on the general and group factors (respectively), and V_X is the observed score variance of the test, assuming that *X* is a unit-weighted sum of standardized items (Zinbarg et al., 2005). Coefficient ω_{t} can be used to disattenuate effect sizes in the usual way, to estimate correlations and mean differences between true scores in absence of measurement error:

$$r_{\mathrm{c}} = \frac{r_{\mathrm{obs}}}{\sqrt{\omega_{\mathrm{t}X}}\sqrt{\omega_{\mathrm{t}Y}}} \quad (5)$$

and

$$d_{\mathrm{c}} = \frac{d_{\mathrm{obs}}}{\sqrt{\omega_{\mathrm{t}X}}}. \quad (6)$$

[2] The simulations reported by Bell et al. (2024) illustrate the practical equivalence of hierarchical and bifactor (Schmid-Leiman) estimates of the general factor variance, despite the different meaning of the general factor in the two kinds of models.

In contrast with omega-total, *omega-general* (ω_g) estimates the proportion of observed score variance explained by the general factor alone, leaving out the additional secondary factors:[3]

$$\omega_\mathrm{g} = \frac{\left(\sum_{i=1}^{k} \lambda_{\mathrm{g}i}\right)^2}{\mathrm{Var}(X)} \,. \qquad (7)$$

Zinbarg et al. (2005) proposed *omega-hierarchical* (ω_h) as an alternative label for ω_g to use specifically when factor loadings are estimated with the Schmid-Leiman transformation from an exploratory hierarchical model (as contrasted with a confirmatory bifactor model; see Revelle, 2024b; Revelle & Condon, 2018). Because of the different constraints imposed on item cross-loadings, ω_h tends to yield slightly smaller values than the confirmatory bifactor ω_g. However, the meaning of the coefficients is the same; here, I use ω_g as a generic label that includes ω_h as a special case (for a thorough discussion of these and other estimation methods see Cho, 2022). Note that estimating ω_g with precision requires larger samples compared with ω_t; also, ω_g estimators perform better as the number of secondary factors increases (Cho, 2022). For more details on the performance and limitations of omega indices, see Bell et al. (2024) and Cho (2022).

The ratio between ω_g and ω_t is the proportion of the common (i.e., true score) variance of *X* accounted for by the general factor. Since measurement error vanishes as test length approaches infinity, $\omega_\mathrm{g}/\omega_\mathrm{t}$ provides an estimate of ω_g for a test of infinite length with a structure analogous to that of the observed test. For this reason, it is called *asymptotic* ω_g

[3] Note that, if a test is strictly unidimensional (i.e., a general factor and no additional group factors), ω_t reduces to ω_g. In this case, the formula for ω_g provides an estimate of reliability that is sometimes described as *omega-unidimensional* (ω_u; see e.g., Bell et al., 2024; Flora, 2020).

(*asymptotic* ω_{h} in Revelle, 2024c), *omega-limit* (Trizano-Hermosilla et al., 2021), or ω_{inf} (Revelle, 2024b):

$$\omega_{\mathrm{inf}} = \frac{\omega_{\mathrm{g}}}{\omega_{\mathrm{t}}} = \frac{\left(\sum_{i=1}^{k} \lambda_{\mathrm{g}i}\right)^2}{\left(\sum_{i=1}^{k} \lambda_{\mathrm{g}i}\right)^2 + \sum_{f=1}^{n}\left(\sum_{i=1}^{k} \lambda_{fi}\right)^2} . \quad (8)$$

Helped along by several accessible tutorials (e.g., Flora, 2020; McNeish, 2018; Revelle, 2024b) and easy-to-use statistical functions (for example in the R packages *psych* [Revelle, 2024c] and *fungible* [Waller, 2024]; see also the online app https://chocalc.shinyapps.io/ChoCalc/ [Cho, 2022]), omega coefficients are quickly surging in popularity. It is becoming more common to see ω_{t} and ω_{g} (often in the form of ω_{h}) reported in empirical papers, either by themselves or alongside the more traditional α.

g-Disattenuation

The growing availability of omega coefficients in the literature affords an interesting opportunity that, as far as I know, has not been discussed before. Simply put, ω_{g} can be used to disattenuate correlations and other effect sizes, in place of reliability coefficients such as α or ω_{t}. I call this procedure *g-disattenuation*. Under the appropriate assumptions, g-disattenuation yields an estimate of the corresponding effect size for the underlying general factor(s), net of the contribution of secondary factors. Whenever general factors represent the main constructs of interest underlying the observed variables, the effect sizes of interest are precisely those at the level of those general factors. Disattenuating with ω_{t} (Eqs. 5 and 6) only corrects for the bias due to measurement error, but not for the bias introduced by the presence of additional factors (which can also be regarded as a source of error if one is attempting to measure the construct described by the general factor).

Compared with the alternative option of fitting a full confirmatory model to the data, g-disattenuation is a quick, computationally trivial shortcut to calculate effect sizes at the level of general factors. Most importantly, effect sizes can be g-disattenuated using published summary statistics, even when the raw data are not available; in principle, one may also apply g-disattenuation to meta-analytic correlation matrices obtained from multiple datasets, each containing only a subset of the variables.

Correlation Coefficients

The disattenuated correlation r_{g} estimates the correlation between the general factor components of the two variables, $r(g_X, g_Y)$. This is the correlation that would obtain if *X* and *Y* were pure, error-free measures of the respective general factors:

$$r_{\mathrm{g}} = \frac{r_{\mathrm{obs}}}{\sqrt{\omega_{\mathrm{g}X}}\sqrt{\omega_{\mathrm{g}Y}}} \, . \tag{9}$$

The accuracy of r_{g} depends on coefficients $\omega_{\mathrm{g}X}$ and $\omega_{\mathrm{g}Y}$ being accurate, but also on the pattern of correlations among the group factors that contribute to *X* and *Y*. In particular, r_{g} is an accurate estimate of $r(g_X,\ g_Y)$ if the secondary factors of *X* are uncorrelated with those of *Y*, which is a typical assumption in confirmatory factor analysis when multiple constructs are modeled simultaneously. Even if certain secondary factors are cross-correlated between *X* and *Y*, the effects on r_{g} may partly or fully cancel out if some correlations are positive and others are negative; in other words, what matters is the overall correlation between the secondary factor components of *X* and *Y*, which can be written as $r(F_X,\ F_Y)$. The g-disattenuation formula in Eq. 9 is based on the assumption that $r(F_X,\ F_Y) = 0$.

If F_X and F_Y are correlated, then r_{g} can be inflated, deflated, or potentially reversed depending on the direction and size of both $r(g_X,\ g_Y)$ and $r(F_X,\ F_Y)$. If the correlation between general factor components is positive, a positive correlation between F_X and F_Y will lead to inflated values of r_{g}, whereas a negative correlation may deflate r_{g} or even yield a negative estimate. The opposite applies if the correlation between the general factor components is negative. When the assumption $r(F_X,\ F_Y) = 0$ is violated, the potential impact on r_{g} is stronger to the extent that secondary factors explain a larger portion of the common variance, yielding lower values of ω_{inf}. By contrast, when ω_{inf} is large, most of the variance is accounted for by the general factor, and the potential impact of secondary factors is necessarily limited. However, a large ω_{inf} also means that g-disattenaution is going to yield results very similar to those of regular disattenuation, so that r_{g} will be only slightly larger than r_{c}.

A simple formula (derived in the Appendix) can be used to compute theoretical bounds on the general factor correlation $r(g_X,\ g_Y)$, given the current point estimates of r_{obs}, ω_{t} and ω_{g}:

$$\frac{r_{\mathrm{c}}-\sqrt{1-\omega_{\mathrm{inf}X}}\sqrt{1-\omega_{\mathrm{inf}Y}}}{\sqrt{\omega_{\mathrm{inf}X}}\sqrt{\omega_{\mathrm{inf}Y}}} \leq r(g_X,\ g_Y) \leq \frac{r_{\mathrm{c}}+\sqrt{1-\omega_{\mathrm{inf}X}}\sqrt{1-\omega_{\mathrm{inf}Y}}}{\sqrt{\omega_{\mathrm{inf}X}}\sqrt{\omega_{\mathrm{inf}Y}}}\ . \tag{10}$$

The lower bound corresponds to the limit case in which the secondary factor components of *X* and *Y* are perfectly correlated, so that $r(F_X,\ F_Y) = 1$; the upper bound corresponds to the case of perfectly anti-correlated secondary factor components, with $r(F_X,\ F_Y) = -1$. (Note that, like the disattenuation formulas in Eq. 5 and 9, Eq. 10 may sometimes return values smaller than –1 or larger than 1, to be replaced with ± 1.)

It can be useful to stress that those in Eq. 10 are *not* confidence intervals on r_{g}, but theoretical bounds based on extreme scenarios (correlations of ± 1 between secondary factor components) that are unlikely to occur in reality. If there are grounds to restrict the plausible values of $r(F_X,\ F_Y)$ to a narrower interval $[r_{\mathrm{low}}, r_{\mathrm{high}}]$, the corresponding bounds become:

$$\frac{r_{\mathrm{c}}-r_{\mathrm{high}}\sqrt{1-\omega_{\mathrm{inf}X}}\sqrt{1-\omega_{\mathrm{inf}Y}}}{\sqrt{\omega_{\mathrm{inf}X}}\sqrt{\omega_{\mathrm{inf}Y}}} \leq r(g_X,\ g_Y) \leq \frac{r_{\mathrm{c}}-r_{\mathrm{low}}\sqrt{1-\omega_{\mathrm{inf}X}}\sqrt{1-\omega_{\mathrm{inf}Y}}}{\sqrt{\omega_{\mathrm{inf}X}}\sqrt{\omega_{\mathrm{inf}Y}}}\,. \tag{11}$$

As an illustration, consider two variables *X* and *Y* with an observed correlation $r_{\mathrm{obs}} = .30$. The coefficients for *X* are $\omega_{\mathrm{t}X} = .85$, $\omega_{\mathrm{g}X} = .55$, and $\omega_{\mathrm{inf}X} = \frac{.55}{.85} = .65$. Variable *Y* is less reliable, but with a stronger contribution of the general factor to the common variance: $\omega_{\mathrm{t}Y} = .70$, $\omega_{\mathrm{g}Y} = .60$, and $\omega_{\mathrm{inf}Y} = \frac{.60}{.70} = .85$. The disattenuated correlation based on reliabilities is $r_{\mathrm{c}} = .39$; this is the estimated correlation between the true scores of *X* and *Y* in absence of measurement error. The estimated correlation between the underlying general factors is given by the g-disattenuated correlation, $r_{\mathrm{g}} = .52$. As discussed above, this estimate is based on the assumption of uncorrelated group factor components between *X* and *Y*. If the assumption is justified, the estimate can be taken as valid. If the assumption is violated, the correlation between general factors must lie between .22 and .82 (from Eq. 10), conditional on the current point estimates of the observed correlation and omega coefficients. However, these bounds are exceedingly wide and not very useful in practice. If values of the correlation between secondary factor components could be plausibly restricted to a narrower range of (say) $\pm.20$, then the correlation between general factors would be bounded between .46 and .58 (from Eq. 11).

To conclude this section, a brief note on standard errors. Disattenuation is not a free lunch: the increased accuracy of the estimated effect size is paid for with a corresponding decrease in the precision of that estimate. Specifically, the standard error (*SE*) of the effect size increases by about the same amount as the effect size itself (see Schmidt & Hunter, 2014; Wiernik & Dahlke, 2020). Naturally, g-disattenuation is no exception to this rule. To a good approximation:

$$SE_{\mathrm{g}} = SE_{\mathrm{obs}}\frac{1}{\sqrt{\omega_{\mathrm{g}X}}\sqrt{\omega_{\mathrm{g}Y}}} = SE_{\mathrm{obs}}\frac{r_{\mathrm{g}}}{r_{\mathrm{obs}}}\,. \tag{12}$$

Note that the standard adjustment formula in Eq. 12 simply takes the coefficients at the denominator as given; hence, it does not take into account the sampling error associated with ω_{gX} and ω_{gY}. If one wishes to include all the relevant uncertainty and the raw data are available, one option is to bootstrap the entire procedure, from the initial estimation of r_{obs}, ω_{gX}, and ω_{gY} to the calculation of r_g.

Standardized Mean Differences

For standardized mean differences such as Cohen's *d*, the g-disattenuation formula is:

$$d_g = \frac{d_{obs}}{\sqrt{\omega_{gX}}}, \quad (13)$$

where d_g is the estimated difference between two groups on the general factor that underlies variable *X*. For d_g to be an accurate estimate of the underlying effect size, observed scores must satisfy measurement invariance with respect to the general factor—that is, the same latent score on the general factor must correspond to the same expected observed score, regardless of group membership (see e.g., Meredith, 1993; Putnick & Bornstein, 2016; Rudnev et al., 2018). This implies either that all the secondary factors have equal means across groups, or that the effects of different secondary factors cancel out. In both cases, the result is that the secondary factor component F_X has equal means between groups, and the observed difference is entirely due to g_X. Note that measurement invariance is assumed, at least implicitly, whenever one calculates an effect size such as d_{obs} and interprets it as a group difference on a unitary construct (in this case, the general factor).[4] Hence, d_g can be regarded as accurate to the extent that ω_{gX} is accurate and the observed d_{obs} is a valid effect size with respect to the construct measured by the general factor.

[4] It is worth stressing again that, while this is the dominant conception of psychometric constructs, it is not the only game in town (Revelle, 2024a).

As with correlations, violations of the assumption that F_X does not differ between groups have a stronger impact when ω_{inf} is small, i.e., when secondary factors explain a larger portion of the common variance. When ω_{inf} is large, most of the common variance is explained by the general factor, and violations at the level of secondary factors tend to have a smaller impact on d_{g}.

To a good approximation, the standard error of the g-disattenuated d_{g} is given by:

$$SE_{\text{g}} = SE_{\text{obs}} \frac{1}{\sqrt{\omega_{\text{g}X}}} = SE_{\text{obs}} \frac{d_{\text{g}}}{d_{\text{obs}}} . \quad (14)$$

The same observations made with respect to the standard error of r_{g} apply here as well.

Multivariate Differences

Like regular disattenuation, g-disattenuation may be applied to other effect sizes besides *r* and *d*. In particular, it can be used to calculate the general-factor version of multivariate effect sizes such as Mahalanobis' *D*. The *D* index is the multivariate generalization of Cohen's *d*, and compares the average profiles of two groups on a set of correlated variables (see Del Giudice, 2009, 2022, 2023; Olejnik & Algina, 2000).[5]

To calculate the g-disattenuated index D_{g}, one must g-disattenaute both the vector of standardized mean differences and the (pooled) correlation matrix. This requires the variables involved to satisfy the assumptions for both kinds of effect sizes—namely, the secondary factor components of all the variables should have equal means across groups *and* be uncorrelated between pairs of variables. This invites some prudence, because assumption violations across

[5] The formula is $D = \sqrt{\mathbf{d}^{\text{T}}\mathbf{R}^{-1}\mathbf{d}}$, where **d** is a column vector of standardized mean differences between groups and **R** is the pooled correlation matrix. A simple R function to calculate and disattenuate *D* is available at https://doi.org/10.6084/m9.figshare.7934942.

multiple variables—as well as inaccuracies in the estimation of ω_{g} coefficients—may end up having a cumulative effect, potentially causing substantial distortions in the overall effect size.

Another issue is that g-disattenuation increases the magnitude of correlations more than regular disattenuation does; for this reason, there is a higher chance that the resulting correlation matrix $\mathbf{R_g}$ will be non-positive definite (NPD), thus distorting or precluding the calculation of D_{g}. There are a variety of numerical methods to smooth an NPD correlation matrix into a neighboring, positive definite matrix (e.g., Bentler & Yuan, 2011; Higham, 2002). In this particular task, however, standard smoothing methods may yield poor results. If the original g-disattenuated matrix contains artifactual patterns (due to violations of assumptions and/or inaccurate omega coefficients), a smoothed matrix is likely to closely reflect the same patterns, resulting in distorted or grossly unrealistic values of D_{g}.

A more conservative "averaging" approach can be used when matrix smoothing fails, leveraging the availability of regular disattenuated correlations to perform a simple but effective correction. Specifically, if the disattenuated matrix $\mathbf{R_c}$ is positive definite, one can take a weighted average of $\mathbf{R_g}$ and $\mathbf{R_c}$:

$$\mathbf{R_{avg}} = w\mathbf{R_g} + (1 - w)\mathbf{R_c} \,, \tag{15}$$

choosing the largest value of w between 0 and 1 that yields a positive definite matrix $\mathbf{R_{avg}}$. (The largest suitable w is easy to find iteratively, by trying out decreasing values in small steps starting from 1 and testing the resulting matrices for positive definiteness.) This amounts to correcting the correlations for *as much as possible* of the variance explained by secondary factors, on top of measurement error. The averaged matrix $\mathbf{R_{avg}}$ can then be used in place of $\mathbf{R_g}$ to calculate D_{g}. Naturally, the effect size calculated in this way can only be treated as a ballpark approximation. The bottom line is that estimating multivariate effect sizes stretches g-disattenuation to its limit, and should be done with caution and careful judgment.

An Empirical Example

I conclude with a real-world example that illustrates both the usefulness and the challenges of g-disattenuation. The data come from the U.S. subsample of a large online study of personality carried out by the *Open Source Psychometrics project* (https://openpsychometrics.org). In the study, the 15 primary personality factors of Cattell's 16PF model were measured using public domain items (10 items per scale, for a total of 150 items). After cleanup (see Kaiser et al., 2020), the U.S. subsample comprises 7,963 males and 13,604 females (16-90 years). The R code of the present analysis is available at https://doi.org/10.6084/m9.figshare.27015079.

In a previous analysis of these data, my colleagues and I found a multivariate difference between males and females of D = 1.18 with observed scores, which increased to 1.68 after disattenuation with Cronbach's α (Kaiser et al., 2020). We then fitted a multigroup confirmatory factor analysis model to the data, specifying a single general factor for each scale and merging the corresponding items into three parcels with the Single Factor method (Landis et al., 2000). This method is especially suitable for unidimensional tests with a strong general factor, as it tends to distribute any existing secondary factors across parcels; if a test has a substantially multidimensional structure, the (weak) general factor becomes partly confounded with the secondary factors. As a result, the estimated correlations and/or group differences at the level of general factors are going to be deflated, to the extent that secondary factors are uncorrelated between tests and/or have equal means between groups (i.e., to the extent that the assumptions of g-disattenuation are met; see Hall et al., 1999; Little et al., 2002, 2013). Based on latent correlations and univariate differences from this model, we estimated a multivariate sex difference of D = 2.06 at the level of general factors (see Kaiser et al., 2020).

To subject the same dataset to g-disattenuation, I calculated ω_{t}, ω_{h} (omega-hierarchical; see above) and ω_{inf} coefficients for each of the 15 scales, using function *omega()* in the *psych* package (v. 2.4.6.26; Revelle, 2024c). The number of lower-order factors was set at the default of n = 3, with three exceptions. For Dutifulness and Imagination, four factors yielded a cleaner solution, as judged from bifactor loading plots; in both cases, coefficients were very similar in the two solutions (for Dutifulness, $\omega_{\mathrm{t}} = .89$ and $\omega_{\mathrm{h}} = .78$ with three factors, $\omega_{\mathrm{t}} = .90$ and $\omega_{\mathrm{h}} = .73$ with four factors; for Imagination, $\omega_{\mathrm{t}} = .85$ and $\omega_{\mathrm{h}} = .57$ with three factors, $\omega_{\mathrm{t}} = .86$ and $\omega_{\mathrm{h}} = .61$ with four factors). For Introversion, extracting three or more lower-order factors yielded a Heywood case, which was fixed by switching to two factors with equal loadings (for details see Revelle, 2024c).

Table 1 shows the omega coefficients as well as the observed, disattenuated, and g-disattenuated sex differences for the 15 scales. The scales with the lowest values of ω_{inf}, and hence with the strongest multidimensionality, are Gregariousness and Sensitivity. While Gregariousness shows virtually no sex differences, Sensitivity is the most sexually differentiated scale in the 16PF (Del Giudice et al., 2012). In combination with the small proportion of variance explained by the general factor ($\omega_{\mathrm{h}} = .44$), this yields a particularly large g-disattenuated sex difference on this scale ($d_{\mathrm{g}} = -1.22$; for comparison, the latent estimate in Kaiser et al. [2020] was – 0.93).

Table 1. Omega coefficients, univariate sex differences (Cohen's *d*), and multivariate sex differences (Mahalanobis' *D*) for the 15 personality scales.

	ω_t	ω_h	ω_{inf}	d_{obs}	d_c	d_g
A. Warmth	.87	.70	.81	– 0.35	– 0.38	– 0.42
C. Emot. Stability	.89	.74	.83	0.31	0.32	0.36
E. Assertiveness	.87	.72	.83	0.19	0.21	0.23
F. Gregariousness	.85	.47	.55	0.01	0.01	0.01
G. Dutifulness	.90	.73	.82	0.32	0.34	0.38
H. Friendliness	.92	.82	.89	0.00	0.00	0.00
I. Sensitivity	.75	.44	.59	– 0.80	– 0.93	– 1.22
L. Distrust	.89	.74	.83	0.05	0.05	0.06
M. Imagination	.86	.61	.70	0.11	0.12	0.14
N. Reserve	.90	.81	.90	0.15	0.16	0.17
O. Anxiety	.87	.75	.86	– 0.51	– 0.55	– 0.59
Q1. Complexity	.82	.75	.75	– 0.20	– 0.22	– 0.26
Q2. Introversion	.87	.59	.68	– 0.01	– 0.01	– 0.01
Q3. Orderliness	.86	.53	.62	0.09	0.09	0.12
Q4. Emotionality	.85	.60	.70	– 0.09	– 0.10	– 0.11
				D_{obs}	D_c	D_g
Multivariate				1.18	1.49	2.80

Note. d_{obs}, D_{obs} = observed mean differences; d_c, D_c = disattenuated mean differences; d_g, D_g = g-disattenuated mean differences. Positive values of *d* indicate that males score higher than females; the multivariate *D* is unsigned. The letters associated with each scale are the conventional trait identifiers in the 16PF model.

The observed (r_{obs}) and disattenuated correlations (r_c) are shown in Table 2, while Table 3 shows the g-disattenuated correlations (r_g, below the diagonal). Note that observed correlations are pooled from the male and female subsamples, to avoid double-counting the effect of sex differences in means; disattenuated and g-disattenuated correlations are then calculated from the pooled observed correlations.

Table 2. Observed and disattenuated correlations among the 15 personality scales.

	A.	C.	E.	F.	G.	H.	I.	L.	M.	N.	O.	Q1.	Q2.	Q3.	Q4.
A. Warmth		.27	.19	.46	–.26	.56	.33	–.50	–.05	–.55	–.04	–.33	–.48	–.09	–.47
C. Emot. Stability	.24		.42	.23	–.28	.49	–.03	–.49	–.26	–.31	–.85	–.14	–.26	–.21	–.62
E. Assertiveness	.16	.36		.36	.08	.58	.03	–.01	.00	–.35	–.45	–.30	–.16	–.28	.02
F. Gregariousness	.40	.20	.31		.13	.72	.00	–.22	.18	–.47	–.17	–.20	–.60	.13	–.12
G. Dutifulness	–.23	–.25	.07	.11		–.11	.05	.38	.59	.14	.06	–.30	.20	.41	.30
H. Friendliness	.51	.44	.52	.64	–.10		.06	–.40	–.13	–.70	–.40	–.22	–.63	–.12	–.29
I. Sensitivity	.27	–.03	.03	.00	.04	.05		–.16	.29	–.18	.06	–.63	.12	.02	–.20
L. Distrust	–.44	–.44	–.01	–.20	.34	–.37	–.13		.28	.48	.33	.09	.43	.04	.62
M. Imagination	–.04	–.23	.00	.16	.52	–.12	.23	.24		.12	.18	–.53	.26	.40	.14
N. Reserve	–.49	–.27	–.31	–.41	.13	–.64	–.15	.43	.11		.17	.18	.55	.06	.22
O. Anxiety	–.04	–.75	–.39	–.15	.06	–.36	.05	.29	.15	.16		.16	.13	.09	.54
Q1. Complexity	–.28	–.12	–.25	–.17	–.26	–.19	–.49	.08	–.45	.15	.14		–.06	–.09	.29
Q2. Introversion	–.42	–.23	–.14	–.52	.18	–.56	.10	.38	.22	.49	.12	–.05		–.01	.24
Q3. Orderliness	–.08	–.18	–.24	.11	.36	–.10	.02	.04	.35	.05	.07	–.08	–.01		.05
Q4. Emotionality	–.41	–.54	.01	–.10	.27	–.26	–.16	.54	.12	.19	.46	.24	.21	.04	

Note. Observed correlations (r_{obs}, pooled from the male and female subsamples) below the diagonal; disattenuated correlations (r_{c}) above the diagonal.

Interestingly, some of the g-disattenuated correlations approach or even reach unity. Taken at face value, this would imply that Gregariousness, Friendliness, and (reverse) Introversion actually share the same general factor, and that scores on these scales are only differentiated by the contribution of secondary factors (plus measurement error). Likewise, at the level of general factors, Anxiety appears to be just the reverse of Emotional Stability, whereas Complexity is essentially the reverse of Sensitivity. Since the 15 scales are meant to measure distinct underlying factors, these correlations highlight a potential problem, either with the specific questionnaire used in this study or with the 16PF model more generally.

Of course, these interpretations of g-disattenuated coefficients rest on the assumption that secondary factors are uncorrelated among scales, which—considering the sheer number of scales and modeled secondary factors—is unlikely to be strictly true. Indeed, a comparison between standardized mean differences (Table 1) and correlations (Table 3) provides indirect evidence of assumption violations for some scales. In particular, if the general factors underlying Complexity and (reverse) Sensitivity were essentially the *same* factor (as indicated by $r_{\mathrm{g}} = -.95$), the g-disattenuated sex differences on these two scales should have about the same size, but with opposite signs. Instead, the effects are substantially different in size and go in the same direction (higher means in females), with $d_{\mathrm{g}} = -0.26$ for Complexity and -1.22 for Sensitivity. This means that some secondary factors of Complexity and Sensitivity must be correlated with one another in a way that inflates the estimated r_{g}, and/or have different means between the sexes in a way that violates measurement invariance.

Table 3. g-disattenuated and averaged correlations among the 15 personality scales.

	A.	C.	E.	F.	G.	H.	I.	L.	M.	N.	O.	Q1.	Q2.	Q3.	Q4.
A. Warmth		.29	.20	.55	–.28	.60	.39	–.54	–.05	–.59	–.04	–.37	–.54	–.11	–.53
C. Emot. Stability	.33		.45	.28	–.31	.52	–.04	–.53	–.30	–.33	**–.90**	–.15	–.29	–.24	–.69
E. Assertiveness	.23	.50		.42	.09	.61	.04	–.01	.00	–.37	–.48	–.33	–.18	–.32	.02
F. Gregariousness	.69	.35	.53		.15	.83	–.01	–.26	.23	–.55	–.20	–.24	–.74	.17	–.15
G. Dutifulness	–.32	–.35	.10	.19		–.12	.06	.41	.66	.15	.07	–.33	.23	.48	.34
H. Friendliness	.67	.57	.67	**1.00**	–.13		.07	–.43	–.15	–.73	–.42	–.23	–.70	–.13	–.32
I. Sensitivity	.48	–.05	.05	–.01	.08	.09		–.19	.35	–.21	.07	–.75	.15	.03	–.25
L. Distrust	–.61	–.59	–.01	–.33	.46	–.47	–.23		.31	.51	.36	.10	.49	.05	.70
M. Imagination	–.06	–.35	.00	.30	.78	–.17	.46	.36		.13	.19	–.61	.30	.48	.17
N. Reserve	–.65	–.36	–.40	–.67	.17	–.78	–.25	.55	.15		.18	.19	.61	.06	.24
O. Anxiety	–.05	**–1.00**	–.53	–.25	.08	–.46	.09	.39	.22	.20		.18	.15	.10	.60
Q1. Complexity	–.43	–.17	–.38	–.31	–.38	–.26	**–.95**	.11	–.73	.21	.20		–.07	–.11	.33
Q2. Introversion	–.65	–.35	–.21	**–.98**	.27	–.81	.20	.58	.37	.71	.17	–.09		–.02	.29
Q3. Orderliness	–.13	–.29	–.39	.23	.58	–.16	.04	.06	.61	.08	.12	–.14	–.02		.06
Q4. Emotionality	–.63	–.81	.02	–.20	.40	–.37	–.32	.81	.20	.28	.69	.40	.35	.07	

Note. g-disattenuated correlations (r_{g}) below the diagonal; averaged correlations (r_{avg}) above the diagonal. Values of r_{avg} are weighted averages between disattenuated and g-disattenuated correlations (see the main text for details). Correlations of $\pm.90$ and beyond are shown in boldface.

With these caveats in mind, one can try to estimate multivariate sex differences at the level of the general factors underlying the personality scales (bottom line of Table 1). The large sample size relative to the number of variables obviates the need for small-sample corrections (see Del Giudice, 2022). The effect size is $D_{\mathrm{obs}} = 1.18$ from observed scores, and $D_{\mathrm{c}} = 1.49$ after correcting for unreliability with ω_{t}.

The g-disattenuated correlation matrix $\mathbf{R_g}$ (correlations below the diagonal in Table 3) turns out to be NPD, so the g-disattenuated effect size D_{g} cannot be calculated directly. Smoothing the $\mathbf{R_g}$ matrix with Higham's (2002) method (implemented in package *Matrix* v. 1.7-0; Bates et al., 2024) yields a preposterous estimate of $D_{\mathrm{g}} = 2{,}939.18$, whereas Bentler and Yuan's (2011) method (implemented in package *fungible* v. 2.4.4; Waller, 2024) returns a more realistic, but still suspiciously large $D_{\mathrm{g}} = 5.01$. Following the alternative correction procedure described earlier, I averaged $\mathbf{R_g}$ and $\mathbf{R_c}$ with $w = .38$, the smallest value that resulted in a positive definite matrix $\mathbf{R_{avg}}$ (correlations above the diagonal in Table 3; note the absence of correlations stronger than $\pm .90$). The averaged matrix yields a multivariate effect size of $D_{\mathrm{g}} = 2.80$, noticeably larger than the latent estimate of 2.06 in Kaiser et al. (2020), but (intriguingly) close to the latent estimate of 2.71 based on the validation sample of the original 16PF questionnaire (Del Giudice et al., 2012). This effect size must be regarded as nothing more than a tentative estimate—both because the g-disattenuated correlation matrix had to be corrected, and because there are indications that the assumptions of g-disattenuation may not hold, at least for some of the scales.

Conclusion

Omega coefficients (ω_t and ω_g/ω_h) are becoming common in the psychometric literature; in this paper, I noted that they can be used to estimate effect sizes at the level of the underlying general factors, without the need for complex modeling or even access to the raw data. This is in line with other extensions of the original disattenuation procedure that seek to account and correct for multiple sources of error (e.g., Le et al., 2009; Vispoel et al., 2018). While g-disattenuation is a practical and cost-effective shortcut, its accuracy depends on certain assumptions about the behavior of secondary factors. In a deeper sense, g-disattenuation is meaningful to the extent that general factors (to the exclusion of any lower-order/group factors) are, in fact, the constructs of interest with respect to the question at hand. If these conditions are met, omega coefficients offer a convenient way to estimate what the effect sizes would be if the tests were pure, error-free measures of their respective general factors. After 120 years, Spearman's idea has yet to exhaust its potential, and continues to inspire new ways to extract useful information from psychometric data.

References

Bates, D., Maechler, M., & Jagan, M. (2024). *Matrix* v. 1.7-0. https://cran.r-project.org/package=Matrix

Bell, S. M., Chalmers, R. P., & Flora, D. B. (2024). The impact of measurement model misspecification on coefficient omega estimates of composite reliability. *Educational and Psychological Measurement, 84,* 5-39. https://doi.org/10.1177/00131644231155804

Bentler, P. M. & Yuan, K. H. (2011). Positive definiteness via off-diagonal scaling of a symmetric indefinite matrix. *Psychometrika, 76,* 119–123

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

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Del Giudice, M. (2009). On the real magnitude of psychological sex differences. *Evolutionary Psychology, 7,* 264-279. https://doi.org/10.1177/147470490900700209

Del Giudice, M. (2022). Measuring sex differences and similarities. In D. P. VanderLaan & W. I. Wong (Eds.), *Gender and sexuality development: Contemporary theory and research.* Springer. https://doi.org/10.1007/978-3-030-84273-4_1

Del Giudice, M. (2023). Individual and group differences in multivariate domains: What happens when the number of traits increases? *Personality and Individual Differences, 213,* 112282. https://doi.org/10.1016/j.paid.2023.112282

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105,* 399-412. https://doi.org/10.1111/bjop.12046

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,* 484-501. https://doi.org/10.1177/2515245920951747

Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods, 2,* 233-256. https://doi.org/10.1177/109442819923002

Higham, N. J. (2002). Computing the nearest correlation matrix: A problem from finance. *IMA Journal of Numerical Analysis, 22,* 329–343. https://doi.org/10.1093/imanum/22.3.329

Kaiser, T., Del Giudice, M., & Booth, T. (2020). Global sex differences in personality: Replication with an open online dataset. *Journal of Personality, 88,* 415-429. https://doi.org/10.1111/jopy.12500

Landis, R. S., Beal, D. J., & Tesluk, P. E. (2000). A comparison of approaches to forming composite measures in structural equation models. *Organizational Behavioral Research, 3,* 186–207. https://doi.org/10.1177/109442810032003

Le, H., Schmidt, F. L., & Putka, D. J. (2009). The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organizational Research Methods, 12,* 165-200. https://doi.org/10.1177/1094428107302900

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9,* 151-173. https://doi.org/10.1207/S15328007SEM0902_1

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods, 18,* 285-300. https://doi.org/10.1037/a0033266

McDonald, R. P. (1999). *Test theory: A unified treatment.* Erlbaum Associates.

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

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525-543. https://doi.org/10.1007/BF02294825

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25,* 241-286. https://doi.org/10.1006/ceps.2000.1040

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41,* 71-90. https://doi.org/10.1016/j.dr.2016.06.004

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

Revelle, W. (2024a): The seductive beauty of latent variable models: Or why I don't believe in the Easter Bunny. *Personality and Individual Differences, 221,* 112552. https://doi.org/10.1016/j.paid.2024.112552

Revelle, W. (2024b). *Using R and the* psych *package to find* ω. Retrieved on September 2, 2024 from: http://personality-project.org/r/psych/HowTo/omega.pdf

Revelle, W. (2024c). *psych* v. 2.4.6.26. https://cran.r-project.org/package=psych

Revelle, W., & Condon, D. M. (2018). Reliability. In P. Irwing, T. Booth, & D. J. Hughes (Eds.). *The Wiley handbook of psychometric testing* (pp. 709–749). Wiley.

Rudnev, M., Lytkina, E., Davidov, E., Schmidt, P., & Zick, A. (2018). Testing Measurement Invariance for a Second- Order Factor: A Cross-National Test of the Alienation Scale. *Methods, Data, Analyses, 12,* 47-76. https://doi.org/10.12758/mda.2017.11

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Sage.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15,* 72–101. https://doi.org/10.2307/1412159

Trizano-Hermosilla, I., Gálvez-Nieto, J. L., Alvarado, J. M., Saiz, J. L., & Salvo-Garrido, S. (2021). Reliability estimation in multidimensional scales: Comparing the bias of six estimators in measures with a bifactor structure. *Frontiers in Psychology, 12,* 508287. https://doi.org/10.3389/fpsyg.2021.508287

Vispoel, W. P., Morris, C. A., & Kilinc, M. (2018). Using generalizability theory to disattenuate correlation coefficients for multiple sources of measurement error. *Multivariate Behavioral Research, 53,* 481-501. https://doi.org/10.1080/00273171.2018.1457938

Waller, N. (2024). *fungible* v. 2.4.4. https://cran.r-project.org/package=fungible

Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science, 3,* 94-123. https://doi.org/10.1177/2515245919885611

Wolff, H. G., & Preising, K. (2005). Exploring item and higher order factor structure with the Schmid-Leiman solution: Syntax codes for SPSS and SAS. *Behavior Research Methods, 37,* 48-58. https://doi.org/10.3758/BF03206397

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*, 123-133. https://doi.org/110.1007/s11336-003-0974-7

Appendix

Without loss of generality, let T_X and T_Y be the standardized true scores for variables *X* and *Y*. Each true score can be written as a linear combination of a standardized general factor component g and a standardized secondary factor component F. The proportion of true score variance explained by g corresponds to ω_{inf} (Eq. 8 in the main text), while the proportion of variance explained by F necessarily equals $(1 - \omega_{\text{inf}})$. Hence, the weights for g and F are $\sqrt{\omega_{\text{inf}}}$ and $\sqrt{1 - \omega_{\text{inf}}}$, respectively:

$$\begin{cases} T_X = g_X\sqrt{\omega_{\text{inf}X}} + F_X\sqrt{1 - \omega_{\text{inf}X}} \\ T_Y = g_X\sqrt{\omega_{\text{inf}Y}} + F_X\sqrt{1 - \omega_{\text{inf}Y}} \end{cases} . \quad \text{(A1)}$$

The correlation between the true scores of *X* and *Y* is then equal to:

$$r(T_X, T_Y) = r(g_X, g_Y)\sqrt{\omega_{\text{inf}X}}\sqrt{\omega_{\text{inf}Y}} + r(F_X, F_Y)\sqrt{1 - \omega_{\text{inf}X}}\sqrt{1 - \omega_{\text{inf}Y}} . \quad \text{(A2)}$$

Note that the correlation between true scores is precisely the quantity estimated by the disattenuated correlation coefficient, i.e., $r_{\text{c}} = \hat{r}(T_X, T_Y)$. Substituting r_{c} in Eq. A2 and rearranging yields:

$$r(g_X, g_X) = \frac{r_{\text{c}} - r(F_X, F_Y)\sqrt{1-\omega_{\text{inf}X}}\sqrt{1-\omega_{\text{inf}Y}}}{\sqrt{\omega_{\text{inf}X}}\sqrt{\omega_{\text{inf}Y}}} . \quad \text{(A3)}$$

For $r(F_X, F_Y)$ in the interval $[-1, 1]$, the corresponding bounds on $r(g_X, g_X)$ can be found as:

$$\frac{r_{\text{c}} - \sqrt{1-\omega_{\text{inf}X}}\sqrt{1-\omega_{\text{inf}Y}}}{\sqrt{\omega_{\text{inf}X}}\sqrt{\omega_{\text{inf}Y}}} \leq r(g_X, g_Y) \leq \frac{r_{\text{c}} + \sqrt{1-\omega_{\text{inf}X}}\sqrt{1-\omega_{\text{inf}Y}}}{\sqrt{\omega_{\text{inf}X}}\sqrt{\omega_{\text{inf}Y}}} , \quad \text{(A4)}$$

which is Eq. 10 in the main text. Arbitrary values for the bounds on $r(F_X, F_Y)$ yield Eq. 11 in the main text.