Title
New Methods for Test Reliability based on Structural Equation Modeling

Permalink
https://escholarship.org/uc/item/6v61v1pk

Author
Bentler, Peter M

Publication Date
2015-06-03

Copyright Information
This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at https://creativecommons.org/licenses/by-nc-nd/4.0/
New Methods for Test Reliability based on Structural Equation Modeling


Peter M. Bentler

University of California, Los Angeles
Topics

After a short overview of reliability and structural equation modeling, 2 new reliability methods are presented:

- *Specificity-enhanced* coefficients for improved lower-bound reliability determination

- *Covariate-free* and *covariate-dependent* reliability coefficients for eliminating spurious sources of internal consistency
Reliability

Let \( X \) be an item or a composite score. Test theory posits that \( X \) is the sum of 2 uncorrelated latent variables

\[
X = T + E.
\]

Thus we have additive variances \( \sigma_X^2 = \sigma_T^2 + \sigma_E^2 \) and define

\[
\rho_{XX} = \frac{\sigma_T^2}{\sigma_X^2}.
\]

Such a coefficient holds for an item, or a test/scale, here taken simply as \( X = \sum_i^p X_i \). Today, I concentrate on the reliability of a scale or test, based on the qualities of its items (internal consistency). For simplicity, I assume that errors on different items are uncorrelated.
Factor Analytic Decomposition in a Picture

There are 4 variables A, B, C, D. Each has Common, Specific, and Error Variance, grouped variously:

Factor analysis approach:
Common = True - Specific.
Unique = Specific + Error.

Test theory approach:
True = Common + Specific
Error = Unique - Specific
Equations for FA Variance Decomposition

\[ X = T + E \], but

\[ T = C + S \] (common plus specific, uncorrelated), so

\[ X = C + S + E = C + U \],

with \( \sigma^2_X = \sigma^2_C + \sigma^2_S + \sigma^2_E \). Thus (Bentler, 1968, 2009, 2015)

\[
\rho_{xx} = \frac{\sigma^2_C}{\sigma^2_X} = 1 - \frac{\sigma^2_U}{\sigma^2_X} \leq \frac{\sigma^2_T}{\sigma^2_X} = \rho_{xx} + \frac{\sigma^2_S}{\sigma^2_X} = 1 - \frac{\sigma^2_E}{\sigma^2_X} = \rho_{XX}.
\]

All internal consistency coefficients -- whose history goes back to 1910 (Spearman and Brown) -- are of the form \( \rho_{xx} \). Today, I introduce estimators of \( \sigma^2_S \) to yield specificity-enhanced reliability that will improve these coefficients.
Coefficient Alpha

Let $\Sigma_{xx} = E(x - \mu)(x - \mu)'$ be the population covariance matrix of $X_i$ ($i = 1, \ldots, p$). If $I$ is a unit vector, the variance of the sum $X = I'X_i = \sum_i^p X_i$ is $\sigma_x^2 = I'\Sigma_{xx} I$. Let $\sigma_c^2 \approx p^2 \bar{\sigma}_{ij}$, where $\sigma_{ij}$ is an off-diagonal element of $\Sigma_{xx}$ and $\bar{\sigma}_{ij}$ is the average of all $\sigma_{ij}$. Then

$$\alpha = \frac{p^2 \bar{\sigma}_{ij}}{\sigma_x^2} \leq \rho_{xx}.$$  

In practice, the sample covariance matrix $S_{xx}$ (not $R_{xx}$) is used. Model-based coefficients get closer to $\sigma_c^2$ and hence $\rho_{xx}$ (e.g., Bentler, 2009; Cho & Kim, 2015).
Model-based Coefficients

Applying $X = C + S + E = C + U$ to a set of items, and assuming zero means, the vector of item scores has decomposition

$$x = c + s + e = c + u,$$

This leads to the covariance structure

$$\Sigma_{xx} = \Sigma = \Sigma_c + \Delta_s + \Delta_e = \Sigma_c + \Psi,$$

where $\Sigma_c$ is the covariance matrix of common scores and $\Psi$ is a (typically diagonal) unique variance matrix. Typically, the $c$ are functions of latent variables - in the factor model $c = \Lambda \xi$ so $\Sigma_c = \Lambda \Phi \Lambda'$ -- but could arise from LISREL, Bentler-Weeks, or other models.
When $\Sigma_c$ is well-structured (e.g., $\Sigma_c = \Lambda \Phi \Lambda'$), improved estimates of $\sigma_c^2 = I' \Sigma_c I$ and hence $\rho_{xx} = \sigma_c^2 / \sigma_x^2$ are possible.

Note that $\rho_{xx}$ (RHO in EQS) is one of many coefficients. If $\Sigma_c = \Lambda \Lambda'$, this is Heise & Bohnstedt’s (1970) $\Omega$ and McDonald’s (1970) $\theta$. If $\Lambda$ is a 1-factor model, this is Jöreskog’s (1971) coefficient (McDonald’s 1999 $\omega$). If $\Sigma_c$ is based on an arbitrary – but fitting -- SEM model (Bentler, 2007), it is a unique coefficient that has no added special name.

Essentially always $\alpha \leq \rho_{xx} \leq \rho_{XX}$. Next, I show how to obtain $\alpha^+$ and $\rho_{xx}^+$ such that $\alpha \leq \alpha^+$ and $\rho_{xx} \leq \rho_{xx}^+$. 


Specificity-enhanced Reliability
The Kaufman Assessment Battery for Children (Kline, 2011, p. 235) has correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
</tr>
</thead>
<tbody>
<tr>
<td>V4</td>
<td>1.000</td>
<td>0.390</td>
<td>0.350</td>
<td>0.210</td>
<td>0.320</td>
</tr>
<tr>
<td>V5</td>
<td>0.390</td>
<td>1.000</td>
<td>0.670</td>
<td>0.110</td>
<td>0.270</td>
</tr>
<tr>
<td>V6</td>
<td>0.350</td>
<td>0.670</td>
<td>1.000</td>
<td>0.160</td>
<td>0.290</td>
</tr>
<tr>
<td>V7</td>
<td>0.210</td>
<td>0.110</td>
<td>0.160</td>
<td>1.000</td>
<td>0.380</td>
</tr>
<tr>
<td>V8</td>
<td>0.320</td>
<td>0.270</td>
<td>0.290</td>
<td>0.380</td>
<td>1.000</td>
</tr>
</tbody>
</table>

A model for 5 visual-spatial reasoning variables V4-V8 is:
It fits the covariances well ($\chi^2_{5(ML)} = 2.3, CFI = 1.0$). The unstandardized factor loadings are

$$[1.000 \ 1.421 \ 1.950 \ 1.144 \ 1.675]'$$

with factor variance $\sigma^2_{F1} = 1.956$ and unique variances

$$[5.334 \ 3.341 \ 10.200 \ 5.280 \ 3.510]$$

We have $\hat{\sigma}^2_u = 27.665$, $\hat{\sigma}^2_x = 128.789$, $\hat{\rho}_{xx} = .785$.

Next, keep this model as is, with fixed parameters. We augment it with V1-V3 that may correlate with the unique scores E4 to E8. If the unique scores are just random residuals, they won’t correlate with V1-V3. If they do correlate, the uniquenesses must contain true scores – that is, specificity. Definite nonzero $r_s$ obtain:
Can the E’s be predicted from the auxiliary Vs? Doing stepwise regression of each Ei on V1-V3 yields:

\[ R^2_{E4,V1} = .061, \quad R^2_{E5,V1,V3} = .302, \quad R^2_{E6,V1,V3} = .300, \]

\[ R^2_{E7,V1,V2} = .292, \quad R^2_{E8,V1,V3} = .562 \]

Next we compute, for each E4-E8, the proportion of unique variance that is actually specificity \( (= R^2 \times \sigma_u^2) \) and error variance \( (= \{1 - R^2\} \times \sigma_u^2) \). Computations give
specific, error, and original unique variances:

<table>
<thead>
<tr>
<th>Vi</th>
<th>$\sigma^2_{si}$</th>
<th>$+ \sigma^2_{ei}$</th>
<th>$= \Psi_{ii}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>V4</td>
<td>0.325</td>
<td>5.009</td>
<td>5.334</td>
</tr>
<tr>
<td>V5</td>
<td>1.009</td>
<td>2.332</td>
<td>3.341</td>
</tr>
<tr>
<td>V6</td>
<td>3.060</td>
<td>7.140</td>
<td>10.2</td>
</tr>
<tr>
<td>V7</td>
<td>1.542</td>
<td>3.738</td>
<td>5.280</td>
</tr>
<tr>
<td>V8</td>
<td>1.973</td>
<td>1.537</td>
<td>3.510</td>
</tr>
<tr>
<td>SUM</td>
<td>7.909</td>
<td>19.756</td>
<td>27.665</td>
</tr>
</tbody>
</table>

Having the new estimate $\hat{\sigma}^2 = 7.909$, RHO$^+$ is

$$\hat{\rho}^+_{xx} = \hat{\rho}_{xx} + \frac{\hat{\sigma}_s^2}{\hat{\sigma}_x^2} = .7852 + \frac{7.909}{128.789} = .847 \text{ or}$$

$$\hat{\rho}^+_{xx} = 1 - \frac{\hat{\sigma}_e^2}{\hat{\sigma}_x^2} = 1 - \frac{19.756}{128.789} = .847.$$
The specificity-corrected $\hat{\rho}_{xx}^+ (= \hat{\omega}^+)$ improves the reliability estimate by almost 8%.

Next, consider a 2\textsuperscript{nd} approach to specificity-corrected reliability: We augment the original model with doublet factors. Each doublet factor is associated with a given item and an auxiliary variable, and its variance is $\hat{\sigma}_s^2$.

This expanded model reproduces exactly the same $\hat{\Sigma}$ as the original one that yields $\hat{\rho}_{xx}$.

We also add constraints so that each factor $\hat{\sigma}_s^2$ plus unique $\hat{\sigma}_s^2$ in the augmented model equals the fixed unique $\sigma^2$ from the original model. We specify:
/EQUATIONS
V1 = *F1 + F4 + F5 + F6 + F7 + F8 + E1;
V2 = *F1 + E2;
V3 = *F1 + E3;
V4 = 1.000F2 + F4 + E4;
V5 = 1.421F2 + F5 + E5;
V6 = 1.950F2 + F6 + E6;
V7 = 1.144F2 + F7 + E7;
V8 = 1.675F2 + F8 + E8;

/VARIANCES
F1 = 1; F2 = 1.956;
F4 TO F8 = *; E1 TO E8 = *

/COVARIANCE
F1,F2 = *;

/CONSTRAINTS
(F4,F4) + (E4,E4) = 5.334;
(F5,F5) + (E5,E5) = 3.341;
(F6,F6) + (E6,E6) = 10.2;
(F7,F7) + (E7,E7) = 5.280;
(F8,F8) + (E8,E8) = 3.510;
Notice that:

- F4, F5, F6, F7, F8 are *common* factors in the space of all variables
- F4 - F8 are *not* common factors in the space of the items V4-V8 making up our scale
- In principle, there are as many possible doublets as the product of # auxiliary vars \( \times \) # items
- Doublets whose variances are not significant should be removed, to avoid capitalizing on chance
- If a doublet variance is constrained at zero, a reparameterization should be considered to allow a possibly negative doublet correlation

The model fits well \( (\chi^2_{24(ML)} = 13.2, CFI = 1.0) \).
Specific, error and original unique variances are:

<table>
<thead>
<tr>
<th>Vi</th>
<th>Fi,Fi</th>
<th>+ Ei,Ei</th>
<th>= Ψ_{ii} (fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V4</td>
<td>.872</td>
<td>4.462</td>
<td>5.334</td>
</tr>
<tr>
<td>V5</td>
<td>1.259</td>
<td>2.082</td>
<td>3.341</td>
</tr>
<tr>
<td>V6</td>
<td>3.111</td>
<td>7.089</td>
<td>10.2</td>
</tr>
<tr>
<td>V7</td>
<td>2.073</td>
<td>3.207</td>
<td>5.280</td>
</tr>
<tr>
<td>V8</td>
<td>1.952</td>
<td>1.558</td>
<td>3.510</td>
</tr>
<tr>
<td>SUM</td>
<td>9.267</td>
<td>18.398</td>
<td>27.665</td>
</tr>
</tbody>
</table>

\( \hat{\rho}_{xx} = 1 - (27.665 / 128.789) = .785 \)

\( \hat{\rho}_{xx}^+ = 1 - (18.398 / 128.789) = .857 \),

about a 9% improvement. The specific \( \hat{\sigma}_{sV4}^2 = \hat{\sigma}_{F4}^2 \) is not

significant – if we set it to zero, we get

\( \hat{\rho}_{xx}^+ = 1 - (19.704 / 128.789) = .847 \) (a .01 reduction)
We may similarly compute $\hat{\alpha}$ and $\hat{\alpha}^+$. The runs are identical to the above (keeping all 5 specific factors), except that to get $\alpha$ from a factor model rather than just the sample covariances:

1. The 1-factor model has all fixed 1.0 loadings
2. METHOD = LS; (least squares estimation).

The model fits so-so ($\chi^2_{9(LS)} = 21.6, CFI = .95$)

$$\hat{\alpha} = 1 - \frac{29.11}{128.854} = .774$$

The enlarged model fits so-so ($\chi^2_{24(LS)} = 56.2, CFI = .93$)

$$\hat{\alpha}^+ = 1 - \frac{19.786}{128.854} = .846.$$ 

These are almost as high as those from the unrestricted 1-factor model.
These approaches also extend to various other coefficients. An important example is the greatest lower bound (glb) (Bentler, 1972; Woodhouse & Jackson, 1977; Bentler & Woodward, 1980). This is based on a factor model with an unspecified # of factors that explains all covariances.

Using the doublet approach as before, we get:

\[ \hat{\rho}_{\text{glb}} = .805 \]

\[ \hat{\rho}_{\text{glb}}^+ = .876 \]

The new glb\(^+\) exceeds the glb by about 9%.
Covariate-free and Covariate-dependent Reliability Coefficients

Is $\rho_{xx}$ invariant to changes in populations? The APA Task Force on Statistical Inference (Wilkinson & APA, 1999): “…a test is not reliable or unreliable. Reliability is a property of the scores on a test for a particular population of examinees.” This implies there may be several, or even dozens, of reliability coefficients [of any fixed definition] for a given scale: for males (females), old (young), low (high) SES, highly (little) educated, etc.

Not a new idea: Generalizability theory has long held that various sources of error may imply different variance ratios.
How serious is this problem, and how can influences on $\rho_{xx}$ be evaluated? In a previous talk (Bentler, 2014), I reviewed several possible approaches to this problem:

1. Reliability generalization. This is a meta-analysis method that seeks correlates and predictors of $\rho_{xx}$ size, such as gender.

2. Multiple group models. Invariance or near invariance of parameters implies (near) invariance of $\rho_{xx}$ across groups.

3. Multilevel models. These provide both Between-group ($\Sigma_B$) and Within-group ($\Sigma_W$) covariance matrices that can be used to obtain $\rho_{xx}$ coefficients. Within-group $\rho_{xx}$ eliminates cluster differences.

I also proposed a new covariate-based methodology.
A Covariate-based Approach to Reliability

As before, we start with

\[ X = T + E \]

and make the usual assumptions to obtain

\[ \rho_{xx} = \frac{\sigma_T^2}{\sigma_X^2}. \]

(For simplicity, I drop the distinction between \( \rho_{xx} \) and \( \rho_{XX} \). Context will clarify.) Now assume there is a set of covariates \( Z \), which may be one or many variables, latent or observed, categorical or continuous, and consider the regression (linear or nonlinear) of \( T \) on \( Z \) such that there exists the orthogonal decomposition
\[ T = \hat{T} + \tilde{T}, \]

with \( \hat{T} = T(Z) \) the covariate-dependent part of \( T \), and \( \tilde{T} = T - T(Z) \) the covariate-free part of \( T \). It follows that \( \sigma^2_T = \sigma^2_{\hat{T}} + \sigma^2_{\tilde{T}} \) and hence

\[
\rho_{xx} = \frac{\sigma^2_T}{\sigma^2_x} = \frac{\sigma^2_{\hat{T}}}{\sigma^2_x} + \frac{\sigma^2_{\tilde{T}}}{\sigma^2_x} = \rho^{(z)}_{xx} + \rho_{xx}^{\perp z}.
\]

\( \rho^{(z)}_{xx} \) is covariate-dependent reliability and \( \rho_{xx}^{\perp z} \) is covariate-free reliability.

In practice, the score decomposition \( T = \hat{T} + \tilde{T} \) is not needed; only the variance decomposition is necessary.
This decomposition can be applied to each of multiple T scores, or to Ts that are based on a factor model, and hence a linear compound of factors F.

If covariate-free reliability $\rho_{xx}^{\perp z}$ is large compared to $\rho_{xx}$, we have high reliability generalization. Reliability then hardly depends on covariates.

If covariate-dependent reliability $\rho_{xx}^{(z)}$ is large compared to $\rho_{xx}$ (alternatively, absolutely large), reliability is highly population-dependent. Separate coefficients would be needed for different populations.
Covariate-free & Covariate-dependent Alpha

Based on $\Sigma_{xx}$, the population covariance matrix among items, we have already encountered

$$\alpha = \frac{p^2 \bar{\sigma}_{ij}}{\sigma_x^2}.$$ 

With covariates, we also have

$$
\begin{pmatrix}
\Sigma_{xx} & \Sigma_{xz} \\
\Sigma_{zx} & \Sigma_{zz}
\end{pmatrix}
$$

The regression of $X_i$ on $Z$ yields the matrix identity

$$
\Sigma_{xx} = (\Sigma_{xx} - \Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{zx}) + (\Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{zx}),
$$

the residual and predictable parts of $X_i$. Hence, their off-diagonal elements obey the equality
$$mean\{\text{offdiag} (\Sigma_{xx})\} = mean\{\text{offdiag} (\Sigma_{xx} - \Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{zx})\}$$

$$+ mean\{\text{offdiag} (\Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{zx})\}$$

and specifically,

$$\bar{\sigma}_{ij} = \bar{\sigma}_{ij}^{\perp z} + \bar{\sigma}_{ij}^{(z)}.$$ 

It follows that alpha can be decomposed into

$$\alpha = \alpha^{\perp z} + \alpha^{(z)},$$

where

$$\alpha^{\perp z} = p^2 \bar{\sigma}_{ij}^{\perp z} / \sigma_x^2$$ is covariate-free alpha and

$$\alpha^{(z)} = p^2 \bar{\sigma}_{ij}^{(z)} / \sigma_x^2$$ is covariate-dependent alpha.
Model-based Coefficients

We also have already seen the decomposition

$$\Sigma_{xx} = \Sigma = \Sigma_{c} + \Psi,$$

based on orthogonal common and unique $p \times 1$ random vectors in deviation form $x = c + u$. Now we would like to partial the $q \times 1$ vector of covariates $z$ out of $c$.

Similarly as before, we may write the partial covariance identity

$$\Sigma_{cc} = (\Sigma_{cc} - \Sigma_{cz} \Sigma_{zz}^{-1} \Sigma_{zc}) + (\Sigma_{cz} \Sigma_{zz}^{-1} \Sigma_{zc}).$$

To make this operational, we assume that $E(uz') = 0$ and we obtain
\[ E(xz') = E(cz') \text{ or } \Sigma_{xz} = \Sigma_{cz}. \]

Now we can substitute \( \Sigma_{xz} \) in the previous formula:

\[ \Sigma_c = \Sigma_{cc} - \Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{zx} + (\Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{zx}) = \Sigma_{c}^{\perp} + \Sigma_{c}^{(z)}. \]

It immediately follows that

\[ \rho_{xx} = \frac{l' \Sigma_c 1}{l' \Sigma l} = \frac{l' \Sigma_{c}^{\perp} 1}{l' \Sigma l} + \frac{l' \Sigma_{c}^{(z)} 1}{l' \Sigma l} = \rho_{xx}^{\perp} + \rho_{xx}^{(z)} \]

where

\( \rho_{xx}^{\perp} \) is covariate-free reliability

\( \rho_{xx}^{(z)} \) is covariate-dependent reliability.
$\Sigma_c$ represents the common score covariance matrix for many models, such as

- **EFA**: $\Sigma_c = \Lambda \Lambda'$
- **CFA**: $\Sigma_c = \Lambda \Phi \Lambda'$
- **FA/SEM**: $\Sigma_c = \Lambda (I - B)^{-1} \Phi (I - B)^{-1'} \Lambda'$

- **blb** (Bentler, 1972): $\min \text{tr}(\Sigma_c)$ psd, $\Psi$ diagonal
- **glb** (Woodhouse & Jackson, 1977; Bentler & Woodward, 1980): $\min \text{tr}(\Sigma_c)$ psd, $\Psi$ diagonal & psd

Also, $\Sigma$ may be a submatrix of a much larger structural model $\Sigma(\theta)$. The rank of $\Sigma_c$ -- the number of factors -- is typically greater than 1. But the 1-factor case is interesting:
Covariate-based 1-Factor Reliability

Let $x = \Lambda_1 \xi + \epsilon$ be the factor model with $\Sigma_c = \Lambda_1 \phi \Lambda'_1$. The factor variance $\phi$ is a scalar (possibly $\phi = 1$). Hence

$$\rho_{xx} = \rho_{11} = \frac{\phi (\Lambda'_1 \Lambda_1)^2}{\sigma_x^2} \quad (= \omega ).$$

Now let the factor $\xi$ be predicted by covariates $z$, with the $R^2$ for predicting $\xi$ being $R^2_{\xi(z)}$. It follows that

$$\varphi = R^2_{\xi(z)} \phi + (1 - R^2_{\xi(z)}) \phi = \varphi^{\xi(z)} + \varphi^\perp_z.$$

With the factor variance partitioned, we may write

$$\rho_{11} = \frac{\varphi^{\xi(z)} (\Lambda'_1 \Lambda_1)^2}{\sigma_x^2} + \frac{\varphi^\perp_z (\Lambda'_1 \Lambda_1)^2}{\sigma_x^2} = \rho^{(z)}_{11} + \rho^\perp_{11}.$$
This partition of reliability can be obtained in two ways:

(1) a simultaneous mimic-type setup such as

where the equation predicting F1 yields $R^2_{\xi}(z)$ and $\phi^{\perp z}$ is the variance of D1;

(2) a 2-step approach, where $\rho_{11}$ is first obtained from only the factor model (no covariates); in step 2, the model is run with loadings and error variances fixed at step-1 values, and other parameters free.
Covariate-based Reliability with LISREL

The LISREL model easily permits a covariate-based partitioning of reliability. Assume we want the reliability of the endogenous $y$ variables, and $x$ variables and its factors are covariates.
The covariance matrix of the $y$ is

$$\Sigma_{yy} = \Lambda_y (I - B)^{-1} (\Gamma \Phi \Gamma' + \Psi)(I - B)'^{-1} \Lambda_y' + \Theta \varepsilon.$$ 

We immediately see that covariate-based reliability is

$$\rho_{yy}^{(x)} = \frac{l' \Lambda_y (I - B)^{-1} (\Gamma \Phi \Gamma')(I - B)'^{-1} \Lambda_y' l}{l' \Sigma_{yy} l}$$

and covariate-free reliability is

$$\rho_{yy}^{\perp x} = \frac{l' \Lambda_y (I - B)^{-1} (\Psi)(I - B)'^{-1} \Lambda_y' l}{l' \Sigma_{yy} l}.$$
Example: Brain Size and IQ
Did you know that “Big-brained people are smarter” (McDaniel, 2005)? He reported:

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Number of studies</th>
<th>Sample size</th>
<th>Observed mean correlation</th>
<th>Mean correlation corrected for range restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>All correlations</td>
<td>37</td>
<td>1530</td>
<td>0.29</td>
<td>0.33</td>
</tr>
<tr>
<td>Analyses by whether the degree of range restriction was interpolated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpolation</td>
<td>21</td>
<td>963</td>
<td>0.29</td>
<td>0.32</td>
</tr>
<tr>
<td>No interpolation</td>
<td>16</td>
<td>567</td>
<td>0.30</td>
<td>0.34</td>
</tr>
<tr>
<td>Analyses by sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>12</td>
<td>438</td>
<td>0.36</td>
<td>0.40</td>
</tr>
<tr>
<td>Males</td>
<td>17</td>
<td>651</td>
<td>0.30</td>
<td>0.34</td>
</tr>
<tr>
<td>Mixed sex</td>
<td>8</td>
<td>441</td>
<td>0.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Analyses by age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adults</td>
<td>24</td>
<td>1120</td>
<td>0.30</td>
<td>0.33</td>
</tr>
<tr>
<td>Children</td>
<td>13</td>
<td>410</td>
<td>0.28</td>
<td>0.33</td>
</tr>
<tr>
<td>Analyses by age and sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female adults</td>
<td>8</td>
<td>327</td>
<td>0.38</td>
<td>0.41</td>
</tr>
<tr>
<td>Female children</td>
<td>4</td>
<td>111</td>
<td>0.30</td>
<td>0.37</td>
</tr>
<tr>
<td>Male adults</td>
<td>11</td>
<td>470</td>
<td>0.34</td>
<td>0.38</td>
</tr>
<tr>
<td>Male children</td>
<td>6</td>
<td>181</td>
<td>0.21</td>
<td>0.22</td>
</tr>
</tbody>
</table>
Are intelligence measures mainly indirect measures of brain size? Posthuma et al. (2003) found:

Table 2

<table>
<thead>
<tr>
<th></th>
<th>GMV</th>
<th>WMV</th>
<th>CBV</th>
<th>VC</th>
<th>WM</th>
<th>PO</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMV</td>
<td>0.59**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBV</td>
<td>0.47**</td>
<td>0.49**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VC</td>
<td>0.06</td>
<td>0.01</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WM</td>
<td>0.27**</td>
<td>0.28**</td>
<td>0.27**</td>
<td>0.54**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO</td>
<td>0.20*</td>
<td>0.08</td>
<td>0.18*</td>
<td>0.49**</td>
<td>0.51**</td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>0.16</td>
<td>0.25**</td>
<td>0.11</td>
<td>0.28**</td>
<td>0.40**</td>
<td>0.34**</td>
</tr>
</tbody>
</table>

Note: Intra-domain correlations printed in normal text, inter-domain correlations are printed in bold.

* significant at the 0.05 level; ** significant at the 0.01 level. (N = 258 for brain volumes, N = 135 for inter-domain correlations; N = 688 for WAIS III dimensions).
What is the internal consistency reliability of the 4 intelligence measures? Is the total score still reliable if we partial out the effects of the brain matter volumes? We run EQS with the setup:

/RELIABILITY
   SCALE = V4 TO V7;
   COVARIATES = V1 TO V3;

The covariates here are observed variables. They affect an IQ factor. Since there are only 4 intelligence measures, we may not get a very high internal consistency reliability.

We get as output:
RELIABILITY COEFFICIENTS USING DEPENDENT VARIABLES ONLY

CRONBACH'S ALPHA = 0.749
COVARIATE-FREE ALPHA = 0.695
COVARIATE-BASED ALPHA = 0.053

We also get results for 1-factor reliability:

RELIABILITY COEFFICIENT RHO = 0.754
COVARIATE-FREE RHO = 0.678
COVARIATE-BASED RHO = 0.076

The intelligence measures retain 93% and 90% of their reliability when the brain volume measures are controlled. But the model fit is a bit marginal.
If we structure the covariates, we obtain better fit and similar $\rho_{xx}$ results, even when models vary somewhat.

$$\hat{\rho}_{xx} = .763, \quad \hat{\rho}_{xx}^{\perp z} = .698, \quad \hat{\rho}_{xx}^{(z)} = .065$$

(Note: $F2 \rightarrow F1 \rightarrow$ Verbal is positive, but $F2 \rightarrow$ Verbal is negative)
Another model also fits well.

\[ \hat{\rho}_{xx} = .761, \quad \hat{\rho}_{xz} = .709, \quad \hat{\rho}_{xz} = .052 \]

(Note: F2 has no effect on Verbal)
Concluding Comments

The proposed *specificity-enhanced* and *covariate-based* reliabilities provide new ways to evaluate the quality of tests and scales.

Like anything else, these methods can probably be misused, e.g.,
- when meaningless auxiliary variables or covariates are used
- when assumptions are not met
- when models $\hat{\Sigma}$ used to define coefficients do not fit the data.
Your feedback is most welcome.

That’s All.
And, thank you again.
References


Bentler, P. M. (1972). A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research, 1*, 343-357.


Bentler, P. M. (2014). Covariate-free and covariate-dependent reliability. Lifetime Achievement Award Address, IMPS 2014, Madison WI.


