New Methods for Test Reliability based on Structural Equation Modeling

Invited Address, 27th Annual APS Convention, New York, May 2015

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

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

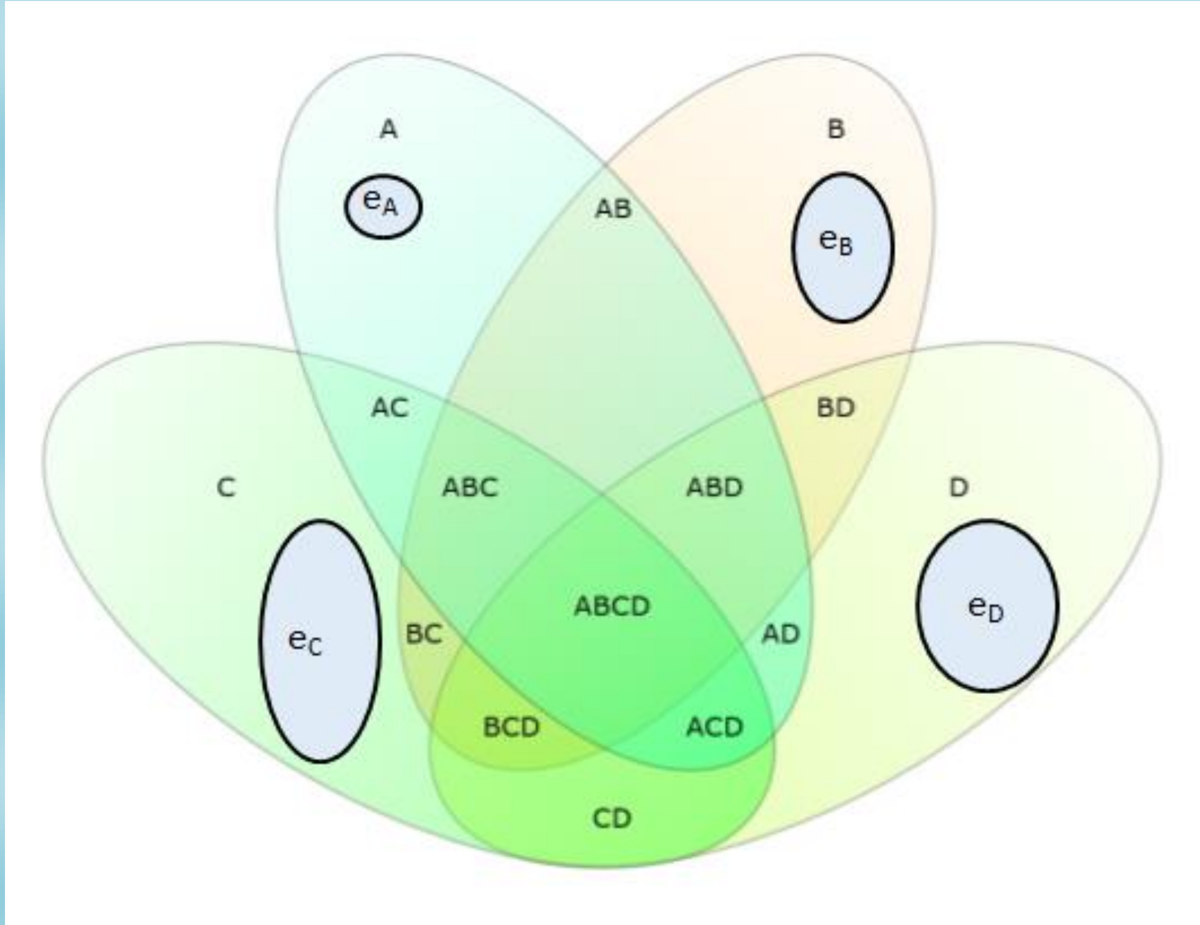
Such a coefficient holds for an item, or a test/scale, here taken simply as $X = \sum_i^p X_i$. Today, I concentrate on the reliability of a scale or test, based on the qualities of its items (*internal consistency*). For simplicity, I assume that errors on different items are uncorrelated.

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


A
B
e_A
AB
e_B
AC
BD
C
ABC
ABD
D
ABCD
e_D
e_C
BC
AD
BCD
ACD
CD

Equations for FA Variance Decomposition

$X = T + E$, but

$T = C + S$ (common plus specific, uncorrelated), so

$X = C + S + E = C + U$,

with $\sigma_X^2 = \sigma_C^2 + \sigma_S^2 + \sigma_E^2$. Thus (Bentler, 1968, 2009, 2015)

$$\rho_{xx} = \frac{\sigma_C^2}{\sigma_X^2} = 1 - \frac{\sigma_U^2}{\sigma_X^2} \le \frac{\sigma_T^2}{\sigma_X^2} = \rho_{xx} + \frac{\sigma_S^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2} = \rho_{XX}.$$

All internal consistency coefficients -- whose history goes back to 1910 (Spearman and Brown) -- are of the form ρ_{xx}. Today, I introduce estimators of σ_S^2 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 $(i = 1, ..., p)$. If $\mathit{1}$ is a unit vector, the variance of the sum $X = \mathit{1}'X_i = \sum_i^p X_i$ is $\sigma_x^2 = \mathit{1}'\Sigma_{xx}\mathit{1}$. Let $\sigma_c^2 \approx p^2\bar{\sigma}_{ij}$, where σ_{ij} is an off-diagonal element of Σ_{xx} and $\bar{\sigma}_{ij}$ is the average of all σ_{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 σ_c^2 and hence ρ_{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 Σ_c is the covariance matrix of common scores and Ψ 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 Σ_c is well-structured (e.g., $\Sigma_c = \Lambda\Phi\Lambda'$), improved estimates of $\sigma_c^2 = 1'\Sigma_c 1$ and hence $\rho_{xx} = \sigma_c^2 / \sigma_x^2$ are possible.

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

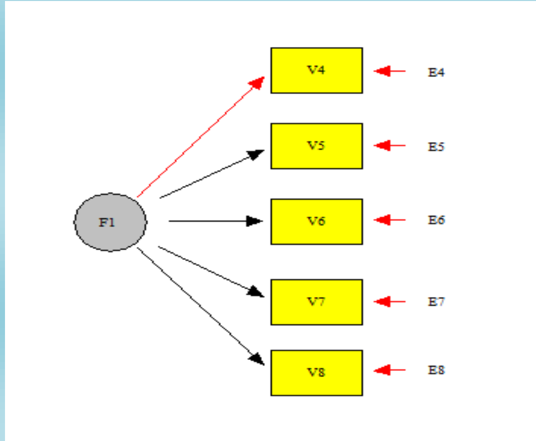
Essentially always $\alpha \leq \rho_{xx} \leq \rho_{XX}$. Next, I show how to obtain α^+ and ρ_{xx}^+ such that $\alpha \leq \alpha^+$ and $\rho_{xx} \leq \rho_{xx}^+$.

Specificity-enhanced Reliability

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

1.0000	0.3900	0.3500	0.2100	0.3200	0.4000	0.3900	0.3900
0.3900	1.0000	0.6700	0.1100	0.2700	0.2900	0.3200	0.2900
0.3500	0.6700	1.0000	0.1600	0.2900	0.2800	0.3000	0.3700
0.2100	0.1100	0.1600	1.0000	0.3800	0.3000	0.3100	0.4200
0.3200	0.2700	0.2900	0.3800	1.0000	0.4700	0.4200	0.5800
0.4000	0.2900	0.2800	0.3000	0.4700	1.0000	0.4100	0.5100
0.3900	0.3200	0.3000	0.3100	0.4200	0.4100	1.0000	0.4200
0.3900	0.2900	0.3700	0.4200	0.5800	0.5100	0.4200	1.0000

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 rs obtain:

	V1	V2	V3	E4	E5	E6	E7	E8
V1	1.0000	0.3890	0.3500	0.2480	0.4740	0.5140	0.4720	0.6330
V2	0.3890	1.0000	0.6700	0.1290	0.4000	0.3680	0.3890	0.4740
V3	0.3500	0.6700	1.0000	0.1830	0.4270	0.3580	0.3700	0.5980
E4	0.2480	0.1290	0.1830	1.0000	0.0000	0.0000	0.0000	0.0000
E5	0.4740	0.4000	0.4270	0.0000	1.0000	0.0000	0.0000	0.0000
E6	0.5140	0.3680	0.3580	0.0000	0.0000	1.0000	0.0000	0.0000
E7	0.4720	0.3890	0.3700	0.0000	0.0000	0.0000	1.0000	0.0000
E8	0.6330	0.4740	0.5980	0.0000	0.0000	0.0000	0.0000	1.0000

Can the E's be predicted from the auxiliary Vs? Doing stepwise regression of each Ei on V1-V3 yields:

$$R^2_{\mathrm{E4.V1}} = .061, R^2_{\mathrm{E5.V1,V3}} = .302, R^2_{\mathrm{E6.V1,V3}} = .300,$$

$$R^2_{\mathrm{E7.V1,V2}} = .292, R^2_{\mathrm{E8.V1,V3}} = .562$$

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:

Vi	$\sigma_{s_i}^2$	$+\ \sigma_{e_i}^2$	$=\ \Psi_{ii}$
V4	0.325	5.009	5.334
V5	1.009	2.332	3.341
V6	3.060	7.140	10.2
V7	1.542	3.738	5.280
V8	1.973	1.537	3.510
SUM	7.909	19.756	27.665

Having the new estimate $\hat{\sigma}_s^2 = 7.909$, 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 2nd 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 $\hat{\sigma}^2$ plus unique $\hat{\sigma}^2$ in the augmented model equals the fixed unique σ^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:

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

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

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

about a 9% improvement. The specific $\hat{\sigma}^2_{sV4} = \hat{\sigma}^2_{F4}$ is not significant – if we set it to zero, we get

$$\hat{\rho}_{xx}^{+} = 1 - (19.704/128.789) = .847 \text{ (a .01 reduction)}$$

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 α 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:

$$\hat{\rho}_{\mathrm{glb}} = .805$$

$$\hat{\rho}^{+}_{\mathrm{glb}} = .876$$

The new glb^{+} exceeds the glb by about 9%.

Covariate-free and Covariate-dependent Reliability Coefficients

Is ρ_{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 ρ_{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 ρ_{xx} size, such as gender.
2. Multiple group models. Invariance or near invariance of parameters implies (near) invariance of ρ_{xx} across groups.
3. Multilevel models. These provide both Between-group (Σ_B) and Within-group (Σ_W) covariance matrices that can be used to obtain ρ_{xx} coefficients. Within-group ρ_{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 ρ_{xx} and ρ_{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 $\sigma_T^2 = \sigma_{\hat{T}}^2 + \sigma_{\tilde{T}}^2$ and hence

$$\rho_{xx} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_{\hat{T}}^2}{\sigma_X^2} + \frac{\sigma_{\tilde{T}}^2}{\sigma_X^2}$$

$$= \rho_{xx}^{(z)} + \rho_{xx}^{\perp z}.$$

$\rho_{xx}^{(z)}$ 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 ρ_{xx}, we have high reliability generalization. Reliability then hardly depends on covariates.

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

Covariate-free & Covariate-dependent Alpha

Based on Σ_{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

$$mean\{offdiag(\Sigma_{xx})\} = mean\{offdiag(\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})\}$$

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

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 Σ_{xz} in the previous formula:

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

It immediately follows that

$$\rho_{xx} = \frac{1'\Sigma_c 1}{1'\Sigma 1} = \frac{1'\Sigma_c^{\perp z} 1}{1'\Sigma 1} + \frac{1'\Sigma_c^{(z)} 1}{1'\Sigma 1}$$
$$= \rho_{xx}^{\perp z} + \rho_{xx}^{(z)}$$

where

$\rho_{xx}^{\perp z}$ is *covariate-free* reliability

$\rho_{xx}^{(z)}$ is *covariate-dependent* reliability.

Σ_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 tr(Σ_c) psd, Ψ diagonal

glb (Woodhouse & Jackson, 1977; Bentler & Woodward, 1980): min tr(Σ_c) psd, Ψ diagonal & psd

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

Covariate-based 1-Factor Reliability

Let $x = \Lambda_1 \xi + \varepsilon$ be the factor model with $\Sigma_c = \Lambda_1 \phi \Lambda_1{}'$. The factor variance ϕ is a scalar (possibly $\phi = 1$). Hence

$$\rho_{xx} = \rho_{11} = \frac{\phi(1'\Lambda_1)^2}{\sigma_x^2} \; (= \omega\ ).$$

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

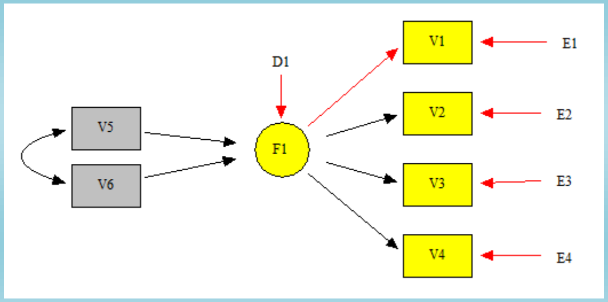
$$\varphi = R^2_{\xi(z)} \phi + (1 - R^2_{\xi(z)}) \phi = \varphi^{\xi(z)} + \varphi^{\perp z}.$$

With the factor variance partitioned, we may write

$$\rho_{11} = \frac{\phi^{\xi(z)}(1'\Lambda_1)^2}{\sigma_x^2} + \frac{\phi^{\perp z}(1'\Lambda_1)^2}{\sigma_x^2} = \rho_{11}^{(z)} + \rho_{11}^{\perp z}.$$

This partition of reliability can be obtained in two ways:

(1) a simultaneous mimic-type setup such as

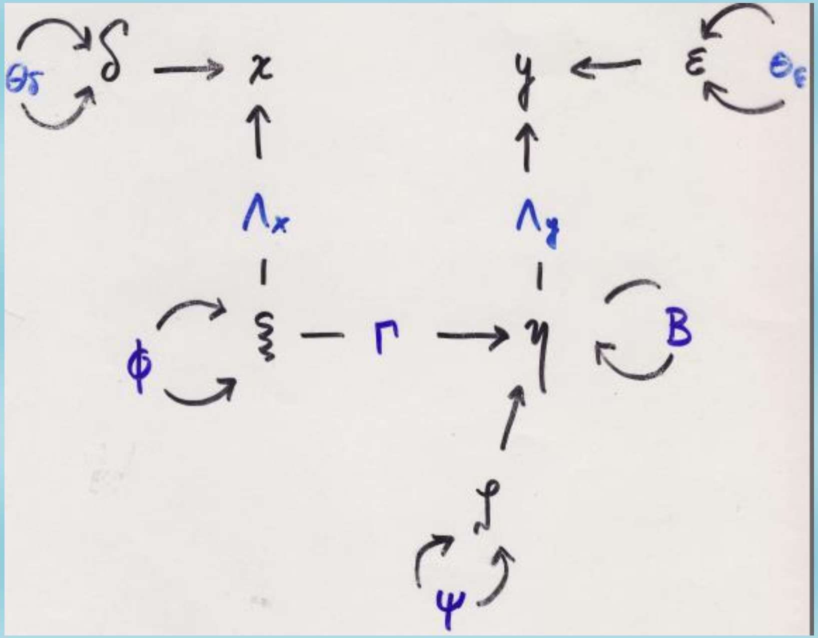


where the equation predicting F1 yields $R^2_{\xi(z)}$ and $\phi^{\perp z}$ is the variance of D1;

(2) a 2-step approach, where ρ_{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

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

and covariate-free reliability is

$$\rho_{yy}^{\perp x} = \frac{1' \Lambda_y (I - B)^{-1} (\Psi)(I - B)'^{-1} \Lambda_y{}' 1}{1' \Sigma_{yy} 1} .$$

Example: Brain Size and IQ

Did you know that "Big-brained people are smarter" (McDaniel, 2005)? He reported:

Table 2
Meta-analytic results for in vivo brain volume and intelligence

Distribution	Number of studies	Sample size	Observed mean correlation	Mean correlation corrected for range restriction
All correlations	37	1530	0.29	0.33
Analyses by whether the degree of range restriction was interpolated				
Interpolation	21	963	0.29	0.32
No interpolation	16	567	0.30	0.34
Analyses by sex				
Females	12	438	0.36	0.40
Males	17	651	0.30	0.34
Mixed sex	8	441	0.21	0.25
Analyses by age				
Adults	24	1120	0.30	0.33
Children	13	410	0.28	0.33
Analyses by age and sex				
Female adults	8	327	0.38	0.41
Female children	4	111	0.30	0.37
Male adults	11	470	0.34	0.38
Male children	6	181	0.21	0.22

Are intelligence measures mainly indirect measures of brain size? Posthuma et al. (2003) found:

Table 2

Pearson Correlations Between Gray Matter Volume, White Matter Volume, Cerebellar Volume, Verbal Comprehension, Working Memory, Perceptual Organization and Processing Speed. Individual Scores on Each Variable Are Adjusted for the Effects of Sex, Age and Cohort

	GMV	WMV	CBV	VC	WM	PO
WMV	0.59**					
CBV	0.47**	0.49**				
VC	**0.06**	**0.01**	**0.03**			
WM	**0.27****	**0.28****	**0.27****	0.54**		
PO	**0.20***	**0.08**	**0.18***	0.49**	0.51**	
PS	**0.16**	**0.25****	**0.11**	0.28**	0.40**	0.34**

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

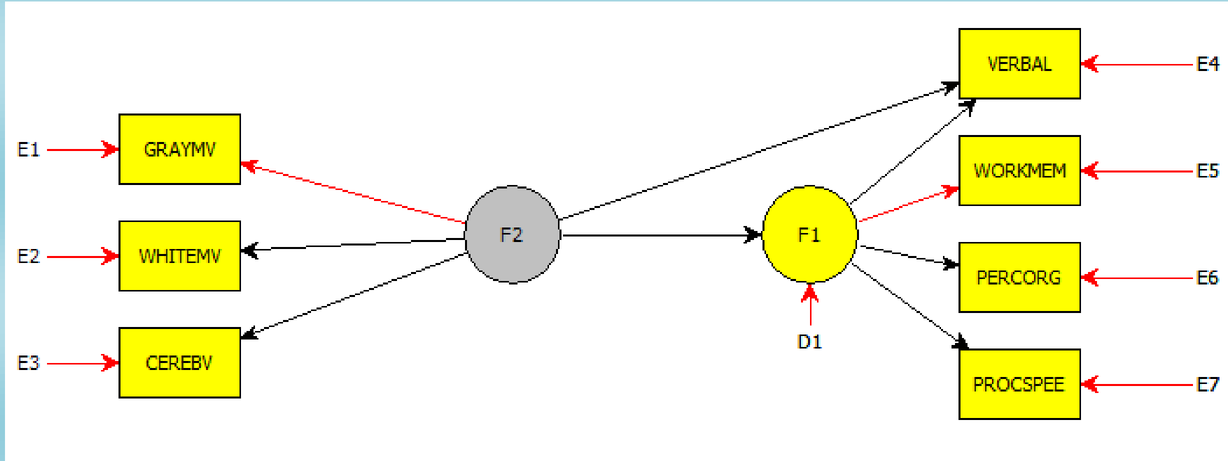
CRONBACH'S ALPHA	=	0.749
COVARIATE-FREE ALPHA	=	0.695
COVARIATE-BASED ALPHA	=	0.053

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.

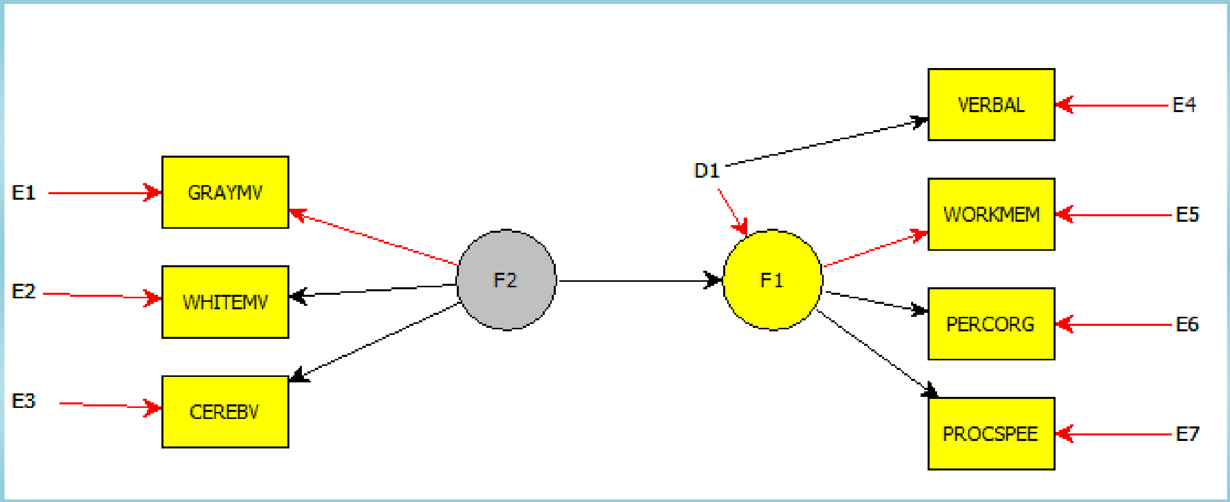
If we structure the covariates, we obtain better fit and similar ρ_{xx} results, even when models vary somewhat.



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

(Note: F2→F1→Verbal is positive, but F2→Verbal is negative)

Another model also fits well.



$$\hat{\rho}_{xx} = .761,\ \hat{\rho}_{xx}^{\perp z} = .709,\ \hat{\rho}_{xx}^{(z)} = .052$$

(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. (1968). Alpha-maximized factor analysis (Alphamax): Its relation to alpha and canonical factor analysis. *Psychometrika*, *33*, 335-345.

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. (2007). Covariance structure models for maximal reliability of unit-weighted composites. In S. –Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 1-19). Amsterdam: North-Holland.

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74*, 137–143.

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

Bentler, P. M. (2015). *Specificity-enhanced reliability coefficients*. Under review.

Bentler, P. M., & Woodward, J. A. (1980). Inequalities among lower bounds to reliability: With applications to test construction and factor analysis. *Psychometrika*, *45*, 249-267.

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods*, *18*, 207-230.

Heise, D. R., & Bohrnstedt, G. W. (1970). Validity, invalidity, and reliability. In E. F. Borgatta & G. W. Bohrnstedt (Eds.) *Sociological methodology* (pp. 104-129). San Francisco: Jossey-Bass.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*, 109-133.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: Guilford.

McDaniel, M. A. (2005). Big-brained people are smarter: A meta analysis of the relationship between in vivo brain volume and intelligence. *Intelligence*, *33*, 337-346.

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*, 1-21.

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

Posthuma, D., Baaré, W. F. C., Hulshoff Pol, H. E., Kahn, R. S., Boomsma, D. I., & De Geus, E. J. C. (2003). Genetic correlations between brain volumes and the WAIS-III dimensions of verbal comprehension, working memory, perceptual organization, and processing speed. *Twin Research*, *6*, 131-139.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604.

Woodhouse, B., & Jackson, P.H. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items. II: A search procedure to locate the greatest lower bound. *Psychometrika*, *42*, 579-591.