"Rassling the hog": the influence of correlated item error on internal consistency, classical reliability, and congeneric reliability

Joseph F. Lucke, State University of New York at Buffalo
"Rassling the Hog": The Influence of Correlated Item Error on Internal Consistency, Classical Reliability, and Congeneric Reliability

Joseph F. Lucke, University of Texas Health Science Center at San Antonio

The properties of internal consistency (\(\alpha\)), classical reliability (\(\rho\)), and congeneric reliability (\(\omega\)) for a composite test with correlated item error are analytically investigated. Possible sources of correlated item error are contextual effects, item bundles, and item models that ignore additional attributes or higher-order attributes. The relation between reliability and internal consistency is determined by the deviance from true-score equivalence. The influence of correlated item error on \(\alpha\), \(\rho\), and \(\omega\) is conveyed strictly through the total item error covariance. As the total item error covariance increases, \(\rho\) and \(\omega\) decrease, but \(\alpha\) increases. The necessary and sufficient condition for \(\alpha\) to be a lower bound to \(\rho\) and to \(\omega\) is that the total item error covariance not exceed the deviance from true-score equivalence. Coefficient \(\alpha\) will uniformly exceed \(\rho\) or \(\omega\) in true-score equivalent tests with positively correlated item error.

Index terms: classical reliability, coefficient alpha, coefficient omega, compound symmetry, congeneric reliability, correlated item error, deviance from true-score equivalence, internal consistency, reliability, total item error covariance, true-score equivalence.

The purpose of this article is to analytically compare internal consistency to reliability for composite tests in which the item errors might be correlated. The psychometric literature variously presents internal consistency as identical to reliability, as another kind of reliability (internal consistency reliability), as a lower bound to reliability, as an approximator to reliability, or as an estimator of reliability. This explication, however, begins at first principles and makes no assumptions regarding the relationship between internal consistency and reliability. The development focuses entirely on internal consistency and reliability as parameters; their estimation is only briefly discussed near the end.

The notation and model specifications, except for the notation for the true score, are borrowed freely from a special case (continuous variables, no exogenous variables) of Muthén’s (2002) structural equation model (see also, Muthén & Muthén, 2001, Appendix 2). This model is equivalent, with slight modifications, to the LISREL Model 3B, supplemented with a means structure (Jöreskog & Sörbom, 1996; Raykov & Penev, 2001).

All covariance matrices are assumed to be positive semidefinite. The covariance matrices need not be of full rank but may contain zero variances or linearly dependent covariances. This assumption
simplifies what follows by permitting sets of items that are linear functions of other items, constant items having zero variances, and items that have zero true-score variances or zero error variances. As no matrix inverses are required in what follows, this assumption entrains no special problems.

Internal Consistency, Classical Reliability, and Congeneric Reliability

Internal Consistency

Suppose one is given a \( p \)-vector \( y = [y_1, \ldots, y_p]' \) of item random variables with positive semidefinite covariance

\[
\text{var } y = \Sigma. \tag{1}
\]

Suppose further that a composite test \( Y \) is the (unit-weighted) sum of these items, \( Y = \mathbf{1}'y \), where \( \mathbf{1} \) is the unit \( p \)-vector. Then,

\[
\text{var } Y = \mathbf{1}'\Sigma \mathbf{1}. \tag{2}
\]

The internal consistency of the composite test \( Y \) is

\[
\alpha \equiv \frac{p}{p-1} \left( 1 - \frac{\text{tr } \Sigma}{\mathbf{1}'\Sigma \mathbf{1}} \right), \tag{3}
\]

where \( \text{tr} \) is the trace operator (Cronbach, 1951; Guttman, 1945). The definition of \( \alpha \) requires no underlying statistical model for the item vector \( y \) other than its covariance be positive semidefinite. The item random variables may have arbitrary (continuous or discrete) distributions subject only to having finite variances.

Classical Reliability

Classical test theory (Lord & Novick, 1968) imposes an additive structure on the items

\[
y = \tau + \epsilon, \tag{4}
\]

where \( \tau \) and \( \epsilon \) are, respectively, \( p \)-vectors of true-score and error random variables. The vector \( \tau \) has finite variance \( \Phi, \epsilon \) has mean 0 with

\[
\text{var } \epsilon = \Theta, \tag{5}
\]

and \( \tau \) and \( \epsilon \) are uncorrelated with each other. No other distributional assumptions are made. The items in \( y \) are usually assumed to be continuous, but there are discrete random variables that satisfy equation (4) (Lord & Novick, 1968, pp. 32-34). Classical test theory also makes the following assumption:

ASSUMPTION 1. (Uncorrelated Item Error). \( \Theta \) is a diagonal matrix.

The composite test is \( Y = \mathbf{1}'y = \mathbf{1}'\tau + \mathbf{1}'\epsilon \), with composite true score \( \tau_Y = \mathbf{1}'\tau \). The variance of the composite true score is then

\[
\text{var } \tau_Y = \mathbf{1}'\Phi \mathbf{1}, \tag{6}
\]

and the variance of the composite test \( Y \) is

\[
\text{var } Y = \mathbf{1}'\Phi \mathbf{1} + \text{tr } \Theta. \tag{7}
\]
The classical reliability of \( Y \) is
\[
\rho = \frac{\text{var} \tau_Y}{\text{var} Y} = \frac{\text{tr} \Phi \Phi' + \text{tr} \Theta}{\text{tr} \Phi \Phi' + \text{tr} \Theta}. \tag{8}
\]

Also under classical assumptions, \( \Sigma = \Phi + \Theta \), and \( \alpha \) specializes to
\[
\alpha = \frac{p}{p-1} \left( \frac{\text{tr} \Phi + \text{tr} \Theta}{\text{tr} \Phi + \text{tr} \Theta} \right). \tag{9}
\]

Furthermore,
\[
\rho = \alpha + \frac{p \text{tr} \Phi - \text{tr} \Phi \Phi'}{(p-1)(\text{tr} \Phi \Phi' + \text{tr} \Theta)} \tag{10}
\]

Adapting an idea from Raykov (1997b, p. 333), the quantity
\[
\delta \equiv \text{tr} \Phi - p^{-1} \text{tr} \Phi \Phi' \tag{11}
\]
is designated as the deviance from true-score equivalence, and the quantity
\[
\frac{p}{p-1} \left( \frac{\delta}{\text{tr} \Phi \Phi' + \text{tr} \Theta} \right)
\]
is designated the relative deviance from true-score equivalence. Equation (29) below ensures
\[
\delta \geq 0. \tag{12}
\]

Although perhaps not intuitive, the name for \( \delta \) is justified by the following:

**Lemma 1.** A vector of items is true-score equivalent if and only if \( \delta = 0 \).

**Proof.** [Proof that true-score equivalence implies \( \delta = 0 \)]. The item vector \( \tau \) is true-score equivalent (essentially \( \tau \) equivalent) if \( \tau = v + \text{tr} \Phi \), where \( v \) is a \( p \)-vector of constants (Lord & Novick, 1968, p. 49). Let \( \text{var} \tau = \phi^2 \), so that \( \Phi = \phi \Phi' \). Then, from equation (11),
\[
\delta = \phi^2 \text{tr} \Phi' \Phi' - p^{-1} \phi^2 \text{tr} \Phi' \Phi' \Phi' = p \phi^2 - p \phi^2 = 0.
\]
The converse is proved in the appendix. \( \square \)

From equation (10), the reliability of a test is its internal consistency plus its relative deviance from true-score equivalence:
\[
\rho = \alpha + \frac{p}{p-1} \left( \frac{\delta}{\text{tr} \Phi \Phi' + \text{tr} \Theta} \right). \tag{13}
\]
Equations (13) and (12) ensure
\[
\rho \geq \alpha. \tag{14}
\]

Lemma 1 shows that internal consistency equals classical reliability when and only when the items are true-score equivalent.
Congeneric Reliability

The generality of classical test theory is achieved at the expense of not specifying how the items are connected to the attributes being measured. Congeneric test theory (Bollen, 1989; Carmines & McIver, 1981; Jöreskog, 1971; McDonald, 1999) is a special case of classical test theory in which the connection between item and attribute is made explicit. Although a special case of classical test theory, its widespread applicability and simplicity entitle it to its own development.

Congeneric test theory specifies that the true score $\tau$ is a linear function of a latent random variable $\eta$ denoting an unobservable attribute:

$$\tau = \nu + \lambda\eta,$$

where $\nu$ is the $p$-vector of item easiness parameters ($-\nu$ is the item difficulties), and $\lambda$ is the $p$-vector of item discriminability parameters with respect to $\eta$. Thus,

$$y = \nu + \lambda\eta + \epsilon.$$  \hspace{1cm} (15)

Congeneric test theory assumes that the attribute $\eta$ has zero mean and unit variance, that the error $\epsilon$ has mean 0 with variance given by equation (5), and that the attribute and error random variables are uncorrelated. Otherwise, no distributional assumptions are made. Also, $\text{var} \tau = \Phi = \lambda\lambda' \Sigma_{\lambda\lambda}$.

Assumption 1 remains in effect. Because the parameter vectors $\nu$ and $\lambda$ may take any real value, the item vector $y$ must be assumed continuous. This restriction incidentally shows the classical item model, equation (4), to be more general than the congeneric model, equation (15).

The composite test is $Y = 1'y = 1'\nu + 1'\lambda\eta + 1'\epsilon$ with composite true score $\tau_Y = 1'\nu + 1'\lambda\eta$. Furthermore, the variance of the composite true score is

$$\text{var} \tau_Y = 1'\lambda\lambda' \Sigma_{\lambda\lambda}.$$  \hspace{1cm} (16)

and the variance of the composite test $Y$ is

$$\text{var} Y = 1'\lambda\lambda' + \text{tr} \Theta.$$  \hspace{1cm} (17)

Additional insight and useful simplifications can be gained by letting $\tilde{\lambda} = p^{-1}1'\lambda$ and then noting that equation (11) becomes

$$\delta = \lambda'\lambda - p\tilde{\lambda}^2 = (\lambda - \tilde{\lambda})'(\lambda - \tilde{\lambda}) \geq 0.$$  \hspace{1cm} (18)

Under congeneric test theory, $\delta$ is the centered sum of squares of the discriminability coefficients. Similar results have been obtained by Raykov (1997b, p. 332) and McDonald (1999, pp. 92-93).

The proof of Lemma 1 is simpler under congeneric test theory. The items in $y$ are true-score equivalent if $\lambda = \lambda I$ (McDonald, 1999, p. 85). In this case, $\tilde{\lambda} = \lambda$, and $\delta = 0$. Conversely, if $\delta = 0$, then from equation (18), $\lambda_i = \tilde{\lambda}$ for $i = 1, \ldots, p$, and $y$ is true-score equivalent.

The congeneric reliability of $Y$ is

$$\omega = \frac{\text{var} \tau_Y}{\text{var} Y} = \frac{1'\lambda\lambda' \Sigma_{\lambda\lambda}}{1'\lambda\lambda' + \text{tr} \Theta} = \frac{p^2\lambda'^2}{p^2\lambda'^2 + \text{tr} \Theta}.$$  \hspace{1cm} (19)

(McDonald, 1985, 1999). Under congeneric assumptions, $\Sigma = \lambda\lambda' + \Theta$, and

$$\alpha = \frac{p}{p-1} \left(1 - \frac{\lambda'\lambda + \text{tr} \Theta}{1'\lambda\lambda' + \text{tr} \Theta}\right) = \frac{p}{p-1} \left(1 - \frac{\delta + p\tilde{\lambda}^2 + \text{tr} \Theta}{p^2\tilde{\lambda}^2 + \text{tr} \Theta}\right).$$  \hspace{1cm} (20)
Furthermore,

\[ \omega = \alpha + \frac{p}{p-1} \left( \frac{\delta}{p^2 \lambda^2 + \text{tr} \Theta} \right) \]  

(21)

(McDonald, 1999, pp. 92-93). Again, the reliability of a test is its internal consistency plus its relative deviance from true-score equivalence. From equation (18),

\[ \omega \geq \alpha. \]  

(22)

Internal consistency equals congeneric reliability when and only when the items are true-score equivalent.

**Correlated Item Error**

Poorly understood is the role that uncorrelated item error, Assumption 1, plays in securing equations (14) and (22). An examination of the proofs of these equations reveals that uncorrelated item error is a only sufficient assumption in the derivations (Lord & Novick, 1968, pp. 88-89; McDonald, 1970, pp. 16-20; McDonald, 1999, pp. 92-93; Novick & Lewis, 1967). The proofs are silent on its necessity. Despite this lacunae in logic, a widespread belief is that correlated item error has little effect on the relation between reliability and internal consistency (Zumbo & Rupp, 2004, p. 79) and that, were it present, “\( \alpha \) would be inflated somewhat, but perhaps not beyond the true value \( \rho \)” (Miller, 1995, p. 266).

**Sources of Correlated Item Error**

One substantive source of correlated item error is contextual effects—“a substantial proportion of the uncontrolled extraneous conditions which are presumably responsible for measurement error [and] will persist from one item to the next” (Rozeboom, 1966, p. 415). However, there appears to be little, if any, research conducted in this area, as item error is rarely the focus of substantive investigations (Zimmerman & Williams, 1980). Consequently, “how severely internal-consistency estimates of reliability are inflated by correlated errors is still an empirical unknown” (Rozeboom, 1989, p. 283).

A second substantive source of correlated item error is item bundles or testlets, a cluster of items that share a common stimulus, contain common content, or possess a common structure (Rosenbaum, 1988; Tuerlinckx & De Boeck, 2004; Wainer & Kiely, 1987; Wilson & Adams, 1995). Items within a bundle can be expected to show intra-bundle correlations in addition to the correlations within the test. Interest in item bundles and testlets originated in item response theory because they violate that theory’s crucial assumption of local independence. An interest in the impact of item bundles and testlets on reliability and internal consistency has recently emerged in classical and congeneric test theory (Feldt, 2002).

The primary methodological source of correlated item error would appear to be model misspecification. Correlated item error can arise from ignoring additional attributes that affect the items (Shevlin, Miles, Davies, & Walker, 2000). Let the true (or empirically adequate) item model be

\[ y = \nu + \lambda_1 \eta_1 + \Lambda_2 \eta_2 + \epsilon^*, \]  

(23)

where \( \eta_2 \) is an \( m_2 \)-vector of additional attributes with zero mean and unit variance, \( \Lambda_2 \) is a \( p \times m_2 \) matrix of discriminabilities, and \( \epsilon^* \) is a \( p \)-vector of item errors, uncorrelated with both \( \eta_1 \) and \( \eta_2 \), with zero mean and

\[ \text{var} \, \epsilon^* = \Theta^*. \]  

(24)
a $p \times p$ diagonal covariance matrix. Assume also that

$$\text{cov}(\eta_1, \eta_2) = \begin{bmatrix} 1 & 0 \\ 0 & \Psi_2 \end{bmatrix},$$

where $\Psi_2$ is an $m_2 \times m_2$ correlation matrix. The true item model is therefore heterogeneous (multidimensional) in that each item reflects more than one attribute, where the second set of attributes is uncorrelated with the first. If the investigator uses a misspecified homogeneous (unidimensional) item model and ignores the additional attributes $\eta_2$, then equation (15) is obtained by setting $\eta = \eta_1$, $\lambda = \lambda_1$, and $\epsilon = \Lambda_2 \eta_2 + \epsilon^*$. But Assumption 1 will not be met because

$$\text{var} \epsilon = \Theta = \Lambda \Psi_2 \Lambda' + \Theta^*$$

will not in general be diagonal.

Correlated item error can also arise from a misspecified structural model, such as might occur with the presence of higher-order attributes. Let the true (or empirically adequate) model be

$$y = \nu + \lambda \eta_1 + \epsilon^* \quad \text{with} \quad \eta_1 = \beta_{12}' \eta_2 + \zeta_1,$$

where $\eta_2$ is an $m_2$-vector of second-order attributes with zero mean and unit variance, $\beta_{12}$ is a $m_2$-vector of regressors of $\eta_1$ on $\eta_2$, and $\zeta_1$ is a disturbance random variable. Assume also that

$$\text{cov}(\zeta_1, \eta_2) = \begin{bmatrix} 1 & 0 \\ 0 & \Psi_2 \end{bmatrix},$$

where $\Psi_2$ is given in equation (25) and $\text{var} \epsilon^*$ is given in equation (24). Again, if the investigator uses a misspecified homogeneous item model and ignores the second-order attributes $\eta_2$, then equation (15) is obtained by setting $\eta = \zeta_1$ and $\epsilon = \lambda \beta_{12}' \eta_2 + \epsilon^*$. But Assumption 1 will not be met because

$$\text{var} \epsilon = \Theta = \lambda \beta_{12}' \Psi_2 \beta_{12} \lambda' + \Theta^*$$

will not in general be diagonal.

### Previous Research

Empirical evidence indicates that correlated item error can render equations (14) and (22) false. Zimmerman, Zumbo, and Lalonde (1993) conducted Monte Carlo comparisons of internal consistency to the known classical reliability of a composite test with true-score equivalent items and positively correlated item error. Upon varying the number of items containing correlated error, the investigators found that $\alpha$ exceeded the known $\rho$ in all but one case. Komaroff (1997) conducted a Monte Carlo investigation of item model misspecification in which the internal consistency of congeneric items containing correlated item error was compared to the reliability of a classical model that assumed uncorrelated item error. In this case, $\alpha$ exceeded the (misspecified) $\rho$ in many but not all cases.

Green and Herschberger (2000) created a true-score equivalent item model with a first-order moving-average item error structure. They then substituted numerical values for the model parameters, obtained the classical reliability, generated the population covariance matrix, and from it obtained the internal consistency. In this numerically specific model, $\alpha$ exceeded $\omega$. These investigators also created a second, Markovian model such that the items were congeneric and the item errors possessed a first-order autoregressive covariance structure. They again substituted numerical
values for the model parameters and obtained the congeneric reliability and internal consistency. And again, \( \alpha \) exceeded \( \omega \).

Intriguing though these results are, their generality is restricted to the parameter space and covariance structures chosen for simulations and substitutions. Many questions remain unanswered. How does correlated item error influence the definitions of \( \alpha \), \( \rho \), and \( \omega \)? What is the functional relationship between the amount of correlated item error and reliability? What is the functional relationship between the amount of correlated item error and internal consistency? What role does the structure of the test (parallel, true-score equivalent, congeneric) play? What role does the structure of the item error covariance matrix (block diagonal, compound symmetric, autoregressive, or moving average) play? What are the necessary conditions to maintain \( \alpha \)'s lower bound property with respect to \( \rho \) and \( \omega \)? Can correlated item error render internal consistency no longer a lower bound to reliability? An analytic analysis is required to acquire a general understanding of the relations among \( \alpha \), \( \rho \), and \( \omega \) under correlated item error.

Guttman (1953) appears to have been the first to conduct an analytic investigation of the influence of correlated item error on internal consistency. His results indicated that the sum of the item error (proper) covariances played a fundamental role in determining the properties of internal consistency. Rozeboom (1966, chap. 9) investigated the impact of correlated item error on classical reliability and internal consistency under an item model very similar to equation (40) in this article (see Rozeboom, 1966, p. 431). His results revealed that a positive sum of item error covariances could decrease \( \rho \) and could also render \( \alpha > \rho \). Raykov (1998b, 2001a) also investigated congeneric reliability and internal consistency under correlated item error. He showed that positively correlated item error could render \( \alpha > \omega \) and that negatively correlated item error could render \( \alpha \) too low a bound for \( \omega \). The purpose of this presentation is to consolidate and extend these analytic results.

**Internal Consistency, Classical Reliability, and Congeneric Reliability Under Correlated Item Error**

From this point forward, the assumption of uncorrelated item error, Assumption 1, is replaced with the following assumption:

**ASSUMPTION 2. (Correlated Item Error).** \( \Theta \) is a positive semidefinite but otherwise arbitrary covariance matrix.

**Special Matrices for Summations**

Two special matrices, when combined with the trace operator, facilitate the summation of the elements of a covariance matrix. Recall that \( I \) is the \( p \)-dimensional unit vector, and let \( I \) denote the \( p \times p \) identity matrix. Define \( J \) and \( H \) as

\[
J = 11' \quad \text{and} \quad H = J - I.
\]

(27)

\( J \) is a \( p \times p \) matrix of ones, and \( H \) is a \( p \times p \) matrix of ones with zeros on the diagonal. Because \( \Sigma \) in equation (1) is square, \( \text{tr} \ J \Sigma \) is the sum of all the elements of \( \Sigma \), \( \text{tr} \ H \Sigma \) is the sum of all the off-diagonal elements, and

\[
\text{tr} \ J \Sigma = \text{tr} \ \Sigma + \text{tr} \ H \Sigma.
\]

(28)

Because \( \Sigma \) is also positive semidefinite,

\[
\text{tr} \ J \Sigma \leq p \ \text{tr} \ \Sigma.
\]

(29)
In other words, the average of the $p^2$ elements of a positive semidefinite covariance matrix cannot exceed the average of the $p$ variances on its diagonal. Proof is given in the appendix. Equations (28) and (29) yield

$$-\text{tr} \Sigma < \text{tr} H \Sigma \leq (p-1)\text{tr} \Sigma. \quad (30)$$

**Internal Consistency**

The definition of internal consistency remains unchanged under correlated item error. Given equations (1) and (2),

$$\alpha \equiv \frac{p}{p-1} \left( 1 - \frac{\text{tr} \Sigma}{\text{tr} J \Sigma} \right) = \frac{p}{p-1} \left( \frac{\text{tr} H \Sigma}{\text{tr} J \Sigma} \right). \quad (31)$$

From equation (31), $\alpha$ is clearly seen as the ratio of the average of the $p(p-1)$ proper covariances to the average of the $p^2$ variances and covariances. If $\text{tr} H \Sigma$ takes on its upper bound $(p-1)\text{tr} \Sigma$ in equation (30), then, from equation (28), $\text{tr} J \Sigma = p \text{tr} \Sigma$. If $\text{tr} H \Sigma$ approaches its lower bound $-\text{tr} \Sigma$, then $\text{tr} J \Sigma$ approaches zero. Using these two results with equation (31) yields

$$-\infty < \alpha \leq 1.$$

**Classical Reliability**

Under classical test theory, the variance of the true score remains as in equation (6), but the variance of the composite test $Y$ is no longer equation (7) but

$$\text{var} Y = \text{tr} J \Phi + \text{tr} \Theta + \text{tr} H \Theta.$$

The classical reliability of $Y$ is then

$$\rho \equiv \frac{\text{var} \tau_y}{\text{var} Y} = \frac{\text{tr} J \Phi}{\text{tr} J \Phi + \text{tr} \Theta + \text{tr} H \Theta}. \quad (32)$$

The quantity $\text{tr} H \Theta$ will be designated as the total item error covariance. Equation (32) reveals a fundamental result:

**PROPOSITION 1.** *The influence of correlated item error on classical reliability is conveyed only by the total item error covariance.*

From this result, it immediately follows that the structure of the $\Theta$, such as having block-diagonal, autocorrelated, moving-average, or compound symmetric errors, is irrelevant. The number of items with correlated error influences classical reliability but only if the total covariance increases as a function of that number. It also immediately follows that a necessary condition to remove the influence of correlated item error on reliability is $\text{tr} H \Theta = 0$, which could occur with a mixture of positively and negatively correlated item errors. Clearly, from equation (32), $0 \leq \rho \leq 1$. As $\text{tr} H \Theta$ increases, $\rho$ decreases. Indeed, from equation (30), as $\text{tr} H \Theta$ increases from $-\text{tr} \Theta$ to $(p-1)\text{tr} \Theta$, $\rho$ decreases as

$$\rho = \begin{cases} 
1 & \text{for } \text{tr} H \Theta = -\text{tr} \Theta; \\
\frac{\text{tr} J \Phi}{\text{tr} J \Phi + \text{tr} \Theta} & \text{for } \text{tr} H \Theta = 0; \\
\frac{\text{tr} J \Phi}{\text{tr} J \Phi + (p-1)\text{tr} \Theta} & \text{for } \text{tr} H \Theta = (p-1)\text{tr} \Theta.
\end{cases}$$
Also under classical test theory assumptions, $\Sigma = \Phi + \Theta$, and

$$\alpha = \frac{p}{p-1} \left( 1 - \frac{\text{tr} \Phi + \text{tr} \Theta}{\text{tr} \bar{J} \Phi + \text{tr} \Theta + \text{tr} \bar{H} \Theta} \right) = \frac{p}{p-1} \left[ \frac{\text{tr} \bar{H} (\Phi + \Theta)}{\text{tr} \bar{J} (\Phi + \Theta)} \right].$$  \hspace{1cm} (33)

Equation (33) exhibits a fundamental result similar to Proposition 1:

**PROPOSITION 2.** The influence of correlated item error on internal consistency under classical test theory is conveyed only by the total item error covariance.

Contrary to $\rho$, as total item error covariance increases, $\alpha$ increases. From equation (30), as $\text{tr} \bar{H} \Theta$ increases from $-\text{tr} \Theta$ to $(p-1)\text{tr} \Theta$, $\alpha$ increases as

$$\alpha = \frac{p}{p-1} \times \begin{cases} 
(1 - \frac{\text{tr} \Phi + \text{tr} \Theta}{\text{tr} \bar{J} \Phi}) & \text{for} \text{ tr} \bar{H} \Theta = -\text{tr} \Theta; \\
(1 - \frac{\text{tr} \Phi + \text{tr} \Theta}{\text{tr} \bar{J} \Phi + \text{tr} \Theta}) & \text{for} \text{ tr} \bar{H} \Theta = 0; \\
\left(1 - \frac{\text{tr} \Phi + \text{tr} \Theta}{\text{tr} \bar{J} \Phi + (p-1)\text{tr} \Theta} \right) & \text{for} \text{ tr} \bar{H} \Theta = (p-1)\text{tr} \Theta.
\end{cases}$$

Comparing $\rho$ to $\alpha$,

$$\rho = \alpha + \frac{p}{p-1} \left[ \frac{\delta - \text{tr} \bar{H} \Theta}{\text{tr} \bar{J} (\Phi + \Theta)} \right].$$  \hspace{1cm} (34)

The necessary and sufficient condition for $\rho \geq \alpha$ is for $\delta \geq \text{tr} \bar{H} \Theta$. If the items are true-score equivalent, $\delta = 0$ and

$$\alpha = \rho + \frac{p}{p-1} \left[ \frac{\text{tr} \bar{H} \Theta}{\text{tr} \bar{J} (\Phi + \Theta)} \right].$$

For internal consistency to remain a lower bound to reliability, the total item error covariance cannot exceed the deviance from true-score equivalence. Equation (34) explains why, as reported in some simulation studies (Komaroff, 1997, p. 347), $\alpha$ remains a lower bound to $\rho$ more often when tests are not true-score equivalent than when they are. The conditions $\delta \geq \text{tr} \bar{H} \Theta$ and therefore $\alpha \leq \rho$ always hold true for items that have nonpositive correlated item error but need not be true items with positively correlated item error. Internal consistency uniformly exceeds classical reliability for true-score equivalent items with positive total item error covariance.

**Congeneric Reliability**

Under congeneric test theory, the variance of the test’s true score remains as in equation (16), but the variance of the composite test $Y$ is no longer equation (17) but

$$\text{var} \ Y = \text{tr} \bar{J} \lambda^t + \text{tr} \Theta + \text{tr} \bar{H} \Theta.$$

The congeneric reliability of $Y$ is

$$\omega = \frac{\text{var} \bar{Y}}{\text{var} \ Y} = \frac{p^2 \bar{\lambda}^2}{p^2 \bar{\lambda}^2 + \text{tr} \Theta + \text{tr} \bar{H} \Theta}. \hspace{1cm} (35)$$

Clearly, $0 \leq \omega \leq 1$. Equation (35) reveals a result virtually identical to Proposition 1:
PROPOSITION 3. The influence of correlated item error on congeneric reliability is conveyed only by the total item error covariance.

Indeed, from equation (30), as \( \text{tr} H \Theta \) increases from \(-\text{tr} \Theta \) to \((p - 1)\text{tr} \Theta \), \( \omega \) decreases as

\[
\omega = \begin{cases} 
1 & \text{for } \text{tr} H \Theta = -\text{tr} \Theta; \\
\frac{p^2 \lambda^2}{p^2 \lambda^2 + \text{tr} \Theta} & \text{for } \text{tr} H \Theta = 0; \\
\frac{p^2 \lambda^2}{p^2 \lambda^2 + \text{tr} \Theta} & \text{for } \text{tr} H \Theta = (p - 1)\text{tr} \Theta.
\end{cases}
\] (36)

Also under congeneric test theory, \( \Sigma = \lambda \lambda' + \Theta \), so that

\[
\alpha = \frac{p}{p - 1} \left( 1 - \frac{\delta + p\lambda^2 + \text{tr} \Theta + \text{tr} H \Theta}{p^2 \lambda^2 + \text{tr} \Theta} \right).
\] (37)

Equation (37) yields a result parallel to Proposition 2:

PROPOSITION 4. The influence of correlated item error on internal consistency under congeneric test theory is conveyed only by the total item error covariance.

From equation (30), as \( \text{tr} H \Theta \) increases from \(-\text{tr} \Theta \) to \((p - 1)\text{tr} \Theta \), \( \alpha \) increases as

\[
\alpha = \frac{p}{p - 1} \times \begin{cases} 
\left( 1 - \frac{\delta + p\lambda^2 + \text{tr} \Theta}{p^2 \lambda^2} \right) & \text{for } \text{tr} H \Theta = -\text{tr} \Theta; \\
\left( 1 - \frac{\delta + p\lambda^2 + \text{tr} \Theta}{p^2 \lambda^2 + \text{tr} \Theta} \right) & \text{for } \text{tr} H \Theta = 0; \\
\left( 1 - \frac{\delta + p\lambda^2 + \text{tr} \Theta}{p^2 \lambda^2 + p\text{tr} \Theta} \right) & \text{for } \text{tr} H \Theta = (p - 1)\text{tr} \Theta.
\end{cases}
\] (38)

Comparing \( \omega \) to \( \alpha \) and using equation (11),

\[
\omega = \alpha + \frac{p}{p - 1} \frac{\text{tr} \lambda \lambda' - \text{tr} J \lambda \lambda' - p \text{tr} H \Theta}{p - 1 \text{tr} J (\lambda \lambda' + \Theta)}
\]

\[
= \alpha + \frac{p}{p - 1} \left( \frac{\delta - \text{tr} H \Theta}{p^2 \lambda^2 + \text{tr} J \Theta} \right).
\] (39)

Equation (39) is similar to equation (34). Raykov (2001a, pp. 71-72) has obtained similar results.

The necessary and sufficient condition for \( \omega \geq \alpha \) is \( \delta \geq \text{tr} H \Theta \). If the items are true-score equivalent, then \( \lambda = \lambda', \delta = 0, \) and

\[
\alpha = \omega + \frac{p}{p - 1} \left( \frac{\text{tr} H \Theta}{p^2 \lambda^2 + \text{tr} J \Theta} \right).
\]

For internal consistency to remain a lower bound to reliability under congeneric test theory, the total item error covariance cannot exceed the deviance from true-score equivalence. This condition always holds true for nonpositive correlated item error but need not be true for positively correlated item error. Internal consistency uniformly exceeds congeneric reliability for true-score equivalent items with positive total item error covariance. These conclusions are identical to those for classical test theory.
Application: The Factor-Analytic Item Model

There need not be any relation between $\rho$ and $\omega$ for classical test theory does not specify the function linking the unobservable attribute random variable to the observed item random variable. However, there is one model in which the relation between the two reliabilities is specified. The factor-analytic item model is similar to the congeneric model but introduces for each item a specific factor that is separate from the item error term. Thus, the item model is

$$y = \nu + \lambda \eta + \zeta + \epsilon,$$

where $\zeta$ is a $p$-dimensional vector of specific factors that has mean $\theta$ and unknown variance $\Psi$ and is uncorrelated with both $\eta$ and $\epsilon$. The specific factors are usually assumed to be uncorrelated with one another—that is, the $p \times p$ matrix $\Psi$ is usually assumed to be diagonal (Bollen, 1989, pp. 218-221; Lord & Novick, 1968, pp. 535-537; McDonald, 1970, pp. 2-3)—but a reviewer requested an analysis allowing them to be correlated. Issues regarding identifiability of the parameters will temporarily be ignored.

An ambiguity in equation (40) turns on whether $\zeta$ is considered part of the true score or the item error (Bollen, 1989, pp. 218-221). Classical test theory considers $\zeta$ to be part of the true score, so the classical true score is $\nu + \lambda \eta + \zeta$ with variance $\lambda \lambda' + \Psi$, and the error is $\epsilon$ with variance $\Theta$ (Bollen, 1989, Equation 6.44). Bollen (1989) argued that reliability should assess only the systematic part of the model, with the remaining random sources of variation, $\zeta$ and $\epsilon$, being error. Congeneric test theory then considers $\zeta$ part of the item error, so that the congeneric true score is $\nu + \lambda \eta$ with variance $\lambda \lambda'$, and the error is $\zeta + \epsilon$ with variance $\Psi + \Theta$ (Bollen, 1989, Equation 6.46). The original motivation for $\omega$ as an alternative to $\rho$ was based on this distinction (McDonald, 1970, pp. 16-20, Equation A.17).

Referring to equation (40) with Assumption 2,

$$\alpha = \frac{p}{p-1} \left[ 1 - \frac{\text{tr} (\lambda \lambda' + \Psi + \Theta)}{\text{tr} (\lambda \lambda' + \Psi + \Theta)} \right],$$

$$\rho = \frac{\text{tr} (\lambda \lambda' + \Psi)}{\text{tr} (\lambda \lambda' + \Psi + \Theta)},$$

and

$$\omega = \frac{\text{tr} \lambda \lambda'}{\text{tr} (\lambda \lambda' + \Psi + \Theta)}.$$

Propositions 1 through 4 still hold true: The effect of the correlated item is conveyed strictly by $\text{tr} H\Theta$ (as part of $\text{tr} J\Theta$). As total item error covariance increases, $\alpha$ increases, but $\rho$ and $\omega$ both decrease.

Comparisons among the three parameters yields

$$\rho = \omega + \frac{\text{tr} J\Psi}{\text{tr} (\lambda \lambda' + \Psi + \Theta)},$$

$$\rho = \alpha + \frac{p}{p-1} \left[ \frac{\delta - \text{tr} H\Theta + \text{tr} \Psi - p^{-1}\text{tr} J\Psi}{\text{tr} (\lambda \lambda' + \Psi + \Theta)} \right],$$

and

$$\omega = \alpha + \frac{p}{p-1} \left[ \frac{\delta - \text{tr} H (\Psi + \Theta)}{\text{tr} (\lambda \lambda' + \Psi + \Theta)} \right].$$
Under the standard assumptions that the specific factors are uncorrelated and the item errors are uncorrelated, \( \alpha \leq \omega \leq \rho \) (McDonald, 1970). Under the assumptions that the specific factors remain uncorrelated but the errors are correlated, \( \alpha \leq \omega \leq \rho \) remains true. For sufficiently large total item error covariance, \( \alpha \) can exceed \( \rho \) or \( \omega \). Coefficient \( \alpha \) exceeds \( \omega \) if \( \text{tr} \ H \Theta > \delta \), the same as in equation (39). Coefficient \( \alpha \) exceeds \( \rho \) (and \( \omega \)) if \( \text{tr} \ H \Theta > \delta + (p - 1) \text{tr} \ \Psi / p \). Unlike in the classical model, \( \alpha \) need not exceed \( \rho \) in true-score equivalent tests with correlated error in the factor-analytic model.

Correlated specific factors have different effects on \( \rho \) and \( \omega \). For \( \alpha \) to exceed \( \rho \), the presence of specific factors (\( \text{tr} \ J \Psi > 0 \)) requires that the total item error covariance be larger than would be required in the absence of specific factors (\( \text{tr} \ J \Psi = 0 \)). For \( \alpha \) to exceed \( \omega \), negatively correlated specific factors (\( \text{tr} \ H \Psi < 0 \)) require the total item error covariance to be larger than would be required for uncorrelated specific factors (\( \text{tr} \ H \Psi = 0 \)), which in turn require it to be larger than would be required for positively correlated factors (\( \text{tr} \ H \Psi > 0 \)).

Although the factor-analytic item model specifies both classical and congeneric reliabilities, its merits remain theoretical as no method has yet been devised to identify \( \zeta \) separately from \( \epsilon \) (McDonald, 1985, pp. 214-215). If the specific factors were assumed to be correlated, then alternative representations might be \( \zeta = \Lambda \eta_1 \) in equation (23) or \( \zeta = \beta \eta_2 \) in equation (26). In practice, to date, \( \zeta \) has invariably been assumed to be identically 0, leaving the congeneric model specified by equation (15) and \( \rho = \omega \).

**Application: Compound Symmetric Item Error Covariance**

Equations (31), (32), (34), (35), and (39) are too general to reveal the influence of correlated item error on the relations between \( \rho \) and \( \alpha \) and between \( \omega \) and \( \alpha \). Because only the total item error covariance influences the behavior of \( \alpha \), \( \rho \), and \( \omega \), a more perspicuous approach would be to develop a special covariance matrix in which the item error structure could be controlled by a single parameter rather than the \( p(p - 1)/2 \) item covariances. One such structure is the compound symmetric or equicorrelated covariance structure, which stipulates that error is equally correlated among items. The compound symmetric item error covariance is

\[
\Theta = [\gamma I + (1 - \gamma) I] \theta^2. 
\]

(41)

The item error variances are \( \theta^2 \), the item error covariances are \( \gamma \theta^2 \), and the item error correlations are \( \gamma \). Partitioning the sum of the item error variances and covariances according to equation (28) yields

\[
\text{tr} \ \Theta = p \theta^2, \\
\text{tr} \ H \Theta = p(p - 1) \gamma \theta^2, \text{ and} \\
\text{tr} \ J \Theta = p[1 + (p - 1) \gamma] \theta^2. 
\]

(42)

Application of equation (30) reveals that for \( \Theta \) to be positive semidefinite,

\[
-\frac{1}{p - 1} \leq \gamma \leq 1. 
\]

Despite its relative simplicity, the compound symmetric matrix is a completely general matrix for investigating the influence of item error correlations on \( \alpha \), \( \rho \), and \( \omega \). From equations (31), (32), and (35), any two covariance matrices \( \Theta^* \) and \( \Theta \) will have the same influence if

\[
\text{tr} \ \Theta^* = \text{tr} \ \Theta \quad \text{and} \quad \text{tr} \ H \Theta^* = \text{tr} \ H \Theta. 
\]

(43)
Thus, if $\Theta^*$ can be transformed into $\Theta$ such that equation (43) obtains, then the influence of $\Theta^*$ can be examined through $\Theta$. In particular, if $\Theta^*$ can be transformed into a compound symmetric matrix $\Theta$, then the influence of $\operatorname{tr} H\Theta^*$ can be investigated by varying $\Theta$'s item error correlation $\gamma$. Any covariance matrix can be transformed to a compound symmetric that satisfies the equalities in equation (43). Let $\Theta^*$ be a $p \times p$ item error covariance. Define the following quantities as

$$\theta^2 = \frac{\operatorname{tr} \Theta^*}{p} \quad \text{and} \quad \gamma = \frac{\operatorname{tr} H\Theta^*}{(p - 1)\operatorname{tr} \Theta^*} = \frac{\operatorname{tr} H\Theta^*}{p(p - 1)\theta^2}.$$  

Define the compound symmetric matrix $\Theta$ using $\theta^2, \gamma$, and the unit $p$-vector $\mathbf{I}$ according to equation (41). From equation (42), $\operatorname{tr} \Theta = \operatorname{tr} \Theta^*$ and $\operatorname{tr} H\Theta = \operatorname{tr} H\Theta^*$. The influence of $\operatorname{tr} H\Theta^*$ can now be investigated using the $\gamma$ of $\Theta$.

From equations (19) and (42), the congeneric reliability of a composite test is

$$\omega = \frac{p\lambda^2}{p\lambda^2 + [1 + (p - 1)\gamma] \theta^2}. \quad (44)$$

Using equation (36) with equation (42) or using equation (44), as $\gamma$ increases from $-(p - 1)^{-1}$ to 1, $\omega$ decreases as

$$\omega = \begin{cases} 1 & \text{for } \gamma = -\frac{1}{(p-1)}, \\ \frac{p\lambda^2}{p\lambda^2 + \theta^2} & \text{for } \gamma = 0; \\ \frac{\lambda^2}{\lambda^2 + \theta^2} & \text{for } \gamma = 1. \end{cases}$$

Negative item error correlation reduces the effect of item error variance, thereby increasing reliability. At the minimum value of the correlation, item error variance is effectively eliminated, yielding perfect reliability. Increasing item error correlation increases the effect of item error variance and effectively reduces the number of items from $p$ to one.

From equations (3) and (42), internal consistency is

$$\alpha = \frac{p(p - 1)(\lambda^2 + \gamma \theta^2) - \delta}{(p - 1)[p\lambda^2 + (1 + (p - 1)\gamma] \theta^2]}. \quad (45)$$

Using equation (38) with equation (42) or using equation (45), as $\gamma$ increases from $-(p - 1)^{-1}$ to 1, $\alpha$ increases as

$$\alpha = \begin{cases} 1 - \frac{p\theta^2 + \delta}{p(p-1)\lambda^2} & \text{for } \gamma = -\frac{1}{(p-1)}, \\ \frac{p(p-1)\lambda^2 - \delta}{(p-1)[p\lambda^2 + \theta^2]} & \text{for } \gamma = 0; \\ 1 - \frac{\delta}{p(p-1)\lambda^2 + \theta^2} & \text{for } \gamma = 1. \end{cases}$$

Comparing $\omega$ to $\alpha$ using equations (39) and (42) or equations (44) and (45) yields

$$\omega = \alpha + \frac{\delta - p(p - 1)\gamma \theta^2}{(p - 1)[p\lambda^2 + (1 + (p - 1)\gamma] \theta^2]}. \quad (46)$$

The necessary and sufficient condition for $\omega \geq \alpha$ is

$$\gamma \leq \min \left\{ \frac{\delta}{p(p - 1)\theta^2}, 1 \right\}. \quad (46)$$
If the items are true-score equivalent, then $\delta = 0$, and

$$\alpha = \omega + \frac{p\gamma\theta^2}{p\lambda^2 + [1 + (p - 1)\gamma]\theta^2}.$$  

For true-score equivalent items, internal consistency will uniformly exceed congeneric reliability when the item error correlation is positive. Furthermore, as the correlation increases to 1, internal consistency will also increase to 1, regardless of the value of the reliability.

A Numerical Example

For purposes of demonstration, consider a set of four hypothetical composite tests—$Y_1, Y_2, Y_3, Y_4$—each with $p = 6$ items, $\lambda = 3$, and $\theta^2 = 9$ but with four different vectors of discriminabilities:

1. $Y_1$ with $\lambda_1 = [3, 3, 3, 3, 3, 3]'$ so that $\delta_1 = 0$,
2. $Y_2$ with $\lambda_2 = [1, 1, 1, 5, 5, 5]'$ so that $\delta_2 = 24$,
3. $Y_3$ with $\lambda_3 = [-1, -1, -1, 7, 7, 7]'$ so that $\delta_3 = 96$, and
4. $Y_4$ with $\lambda_4 = [-4, -4, -4, 10, 10, 10]'$ so that $\delta_4 = 294$.

All four tests have the same congeneric reliability

$$\omega = \frac{6 \times 3^2}{6 \times 3^2 + 9 (1 + 5\gamma)}.$$  

Figure 1 displays the influence of correlated item error on congeneric reliability and internal consistency for the four tests, and Table 1 displays the values of $\omega$ and the $\alpha$s for select values of $\gamma$. As the item error correlation, $\gamma$, increases, the congeneric reliability $\omega$ decreases from 1 at $\gamma = -1/5$ through $6/7 \approx .86$ at $\gamma = 0$ to $1/2$ at $\gamma = 1$. The test $Y_1$ is a true-score equivalent test. Its internal consistency, denoted $\alpha_1$ in the figure, is less than $\omega$ for negative $\gamma$, equals $\omega$ at $\gamma = 0$, and exceeds $\omega$ for all positive values of $\gamma$, with an upper asymptote of 1. The tests $Y_2$ and $Y_3$ show deviance from true-score equivalence. Their respective $\alpha_2$ and $\alpha_3$ remain lower bounds to $\omega$ until

$$\gamma > \delta_2/[p(p - 1)\theta^2] = 24/270 \approx .09$$  

for $Y_2$ and

$$\gamma > \delta_3/[p(p - 1)\theta^2] = 96/270 \approx .36$$  

for $Y_3$, at which points they each exceed $\omega$. The test $Y_4$ is perhaps an extreme case but shows that with sufficiently large deviance from true-score equivalence, internal consistency can remain a lower bound. In this case,

$$\delta_4/[p(p - 1)\theta^2] = 294/270 \approx 1.09$$  

so that equation (46) is always true.

Discussion

Given a $p$-vector of observable random variables possessing second moments, internal consistency, $\alpha$ (equation (3)), is the ratio of the average of the $p(p - 1)$ proper covariances to the average
Table 1
Influence of Correlated Item Error ($\gamma$) on Reliability ($\omega$) and Internal Consistency ($\alpha$) for Four Congeneric Tests

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\omega$</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>$Y_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.2</td>
<td>1.00</td>
<td>0.80</td>
<td>0.71</td>
<td>0.44</td>
<td>-0.29</td>
</tr>
<tr>
<td>0.0</td>
<td>0.86</td>
<td>0.86</td>
<td>0.78</td>
<td>0.55</td>
<td>-0.08</td>
</tr>
<tr>
<td>0.2</td>
<td>0.75</td>
<td>0.90</td>
<td>0.83</td>
<td>0.63</td>
<td>0.08</td>
</tr>
<tr>
<td>0.4</td>
<td>0.67</td>
<td>0.93</td>
<td>0.87</td>
<td>0.70</td>
<td>0.21</td>
</tr>
<tr>
<td>0.6</td>
<td>0.60</td>
<td>0.96</td>
<td>0.91</td>
<td>0.75</td>
<td>0.31</td>
</tr>
<tr>
<td>0.8</td>
<td>0.55</td>
<td>0.98</td>
<td>0.93</td>
<td>0.79</td>
<td>0.39</td>
</tr>
<tr>
<td>1.0</td>
<td>0.50</td>
<td>1.00</td>
<td>0.96</td>
<td>0.82</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Figure 1
Influence of Correlated Item Error ($\gamma$) on Reliability ($\omega$) and Four Values of Internal Consistency ($\alpha_1$, $\alpha_2$, $\alpha_3$, and $\alpha_4$)
of the $p^2$ covariances and variances. This generality is $\alpha$’s greatest advantage and its greatest disadvantage. The advantage is that $\alpha$ can be derived for the item random variables representing almost any psychometric model, be they continuous or discrete, and regardless of their latent structure. The disadvantage is that $\alpha$ itself is derived neither from a psychometric model nor from any psychometric concept. Internal consistency has no intrinsic psychometric interpretation, and its only purpose is to serve as an approximation to reliability (McDonald, 1981, p. 111).

Classical reliability, $\rho$ (equation (8)), is derived from the psychometric model of classical test theory (Lord & Novick, 1968). The model assumes that each item can be additively decomposed into a mutually uncorrelated latent true score and a latent item error. Congeneric reliability, $\omega$ (equation (19)), is derived from the psychometric model of congeneric test theory, an important special case of classical test theory (Carmines & McIver, 1981). The model assumes that each item is linearly related to an unobservable attribute complemented with an additive, uncorrelated, latent item error. Because of its generality, $\alpha$ can be specialized both to classical test theory (equation (9)) and congeneric test theory (equation (20)). Under the additional assumption that item errors are uncorrelated, $\alpha$ is known to be a lower bound to $\rho$ (Lord & Novick, 1968) and $\omega$ (McDonald, 1999), with equality if the items are true-score equivalent.

Raykov’s (1997a) and McDonald’s (1999) independent proposals to employ the congeneric item model to elucidate the relation between $\omega$ and $\alpha$ constituted a significant advance. McDonald’s derivation of the difference between $\omega$ and $\alpha$ as the centered sum of squares of the discriminability coefficients (equations (18) and (21)) yielded additional simplifications. The results for the congeneric item model, in turn, gave impetus for investigating the relation between $\alpha$ and $\rho$ in the more general, classical item model. The new concepts of deviance from true-score equivalence and the relative deviance from true-score equivalence emerged as fundamental to the relation between reliability (classical or congeneric) and internal consistency. Both quantities are nonnegative. A test is true-score equivalent if and only if its deviance from true-score equivalence is zero. Reliability (classical or congeneric) is internal consistency plus the relative deviance from true-score equivalence. The relative deviance from true-score equivalence shows why internal consistency is a lower bound to reliability.

The assumption of uncorrelated item error may be unrealistic. Correlated item error can arise from contextual effects or from misspecified item models. The proof of the lower bound property shows that uncorrelated error is sufficient but is silent on its necessity (Lord & Novick, 1968; Novick & Lewis, 1967). Without the assumption of uncorrelated item error, the possibility remained that the lower bound property reliability need not hold. From the beginning, analytic investigations of correlated item error have focused the sum of the item error (proper) covariances as the prime source of influence on $\rho$ (Guttman, 1953; Rozeboom, 1966). In this investigation, the concept of total item error covariance emerged as fundamental to assessing the effect of correlated item error on $\alpha$, $\rho$, and $\omega$. The influence of correlated item error on $\rho$ (Proposition 1), $\omega$ (Proposition 3), and $\alpha$ (Propositions 2 and 4) is conveyed strictly through the total item error covariance. The structure of the item error covariance matrix (e.g., block diagonal, autocorrelated, moving average, compound symmetric) is irrelevant. Likewise, the structure of the test (e.g., parallel, true-score equivalent, congeneric) is irrelevant.

Correlated item error influences reliability and internal consistency in opposite directions. As the total item error covariance decreases from zero to its lower bound, reliability increases to an upper asymptote of unity, but internal consistency decreases with no lower asymptote. As the total item error covariance increases from zero to its upper bound, reliability decreases to a lower, nonzero asymptote that depends on the average item error variance, but internal consistency increases to an
upper asymptote that depends on the relative deviance from true-score equivalence. If the items are true-score equivalent, \( \alpha \)'s upper asymptote is unity.

Uncorrelated item error is a sufficient but not a necessary condition to establish the lower bound property of internal consistency. A weaker sufficient condition is that the total item error covariance not exceed zero. Surprisingly, the concepts of deviance from true-score equivalence and total item error covariance were both required to establish necessity. The necessary and sufficient condition for internal consistency to be a lower bound to reliability (classical or congeneric) is that the total item error covariance not exceed the deviance from true-score equivalence. True-score equivalent tests have zero deviance and therefore require a nonpositive total item error covariance for the lower bound property to hold. Positively correlated item error will always cause internal consistency to exceed reliability in true-score equivalent tests.

The factor-analytic item model relates \( \rho \) to \( \omega \) and both of them to \( \alpha \). Correlated item error in the factor-analytic item model can cause \( \alpha \) to exceed both \( \rho \) and \( \omega \). Larger total item error covariance is required for \( \alpha \) to exceed \( \rho \) than it is for \( \alpha \) to exceed \( \omega \). In a true-score equivalent test with positively correlated item error, \( \alpha \) will necessarily exceed \( \omega \) but not necessarily \( \rho \). The presence of positively correlated rather than uncorrelated specific factors will reduce the total item error covariance required for \( \alpha \) to exceed \( \omega \).

The compound symmetric matrix is sufficient to study effects of correlated item error. The influence of correlated item error from any item error covariance matrix can be examined by using a corresponding compound symmetric with the same total item error variance and total item error covariance. The influence of total item error can then be investigated by varying the correlation parameter of the compound symmetric matrix.

Correlated item error poses substantial problems for the application of estimated \( \alpha \) to test development. Current psychometric practice maintains that \( \alpha \) should be estimated in all new tests (Nunnally & Bernstein, 1994, p. 235). Implicit in this standard is the assumption that the test’s reliability will be at least as large as the estimated \( \alpha \). But this conclusion is assured only if the total item error covariance is at most zero. Under positively correlated item error, the estimated \( \alpha \) may exceed the reliability and falsely indicate an adequate test. Items for a composite test are often selected with the goal of increasing the estimated \( \alpha \). Again, the assumption is that an increasing \( \alpha \) indicates increasing reliability. Unfortunately, an increasing \( \alpha \) could instead be indicating increasing item error covariance with decreasing reliability as a side effect. Coefficient \( \alpha \) cannot distinguish reliability from correlated item error. “If internal-consistency measures are to [retain] stature in reliability theory . . . explicit provision must be made for the effects of correlated measurement errors.” Other-wise, “the apparent power of internal-consistency . . . estimates is largely illusory . . . [and] for practical applications like putting on a clean shirt to rassle a hog” (Rozeboom, 1966, pp. 438, 415).

Methods now exist for estimating \( \omega \) directly. Assuming the joint multivariate normality of \( \eta \) and \( \epsilon \), equation (15) becomes a normal-theory factor analysis model. The maximum likelihood estimator (MLE) \( \hat{\omega} \) can be constructed from equation (19) using the MLEs of \( \lambda \), \( \Psi \), and \( \Theta \). A first-order approximation to the variance of \( \hat{\omega} \) has been derived (Raykov, 2002). For nonnormal continuous \( \eta \) and \( \epsilon \), consistent estimates of \( \lambda \), \( \Psi \), and \( \Theta \) may be available. The required point and interval estimates can be obtained from software for structural equation modeling (Raykov, 1997a, 1998a, 2001b, 2002; Raykov & Shrout, 2002; Reuterberg & Gustafsson, 1992). In the case of multivariate normality, the structural equation approach provides the added benefit of testing the measurement model’s goodness of fit. Thus, \( \alpha \) is no longer needed for approximating \( \omega \). Perhaps by estimating \( \omega \) and eschewing \( \alpha \), at least in congeneric tests, one can pen the hog while keeping one’s shirt clean.
Appendix

Proof 2. This proof is a matrix version of Lord and Novick’s (1968) proof of their equation (4.4.6). Let \( y_1, \ldots, y_p \) be a vector of random variables. Let \( \Sigma \) be their covariance matrix, with variances \( \sigma_i^2 \) and covariances \( \sigma_{ij} \) for \( i, j = 1, \ldots, p \). Let \( \nu \) be the \( p \)-dimensional vector of variances of \( \Sigma \). And finally, let \( V = \nu^tV + \nu^t \) so that for each \( v_{ij} \in V \), \( v_{ij} = \sigma_i^2 + \sigma_j^2 \). From equation (28), \( \text{tr} J V = \text{tr} V + \text{tr} HV \). But

\[
\text{tr} J V = \text{tr} J v^t + \text{tr} J 1^t v = 2 \text{tr} J v^t = 2 p \text{tr} J v = 2 p \text{tr} \Sigma,
\]
and

\[
\text{tr} V = \text{tr} v^t + \text{tr} 1^t v = 2 \text{tr} v^t = 2 \text{tr} J v = 2 \text{tr} \Sigma.
\]

Thus,

\[
2 p \text{tr} \Sigma = 2 \text{tr} \Sigma + \text{tr} HV.
\]

But for each \( i \neq j \),

\[
v_{ij} = \sigma_i^2 + \sigma_j^2 \geq 2 \sigma_i \sigma_j \geq 2 \sigma_{ij}
\]
by the Cauchy-Schwarz inequality. This implies

\[
\text{tr} HV \geq 2 \text{tr} H \Sigma.
\]
Therefore,

\[
2 p \text{tr} \Sigma \geq 2 \text{tr} H \Sigma = 2 \text{tr} J \Sigma,
\]
which is equation (29).

If \( \delta = 0 \), then the vector of items is true-score equivalent.

Proof 3. To use the notation in the previous proof, let \( \Sigma = \Phi \) in equation (11). From equations (48) and (50), \( 2(p - 1) \text{tr} \Sigma = \text{tr} HV \geq 2 \text{tr} H \Phi \Phi \). Assume \( \delta = 0 \). From equation (11), \( p \text{tr} \Sigma = \text{tr} J \Sigma \), which yields, via equation (28), \( (p - 1) \text{tr} \Phi = \text{tr} J \Phi \). Therefore, \( 2(p - 1) \text{tr} \Sigma = \text{tr} HV = 2 \text{tr} H \Phi \Phi \). Using equation (47),

\[
\text{tr} J V = \text{tr} V + \text{tr} HV = 2 \text{tr} \Sigma + 2 \text{tr} H \Phi \Phi = 2 \text{tr} J \Sigma,
\]

so that \( \text{tr} J (V - 2 \Sigma) = 0 \). By construction, for every \( \sigma_{ij} \in \Sigma \), there is a \( v_{ij} \in V \) such that \( v_{ij} = \sigma_i^2 + \sigma_j^2 \), where \( i, j = 1, \ldots, p \) and \( \sigma_{ii} = \sigma_i^2 \). Each \( v_{ij} \in V \) can be paired with a \( \sigma_{ij} \in \Sigma \) without altering the value of the trace function. But this implies

\[
\sigma_i^2 + \sigma_j^2 \geq 2 \sigma_{ij}
\]
for all \( i \) and \( j \). Otherwise, any nonzero difference would, by the Cauchy-Schwarz inequality, necessarily be positive and render \( \text{tr} J (V - 2 \Sigma) > 0 \). Therefore, \( V = 2 \Sigma \). Combining equations (51) and (49) yields \( \sigma_i^2 + \sigma_j^2 = 2 \sigma_i \sigma_j \) for all \( i \) and \( j \). Therefore, \( \sigma_i = \sigma_j \) for all \( i \) and \( j \). Reverting to the original notation, \( \Sigma = \Phi = \phi^t 1 1' \), from which it follows that \( y \) is true-score equivalent.
References


Acknowledgments
I thank Mark R. Reiser, Mark D. Reckase, three anonymous reviewers, and two anonymous reviewers of a related manuscript for suggesting numerous improvements. Portions of this work were presented at the 2003 Conference of Texas Statisticians at Texas A & M University in College Station, Texas; the 2003 Joint Statistical Meetings in San Francisco, California; and a 2004 meeting of the San Antonio Chapter of the American Statistical Association in San Antonio, Texas. This work was partially supported by NIH-NINR grant R01 NR07891.

Author’s Address
Address correspondence to Joseph F. Lucke, University of Texas Health Science Center at San Antonio, Mail Code 7951, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900; e-mail: lucke@uthscsa.edu.