PSYCHOMETRIKA—VOL. 81, NO. 4, 907–920
DECEMBER 2016
DOI: 10.1007/s11336-016-9524-y

COVARIATE-FREE AND COVARIATE-DEPENDENT RELIABILITY

PETER M. BENTLER

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Classical test theory reliability coefficients are said to be population specific. Reliability generalization, a meta-analysis method, is the main procedure for evaluating the stability of reliability coefficients across populations. A new approach is developed to evaluate the degree of invariance of reliability coefficients to population characteristics. Factor or common variance of a reliability measure is partitioned into parts that are, and are not, influenced by control variables, resulting in a partition of reliability into a covariate-dependent and a covariate-free part. The approach can be implemented in a single sample and can be applied to a variety of reliability coefficients.

Key words: reliability of composites, classical test theory, factor analysis, structural equation modeling, covariance structure analysis.

The idea that test or scale characteristics, especially those associated with classical test theory (see e.g., Raykov & Marcoulides, 2011), may vary depending on the population is hardly novel. One of these characteristics is reliability. Considering that true score variance is liable to be population specific while error variance may be invariant, McDonald (1999, p. 447) notes that an “emphasis on the SE of measurement…warns against regarding the reliability coefficient as invariant.” A similar view was expressed by the APA Task Force on Statistical Inference (Wilkinson & APA, 1999) which states that “…a test is not reliable or unreliable. Reliability is a property of the scores on a test for a particular population of examinees.” This viewpoint (see also Thompson, 1994) 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, and so on. It is certainly justifiable to be interested in reliability of a given scale in certain specific populations. If one wants to go further to characterize how reliability changes across various populations, a methodology should be able to evaluate the degree of consistency across populations of its classical test theory reliability coefficients. This note first reviews the main methodology that has been developed for this purpose. Then it develops a new approach based on the partitioning of true score variance into a part that is uninfluenced by covariates and another part that depends on the covariates.

Reliability generalization is the primary existing methodology that has been used to address the question of consistency of reliability across contexts. As an extension of validity generalization (Schmidt & Hunter, 1977), Vacha-Haase (1998) introduced reliability generalization as a meta-analysis method to determine “(a) the typical reliability of scores for a given test across studies, (b) the amount of variability in reliability coefficients for given measures, and (c) the sources of variability in reliability coefficients across studies” (p. 6). To implement her procedure, she reviewed 628 studies of the Bem Sex Role Inventory and studied 87 coefficients from 57 studies. She obtained different empirical distributions of reliabilities across test forms and, using regression models, was able to explain some 35–44 % of the variance in reliabilities from characteristics such as type of coefficient used, sample size, language used, and response format. Although some

Based on the invited Lifetime Achievement Award Address, International Meeting of the Psychometric Society 2014, Madison WI, July 23, 2014.

Correspondence should be made to Peter M. Bentler, Departments of Psychology and Statistics, University of California, Los Angeles, 4627 Franz Hall, PO Box 951563, Los Angeles, CA 90095-1563 USA.

issues concerning the methodology were raised (Sawilowsky, 2000; Thompson & Vacha-Haase, 2000), the field came to appreciate the importance of the methodology and it has become widely used. Vacha-Haase and Thompson (2011) reviewed 47 reliability generalization meta-analyses and reported, e.g., the majority of studies used multiple regression to predict reliability size, and found that "The most commonly used predictor variables included gender (83.3 % of the 47 studies), sample size (68.8 %), age in years (54.2 %), and ethnicity (52.1 %)" (p. 162). This classical approach, with some improved methodologies (e.g., Beretvas & Pastor, 2003; Bonett, 2010; Botella, Suero, & Gambara, 2010; Brannick & Zhang, 2013; López-López, Botella, Sánchez-Meca, & Marín-Martínez, 2013, Raykov & Marcoulides, 2013), clearly addresses the issue of the stability of reliability coefficients across contexts. However, its implementation requires multiple studies that often may not be available. Methodologies that accomplish somewhat similar goals, even if of a more limited scope, surely would be informative.

Another approach that can be used is generalizability theory (e.g., Brennan, 2001; Shavelson & Webb, 1991). In a generalizability study, analysis of variance methods are used to estimate variance components. These variance components can be adjusted by covariates and hence can be partitioned into predicted and unpredicted parts. Rather than pursue this approach in detail, since reliability coefficients, and especially internal consistency coefficients describing the internal structure of a scale, typically are based on variances and covariances, methods related to structural equation modeling (SEM) or covariance structure analysis would seem especially appropriate to understanding and quantifying effects on reliability. Such methods based on models for a single group are proposed in the next sections. Subsequently, they are illustrated with real data. While methods based on multiple group models with invariance restrictions (e.g., Millsap, 2011) also are relevant, this approach is not implemented herein though it is further described in the Discussion. We now provide our basic definitions.

1. Reliability with Covariates

Let X be the variable of interest, an item or composite variable. Using standard notation, we start with the standard classical test theory decomposition into additive true and error parts

$$X = T + E. \tag{1}$$

With the usual assumption that true and error scores are uncorrelated, the variances are additive

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \quad . \tag{2}$$

The reliability of X is then the well-known ratio

$$\rho_{XX} = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}. \tag{3}$$

Now assume there exists 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

$$T = T^{(Z)} + T^{\perp Z}, \tag{4}$$

with $T^{(Z)} = T(Z)$ the covariate-dependent part of T, and $T^{\perp Z} = T - T(Z)$ the covariate-free part of T. It follows that $\sigma_T^2 = \sigma_{T^{(Z)}}^2 + \sigma_{T^{\perp Z}}^2$ and hence, substituting in (3) yields

$$\begin{aligned}\rho_{XX} &= \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_{T^{(Z)}}^2}{\sigma_X^2} + \frac{\sigma_{T^{\perp Z}}^2}{\sigma_X^2} \\ &= \rho_{XX}^{(Z)} + \rho_{XX}^{\perp Z}. \qquad (5)\end{aligned}$$

We propose to label $\rho_{XX}^{(Z)}$ as the *covariate-dependent reliability* coefficient and $\rho_{XX}^{\perp Z}$ as the *covariate-free reliability* coefficient. In practice, the score decomposition $T = T^{(Z)} + T^{\perp Z}$ is not needed; only the variance decomposition is necessary.

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, if it is absolutely large), reliability is highly population-dependent. Separate coefficients would be needed for different populations. Of course, what is considered to be "large" may depend on the context.

The decomposition (5) is very general and is meant to be applied in a variety of contexts where reliability coefficients are used. An approach to accomplish this in the case of composite variables is described in the next section in a fairly abstract way. To clarify how the abstract decomposition (5) works in practice, we then describe its application to some well-known reliability coefficients. Some examples follow.

2. Lower Bounds to Composite Reliability

Suppose a vector of p item scores x, in deviation form, has decomposition (e.g., Kaiser & Caffrey, 1965)

$$x = c + u, \qquad (6)$$

where c is the vector of common scores and u is the vector of unique scores. We assume that c and u are mutually uncorrelated. The covariance structure of (6) is

$$\Sigma_{xx} = \Sigma_c + \Psi, \qquad (7)$$

where Σ_c is the covariance matrix of common scores and Ψ is the covariance matrix of unique scores. Σ_c is a positive semidefinite matrix of rank $< p$, while Ψ is a full rank matrix often taken as diagonal. In typical covariance structure models, the common scores c are taken as functions of latent variables, e.g., $c = \Lambda\xi$, where Λ is a factor loading matrix and ξ is a vector of factor scores, but c can be structured in terms of any latent variable model (e.g., Bentler & Weeks, 1980; Jöreskog & Sörbom, 1996). When Ψ is diagonal, Σ_c may be considered to be the reduced covariance matrix having communalities in the diagonal, with off-diagonals that are equal to those in Σ_{xx}.

The setup in (6) and (7) allows us to define a coefficient for the reliability of the composite score $X = 1'x = \sum_i^p X_i$. From (6) $X = 1'c + 1'u = C + U$, and from (7) $\sigma_X^2 = 1'\Sigma_{xx}1$, $\sigma_C^2 = 1'\Sigma_c 1$, and $\sigma_U^2 = 1'\Psi 1$, where 1 is a vector of 1s. Since $\sigma_X^2 = \sigma_C^2 + \sigma_U^2$, it is natural to define the coefficient

$$\rho_{xx} = \frac{\sigma_C^2}{\sigma_X^2} = 1 - \frac{\sigma_U^2}{\sigma_X^2}. \qquad (8)$$

TABLE 1.
Internal consistency coefficients.

Σ_c	Comments	References
$\varphi 11'$	α, $\varphi = \bar{\sigma}_{ij}$ average off-diagonal of Σ	Guttman (1945), Cronbach (1951)
$\Lambda_1 \Lambda_1'$	$\rho_{11} = \omega$, Λ_1 is $p \times 1$	Jöreskog (1971), McDonald (1999)
$\Lambda \Lambda'$	$\alpha_0 = \Omega = \theta$, Λ is $p \times k$	Bentler (1968), Heise and Bohrnstedt (1970), McDonald (1970)
$\Lambda \Phi \Lambda'$	Λ is $p \times k$, Φ is symmetric	Werts, Rock, Linn, and Jöreskog (1978)
$\Sigma_c(\theta)$	Any SEM with additive Ψ	Bentler (2007)
$(\Sigma - \Psi)$ psd	Minimum trace ($\Sigma_c = \Lambda\Lambda'$), dimension-free	Bentler (1972)
$(\Sigma - \Psi)$ psd	Minimum trace ($\Sigma_c = \Lambda\Lambda'$), Ψ psd, greatest lower bound	Woodhouse and Jackson (1977), Bentler and Woodward (1980)

Coefficient (8) is a lower bound to reliability, that is, $\rho_{xx} \leq \rho_{XX}$. The inequality is usually strict because the unique variance σ_U^2 is generally too large in a reliability context. That is, since $\sigma_U^2 = \sigma_S^2 + \sigma_E^2$ typically contains specific variance σ_S^2, which is true but unshared variance, $\sigma_U^2 \geq \sigma_E^2$. This downward bias has been known for a long time, and its implications are still being developed (e.g., Bentler, 1968, 2009, 2016). However, the distinction between (3) and (8) does not affect any key features in the current development and so it is not emphasized below.

Various reliability coefficients in the literature are special cases of (8). Some key ones are listed in Table 1 with a few references on each coefficient. For example, when $\Sigma_c = \Lambda_1 \Lambda_1'$, where Λ_1 is a column vector of factor loadings, coefficient (8) is Jöreskog's (1971) single-factor reliability coefficient for an unweighted composite variable, which is now more widely known as McDonald's (1999) ω. The list in Table 1 is illustrative and not exhaustive. For example, the case where $\Sigma_c(\theta)$ is based on any structural model with additive Ψ includes a wide variety of models within broad frameworks such as the Bentler and Weeks (1980) or Jöreskog and Sörbom (1996) models. Some well-known but unlisted coefficients are special cases of those in the table, e.g., Guttman's (1945) λ_4 and its maximized version $\lambda_{4(\max)}$ (Hunt & Bentler, 2015) are split-half coefficients that are computed as α based on the given splits. Others cover more general situations than are required for our analyses (e.g., Brennan, 2001; Tarkkonen & Vehkalahti, 2005).

3. Partition of Common Variance

The partition of composite reliability into covariate-dependent and covariate-free parts is based on the partition of common variance into a part dependent on the covariates and another part that is free from their influence. That is, we will require $\sigma_C^2 = \sigma_{C(Z)}^2 + \sigma_{C\perp Z}^2$ which can be implemented by creating the partition

$$\Sigma_c = \Sigma_c^{(Z)} + \Sigma_c^{\perp Z}. \tag{9}$$

Details on how this can be accomplished in practice are presented in a subsequent section. For now, we note that $1'\Sigma_c 1 = 1'\Sigma_c^{(Z)}1 + 1'\Sigma_c^{\perp Z}1$ which means that

$$\begin{aligned}\rho_{xx} &= \frac{\sigma_C^2}{\sigma_X^2} = \frac{1'\Sigma_c^{(Z)}1}{\sigma_X^2} + \frac{1'\Sigma_c^{\perp Z}1}{\sigma_X^2}\\ &= \rho_{xx}^{(Z)} + \rho_{xx}^{\perp Z}. \end{aligned}\tag{10}$$

4. Equivalence to Simpler Structures

Although Table 1 implies that the composite reliability coefficient (8) describes the reliability of multidimensional sets of variables, applying Bentler's (2007) theorem[1] shows that whatever the rank of Σ_c, coefficients defined on it are equivalent to 1-factor-based reliability for the composite score. In this approach, $\lambda = (1'\Sigma_c 1)^{-.5}\Sigma_c 1$, where λ is the derived factor loading vector for a 1-factor model. It follows that $\sigma_C^2 = 1'\Sigma_c 1 = 1'\lambda\lambda' 1 = (1'\lambda)^2$ applied to (8) is of the form $\rho_{11} = \omega$ as shown in Table 1.

This type of mapping can be extended to the definitions of covariate-dependent and covariate-free reliability. Let $\lambda^{(Z)} = (1'\Sigma_c^{(Z)}1)^{-.5}\Sigma_c^{(Z)}1$ and $\lambda^{\perp Z} = (1'\Sigma_c^{\perp Z}1)^{-.5}\Sigma_c^{\perp Z}1$. Then $1'\Sigma_c^{(Z)}1 = 1'\lambda^{(Z)}\lambda^{(Z)\prime}1$ and $1'\Sigma_c^{\perp Z}1 = 1'\lambda^{\perp Z}\lambda^{\perp Z\prime}1$, so that

$$\begin{aligned}\rho_{xx} &= \frac{1'\lambda^{(Z)}\lambda^{(Z)\prime}1}{\sigma_X^2} + \frac{1'\lambda^{\perp Z}\lambda^{\perp Z\prime}1}{\sigma_X^2}\\ &= \rho_{XX}^{(Z)} + \rho_{XX}^{\perp Z}. \end{aligned}\tag{11}$$

Covariate-dependent reliability can be conceived as based on a unidimensional representation, as can covariate-free reliability. [2]

The above general conceptual methodology for partitioning reliability into covariate-dependent and covariate-free parts will next be applied to some specific reliability coefficients. We begin with the most widely known coefficient, Guttman's (1945) λ_3, now known as Cronbach's (1951) α.[3]

5. Covariate Partition of Alpha and Lambda 4

Let the population covariance matrix Σ_{xx} given in (8) have off-diagonal elements σ_{ij}, with $\bar{\sigma}_{ij}$ being the average of all σ_{ij}. Let $\Sigma_c = \varphi 11' = \bar{\sigma}_{ij}11'$. Then the reliability coefficient (8) becomes

$$\alpha = \frac{p^2\bar{\sigma}_{ij}}{\sigma_x^2}.\tag{12}$$

[1] The theorem states that when Σ_c is not rank 1, composite reliability coefficients defined on it are identical to a 1-factor-based reliability coefficient for the composite based on a rotated factor whose loading vector λ maximizes $(1'\lambda)^2$. Subsequent rotated factors have loadings whose columns sum to zero.

[2] Other approaches are possible. We could take $\tilde{\lambda}^{(Z)} = (1'\Sigma_c 1)^{-.5}\Sigma_c^{(Z)}1$ and $\tilde{\lambda}^{\perp Z} = (1'\Sigma_c 1)^{-.5}\Sigma_c^{\perp Z}1$ but these would not have the desired property of (11).

[3] A recent discussion on the interpretation of α in terms of all possible k-split alphas is given by Warrens (2014).

With a $q \times 1$ vector of covariates z, we also have $\begin{pmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix}$. The regression of x on z yields the well-known matrix identity $\Sigma_{xx} = (\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}) + (\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})$. Hence, their off-diagonal elements obey the equality

$$\begin{aligned} mean\{offdiag(\Sigma_{xx})\} &= mean\{offdiag(\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})\} \\ &\quad + mean\{offdiag(\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})\} \end{aligned}$$

and so, with obvious notation, $\bar{\sigma}_{ij} = \bar{\sigma}_{ij}^{\perp Z} + \bar{\sigma}_{ij}^{(Z)}$. It follows that alpha can be decomposed into

$$\alpha = \alpha^{\perp Z} + \alpha^{(Z)}, \tag{13}$$

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*.

An obvious application of (13) is when the covariates z are observed variables. They could also be latent variables, in which case some preliminary structural equation modeling would be needed to obtain model-implied covariances equivalent to Σ_{xx},Σ_{xz}, and Σ_{zz}. Clearly, the specific nature of these would depend on the particular model being used, as well as the data.

The decomposition (13) also can be applied in other circumstances. For example, suppose the items X_i of vector x are assigned to one of two parts, and part total scores X_A and X_B are obtained so that $X = 1'x = X_A + X_B$. Then Guttman's (1945) λ_4 coefficients are given by $\alpha(X_A, X_B) = \lambda_4$. The split of items that maximizes $\alpha(X_A, X_B)$ is $\lambda_{4(\max)}$. Experience indicates that $\lambda_{4(\max)} \geq \alpha$, but proof exists only for an even number of items (Jackson, 1979). Hunt and Bentler (2015) provided an effective algorithm for computing $\lambda_{4(\max)}$, and also provided a chain of quantile coefficients $\lambda_{4(Q)}$for different optimized splits that are less likely to capitalize on chance in small samples while still improving substantially on α. Interesting quantile values in the distribution of local optima for various splits are $Q = .05, .50, .95, 1.0$ with $\lambda_{4(.05)} \leq \lambda_{4(.50)} \leq \lambda_{4(.95)} \leq \lambda_{4(1.0)}$. Even the smallest of these is typically a better lower bound reliability coefficient than α. It is often closer to reliability than the greatest lower bound (Bentler & Woodward, 1980; Woodhouse & Jackson, 1977).[4] Every $\lambda_{4(Q)}$ coefficient, including $\lambda_{4(Q=1.0)} = \lambda_{4(\max)}$is an α coefficient defined on some partition of items. Hence by the approach above, we may obtain

$$\lambda_{4(Q)} = \lambda_{4(Q)}^{\perp Z} + \lambda_{4(Q)}^{(Z)}, \tag{14}$$

covariate-free and covariate-dependent λ_4 coefficients.

6. Covariate Partition of General Coefficients

Suppose we are dealing with the composite reliability coefficient (8) under some structural model, and we would like to partial the 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 $E(xz') = E(cz')$ or $\Sigma_{xz} = \Sigma_{cz}$. Now we can substitute Σ_{xz} in the previous formula:

$$\Sigma_c = \Sigma_{cc} = \left(\Sigma_{cc} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\right) + \left(\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\right) = \Sigma_c^{\perp Z} + \Sigma_c^{(Z)}. \tag{15}$$

[4] The greatest lower bound can be biased in small samples; Li and Bentler (2011) remove this bias. Note that Jackson and Agunwamba (1977) provide a condition under which $\lambda_{4(\max)}$ is the greatest lower bound.

Inserting (15) into (9) immediately gives the general partition described previously in (10), that is, $\rho_{xx} = 1'\Sigma_c 1/1'\Sigma 1 = 1'\Sigma_c^{\perp Z} 1/1'\Sigma 1 + 1'\Sigma_c^{(Z)} 1/1'\Sigma 1 = \rho_{xx}^{\perp Z} + \rho_{xx}^{(Z)}$, where $\rho_{xx}^{\perp Z}$ is covariate-free reliability and $\rho_{xx}^{(Z)}$ is covariate-dependent reliability.

Clearly, this is a general partition that holds for any latent variable model with additive errors for which $\Sigma_{xz} = \Sigma_{cz}$. To implement it requires computing the model-implied covariance matrices of the SEM, then further specifying c as a function of the latent variables in that specific model, and evaluating the assumption $\Sigma_{xz} = \Sigma_{cz}$. If the assumption does not apply, (15) cannot be used.

Actually, in most circumstances where focus is on a given latent variable model structure such as $c = \Lambda\xi$, the partitioning of variance can and should be done directly on the latent variables rather than on the c variables. Path tracing or its matrix algebra equivalent can be used to generate the required model-implied matrices for the direct use of the basic formulae (9) and (10). A concrete way of accomplishing this, in the context of the most typical model-based reliability model structure, is discussed next.

7. Covariate Partition of Factor-based Reliability

When the common scores have decomposition $c = \Lambda_1\xi$, where Λ_1 is a column vector, $\Sigma_c = \Lambda_1\varphi\Lambda_1'$, where φ is the variance of the factor ξ. Now let the ξ be predicted by covariates z with the R^2 for predicting ξ being $R^2_{\xi(Z)}$. Thus $\varphi = R^2_{\xi(Z)}\varphi + (1 - R^2_{\xi(Z)})\varphi = \varphi^{\xi(Z)} + \varphi^{\perp Z}$, so that the partition (9) of the common covariance matrix under the 1-factor model becomes $\Sigma_c = \Sigma_c^{(Z)} + \Sigma_c^{\perp Z} = \Lambda_1\varphi^{\xi(Z)}\Lambda_1' + \Lambda_1\varphi^{\perp Z}\Lambda_1'$. It follows that we obtain the special case of (10)

$$\rho_{11} = \omega = \frac{\varphi^{\xi(Z)}(1'\Lambda_1)^2}{\sigma_x^2} + \frac{\varphi^{\perp Z}(1'\Lambda_1)^2}{\sigma_x^2} = \rho_{11}^{(Z)} + \rho_{11}^{\perp Z}, \tag{16}$$

a covariate-based partition of coefficient ω.

To obtain the partition (16) in any given application, we can use two slightly different approaches. The first is a standard SEM run including variables of interest plus covariates. This has the advantage of simplicity and clear statistical inference. A possible disadvantage is that the estimate of ρ_{11} might not be identical to that which would be obtained by analyzing only the variables in the reliability model. The second is a two-stage approach that obtains ρ_{11} from only the variables in the reliability model in the first stage, and obtains the quantities required to partition φ in the second stage.

(1) A simultaneous mimic-type model setup as shown in Fig. 1. This represents a 1-factor model for variables V4–V7, with the factor F1 being predicted by three covariates V1–V3. The equation that is used to predict F1 will yield $R^2_{\xi(Z)}$, which needs to be multiplied by the model-reproduced variance of F1 to obtain $\varphi^{\xi(Z)}$. The variance of D1 will be the variance $\varphi^{\perp Z}$.

(2) A 2-step approach, where ρ_{11} and φ are first obtained from running only the factor model with no covariates. Of course, φ could be fixed or free. In step 2, the model is run with loadings and error variances fixed at step-1 values, and the other parameters are considered free to be estimated. This run will produce $R^2_{\xi(Z)}$, which can be applied to φ to yield the required variance partition.[5]

[5] Alternatively, step 2 can produce $\varphi^{\xi(Z)}$ and $\varphi^{\perp Z}$ as described in the first approach, but these values are not guaranteed to precisely add to φ from step 1 in the 2-step approach.

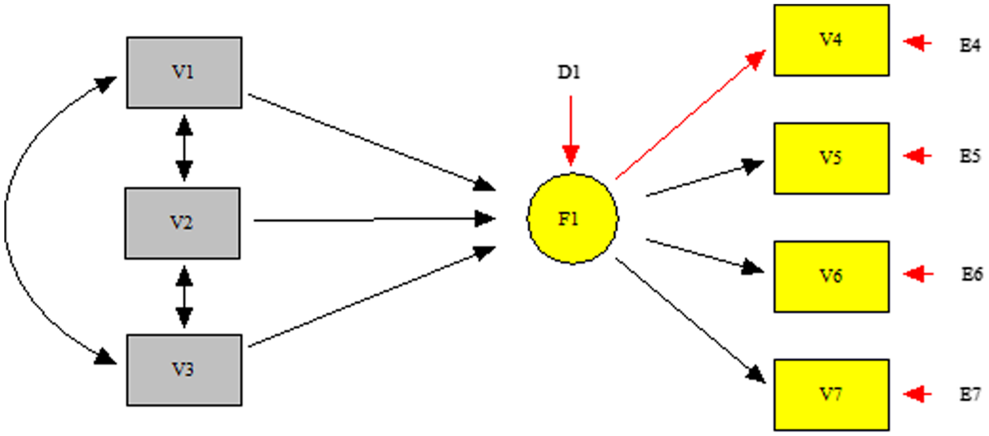

FIGURE 1.
Example of mimic model for partitioning ρ_{11}.

The approaches described for the 1-factor case clearly can be adapted to partition reliability based on multiple factor models. When $\Sigma_c = \Lambda\Phi\Lambda'$, Fig. 1 would have several possibly correlated factors that define reliability, and the covariates would be used to predict all of these factors. Again, both simultaneous and two-step approaches are possible. Since the dimension-free and greatest lower bound computations result in a Gramian Σ_c (Jamshidian & Bentler, 1998), $\Sigma_c = \Lambda\Lambda'$ is possible for any choice of rotation of Λ when doing the two-step approach. Actually, the simultaneous model would not be identified in such a setup, so the first option would not be available.

Next, we clarify how covariate-free and covariate-dependent reliability can be obtained in a general structural equation model.

8. Covariate Partition of Reliability in Structural Equation Models

Since there exists a number of specific types of structural equation models, and the LISREL model (Jöreskog & Sörbom, 1996) can be specialized to cover the vast majority of model types used in practice, we use LISREL to illustrate covariate partitioning of reliability in SEM. In a standard approach with LISREL, there are x variables that have a measurement model with factors $\xi(x = \Lambda_x\xi + \delta)$, there are y variables that have a measurement model with factors $\eta(y = \Lambda_y\eta + \varepsilon)$, and there is a simultaneous equation system that relates the ξ and η factors across these two sets of variables ($\eta = B\eta + \Gamma\xi + \zeta$). We assume that we want the reliability of the endogenous y variables, and that the x variables and its factors are covariates. As is well-known, the covariance matrix of the y variables is

$$\Sigma_{yy} = \Lambda_y(I - B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I - B)'^{-1}\Lambda_y' + \Theta_\varepsilon, \tag{17}$$

where Φ, Ψ, and Θ_ε are the covariance matrices of ξ, ζ, and ε respectively. Then, the composite reliability of $Y = 1'y$ is

$$\rho_{yy} = \frac{1'\Lambda_y(I - B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I - B)'^{-1}\Lambda_y'1}{1'\Sigma_{yy}1}. \tag{18}$$

TABLE 2.
Correlations between brain volumes and WAIS-III variables.

	GRAYMV	WHITEMV	CEREBV	VERBAL	WORKMEM	PERCORG	PROCSPEE
V1 GRAYMV	1.00						
V2 WHITEMV	0.59	1.00					
V3 CEREBV	0.47	0.49	1.00				
V4 VERBAL	0.06	0.01	0.03	1.00			
V5 WORKMEM	0.27	0.28	0.27	0.54	1.00		
V6 PERCORG	0.20	0.08	0.18	0.49	0.51	1.00	
V7 PROCSPEE	0.16	0.25	0.11	0.28	0.40	0.34	1.00

From Posthuma et al. (2003, Table 2), reprinted with permission.

In (17) and (18), $\Gamma\Phi\Gamma'$ is that part of the model-implied covariance matrix of the η that is based on the ξ factors, here taken as covariates. The part of the model-implied covariance matrix of the η that is not predicted by the covariates is Ψ. We immediately see that the predicted part of Σ_c is $\Sigma_c^{(X)} = \Lambda_y(I-B)^{-1}(\Gamma\Phi\Gamma')(I-B)'^{-1}\Lambda_y'$ so that covariate-dependent reliability of the y variables is

$$\rho_{yy}^{(X)} = \frac{1'\Lambda_y(I-B)^{-1}(\Gamma\Phi\Gamma')(I-B)'^{-1}\Lambda_y'1}{1'\Sigma_{yy}1}. \tag{19}$$

Similarly, $\Sigma_c^{\perp X} = \Lambda_y(I-B)^{-1}(\Psi)(I-B)'^{-1}\Lambda_y'$, so that covariate-free reliability is given by

$$\rho_{yy}^{\perp X} = \frac{1'\Lambda_y(I-B)^{-1}(\Psi)(I-B)'^{-1}\Lambda_y'1}{1'\Sigma_{yy}1}. \tag{20}$$

Different model structures, such as those derived from the Bentler–Weeks model, involving different partitions of variables into measurement model variables and covariates, would yield different matrix representations. Those are too specific to be presented here.

9. Illustrative Application to Brain Size and IQ

An interesting literature relates brain volume measurements to intellectual-type variables such as are used to measure IQ. For example, in his provocative article "Big-brained people are smarter," McDaniel (2005) summarized the relation between in vivo brain volume and intelligence across 37 different studies that included subjects of both genders and various ages. Based on 1530 subjects across these studies, he reported that the mean correlation between these two domains was 0.29. This raises the interesting question of whether the reliability of intelligence tests is simply a reflection of individual differences in brain size. Has psychological testing of intelligence simply yielded an indirect measure of brain size? This question can be answered by evaluating whether intelligence measures would remain reliable if brain size differences were controlled.

Posthuma et al. (2003) provided a correlation matrix of the relations among these two domains, adjusted for the effects of sex, age, and cohort. The brain size volumes measured gray matter volume, white matter volume, and cerebellar volume. The intelligence measures were verbal comprehension, working memory, perceptual organization, and processing speed. The correlations among these variables are reproduced in Table 2.[6] Using maximum likelihood estimation in EQS

[6] We treat the correlations as covariances, and ignore the fact that these correlations are based on different sample sizes, $N=258$ for brain volumes, $N=135$ for inter-domain correlations, $N=688$ for WAIS-III variables.

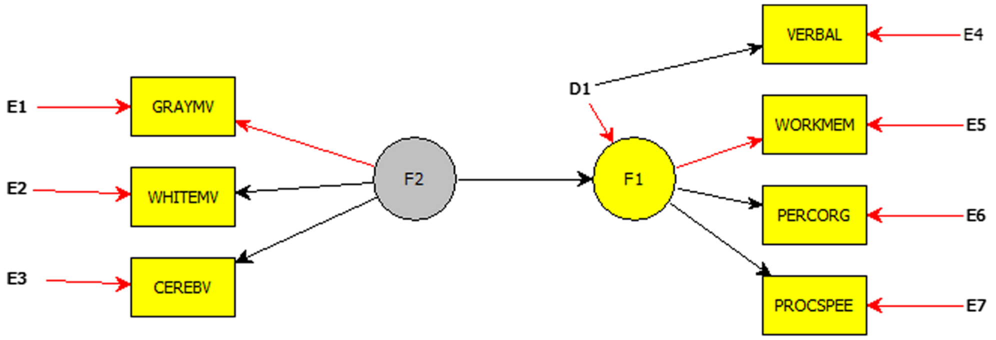

FIGURE 2.
Brain volume and IQ model based on Posthuma et al. data.

(Bentler & Wu, 2015) with the command /RELIABILITY with SCALE = V4–V7 and COVARIATES = V1–V3, EQS created the model of Fig. 1 with the 3 brain volume measures V1–V3 as covariates and a 1-factor model for the 4 WAIS-III variables V4–V7. It produced the output

```
CRONBACH'S ALPHA = 0.749
COVARIATE-FREE ALPHA = 0.695
COVARIATE-BASED ALPHA = 0.053
RELIABILITY COEFFICIENT RHO = 0.754
COVARIATE-FREE RHO = 0.678
COVARIATE-BASED RHO = 0.076
```

The intelligence measures retain 93 % of their α reliability and 90 % of their ω reliability when brain volume measures are controlled. The 1-factor ω model fits reasonably well, with a comparative fit index (CFI) of .955.

In a second model, the correlations among brain measures were modeled by a latent factor, which in turn predicted the IQ factor. This model fits reasonably (CFI = .961), but correlations of brain volume with V4, the verbal comprehension measure, were not reproduced well. Looking at the correlations in Table 2, it is clear that the brain volume is essentially uncorrelated with verbal comprehension. Adding a path from the brain volume factor directly to V4 in the 3rd model led to a near perfect fit with CFI = .999. While fit was good, in this model, the direct effect of the brain volume factor on V4, and its indirect effect via the IQ factor, were opposite in sign to allow the near-zero correlations for V1–V3 with V4 to be reproduced well. The reliability partition was similar to before, $\hat{\omega} = \hat{\rho}_{11} = .763$, $\hat{\rho}_{11}^{\perp Z} = .698$, $\hat{\rho}_{11}^{(Z)} = .065$.

A more satisfactory version of this model is given in Fig. 2, where verbal comprehension is part of the IQ factor, but it cannot be impacted by the brain volume factor. To provide some insight into the symbolic components for $\Sigma_c^{\perp Z}$ and $\Sigma_c^{(Z)}$, let λ_1 be the factor loading for verbal, λ_2 be the 3×1 factor loadings for working memory, perceptual organization, and processing speed, φ be the variance of F2, γ be the path coefficient F2 → F1, and ϕ be the variance of D1. Then it can be shown that

$$\Sigma_c^{\perp Z} = \phi \begin{bmatrix} \lambda_1^2 & \lambda_1\lambda_2' \\ \lambda_1\lambda_2 & \lambda_2\lambda_2' \end{bmatrix} \text{ and } \Sigma_c^{(Z)} = \begin{bmatrix} 0 & 0' \\ 0 & \varphi\gamma^2\lambda_2\lambda_2' \end{bmatrix}.$$

This model fits very well, with CFI = 1.00. Partitioning of 1-factor reliability yielded $\hat{\omega} = \hat{\rho}_{11} = .760$, $\hat{\rho}_{11}^{\perp Z} = .693$, $\hat{\rho}_{11}^{(Z)} = .067$.

Reliability and its partition is very similar among well-fitting alternative models in this example. We would expect such a result to be found more generally. These analyses lead to the expected conclusion that intelligence measures appear to primarily reflect some psychological quality beyond variation in brain volume. Posthuma et al. (2003) raise the possibility that this quality might be substantially genetically driven. Pursuing such a possibility goes beyond the purpose of this illustration.

10. Discussion

The proposed methodology seems to be a useful way to evaluate whether reliability may be generalizable across contexts. While it can be argued from classical test theory that reliability coefficients will always be population specific, it is quite possible that in some circumstances, or with some constructs, variation across populations may have only a small effect. In such a case, a single coefficient could be utilized to simplify discussion of the reliability of a scale. In other cases, e.g., in developmental psychology where different variances on true or total scores may be expected at different ages, varying reliability coefficients may be natural. Of course, the proposed methodology then could quantify the extent to which age might affect reliability.

Like anything else, the proposed methodology requires some thought to be applied meaningfully and it can no doubt be misused. Among the more obvious problems to be avoided are the use of meaningless covariates, or covariates whose control addresses no serious problem related to influences on reliability. Covariates that might be of interest are demographic variables that have been studied in reliability generalization, such as age, gender, and ethnicity, but other types of individual difference covariates (e.g., the brain volume covariate example used in this paper) can be used as well if their use addresses a meaningful substantive question. As in all research with covariates, care must be taken in choice of covariates since omission of key covariates could lead to biased conclusions. Some covariates used in reliability generalization research (e.g., sample size, type of reliability coefficient) cannot be studied with the proposed methodology since they require the existence of many datasets. As noted earlier, the proposed approaches are based on structural equation models, and hence any model being used for reliability purposes must be a plausible model for the given data.

Our approach to partition reliability coefficients into covariate-dependent and covariate-free parts is only one structural modeling-based approach to gain insight into the susceptibility of reliability coefficients to population variability. As noted in the introduction, reliability generalization is another very meaningful approach, but it cannot be applied in the simple one-sample data setup we have been working with. Actually, some other SEM-based approaches could also be considered.

Classical test theory reliability coefficients typically are based on models for the covariance matrix that contain parameters such as factor loadings and error variances. If reliability varies by group, one likely reason is that the item parameters are not invariant. Evaluating the invariance of parameters across discrete groups is an old topic in structural equation modeling (SEM) (e.g., Sörbom, 1974) that remains a thoroughly modern problem (Deng & Yuan, 2015; Merkle & Zeileis, 2013; Millsap, 2011; Raykov, Marcoulides, & Millsap, 2012). Even ignoring invariance of mean parameters, which are not relevant in typical reliability contexts, in practice tests based on many groups will almost always reject equality of all structural parameters across groups, especially when invariance of unique variances is also required. This implies that reliability typically will vary by group, and that separate coefficients will almost always be needed for the various groups. Then there would be no point to using multiple group analysis for evaluating reliability generalization.

However, even in the absence of strong measurement invariance, relaxed versions of invariance may allow a narrow range of reliability to be exhibited across groups. Concepts such as partial measurement invariance (Byrne, Shavelson, & Muthén 1989), approximate measurement invariance as shown via fit indices (Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008), and Bayesian approximate invariance (van de Schoot et al., 2013) may provide justifications for observing high reliability generalization. This is because composite reliability may be invariant even if item parameters are not, at least in some specialized situations. As noted by Labouvie and Ruetsch (1995), the average factor loadings may be equal across groups while individual loadings vary. Similarly, sums of unique variances may be equal even if individual unique variances are unequal. In such a situation, a coefficient such as Jöreskog's (1971) coefficient for an equally weighted sum score based on a one-factor model (i.e., McDonald's 1999 ω) would remain unchanged. We conclude that multiple group modeling as usually applied is an important methodology for understanding the internal structure of items across groups, but the approach may be too strict for evaluating reliability generalization where invariance or near-invariance at the scale level is really of interest.

Another approach that might be considered is multilevel modeling. If data are obtained by a hierarchical sampling plan, typical groups can be identified by a clustering variable. If there is sufficient between-group variability, covariance matrices can be defined for between-group (Σ_B) and within-group (Σ_W) variation. Hence, latent variable modeling is possible at both levels and reliability coefficients can be computed at each level (see e.g., Geldhof, Preacher, & Zyphur, 2014). The within-group reliability coefficient may be celebrated as one that neatly describes the reliability of a scale freed from the effects of group differences.

Unfortunately, multilevel modeling is based on a very strong assumption that, in a way, imposes a desired uniformity on groups that may not really exist. The methodology typically is based on the assumption of homogeneity of within-group covariance matrices. One Σ_W is presumed to summarize within-group covariances for each of dozens or maybe hundreds of level 2 units, although in our reliability context far fewer clusters would usually be used. Nonetheless, this typically untested homogeneity assumption may be too strong—that is, reliability actually could vary by group but the model cannot detect this. While the standard analysis removes any possibility of observing heterogeneous within-group structures, Jak, Oort, and Dolan (2013) show that it is possible to detect violations of measurement invariance across clusters. Perhaps, clusters could be grouped so that reliability is homogeneous within each of those groups. But routine approaches to do this need to be worked out in the future.

Furthermore, tests of the multilevel structure that might be used to define reliability can themselves be problematic. First, it may be difficult to obtain a statistically acceptable model fit due to problems in the between-group structure which is not even of interest in this context. This can be overcome using a saturated model for Σ_B(Ryu & West, 2009) or using a two-stage model segregation approach (Schweig, 2014; Yuan & Bentler, 2007). A more serious problem is that typically one might be interested in reliability variations across only a few fixed groups, not among a large number of random groups as is required for the appropriate use of model-testing statistics in multilevel SEM (e.g., Liang & Bentler, 2004; Yuan & Bentler, 2003). Perhaps, this issue can be ignored by focusing on measures of practical fit rather than statistical fit. However, such fit indices may not be reliable either (Hsu, Kwok, Lin, & Acosta, 2015).

A problem with both multiple group and multilevel modeling approaches to understanding heterogeneity in reliability is that groups and clusters need to be discrete. It would be nice to also be able to evaluate the effects of continuous covariates on reliability. This is possible with our proposed approach, as was illustrated in our example. Of course, our approach can also be applied in multiple populations. In such a case, one could study the invariance or non-invariance not only of the overall reliability coefficient, but also of the covariate-based partition of reliability across populations. Even if overall reliability were invariant in this context, it would be possible for the covariate-based partition not to be. Furthermore, it would be interesting to see whether the relative partition into covariate-dependent and covariate-free reliability might be influenced by moderator variables.

The focus in this paper has been to study influences on reliability coefficients as traditionally defined in classical test theory. Other foci may be equally important, e.g., influences on precision as quantified by the standard error of measurement. Similarly, a methodology to obtain confidence intervals for the proposed coefficients needs to be developed, perhaps extending such methods as given by Kelley and Pornprasertmanit (2016). Such work is left to the future.

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. (2016). Specificity-enhanced reliability coefficients. *Psychological Methods*. http://dx.doi.org/10.1037/met0000092.

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, *45*, 289–308.

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.

Bentler, P. M., & Wu, E. J. C. (2015). *EQS 6.3 structural equations program*. Temple City, CA: Multivariate Software.

Beretvas, S. N., & Pastor, D. A. (2003). Using mixed-effects models in reliability generalization studies. *Educational and Psychological Measurement*, *63*, 75–95.

Bonett, D. G. (2010). Varying coefficient meta-analytic methods for alpha reliability. *Psychological Methods*, *15*, 368–385.

Botella, J., Suero, M., & Gambara, H. (2010). Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychological Methods*, *15*, 386–397.

Brannick, M. T., & Zhang, N. (2013). Bayesian meta-analysis of coefficient alpha. *Research Synthesis Methods*, *4*, 198–207.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Deng, L., & Yuan, K.-H. (2015). Multiple-group analysis for structural equation modeling with dependent samples. *Structural Equation Modeling*. doi:10.1080/10705511.2014.950534.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*, 72–91.

Guttman, L. A. (1945). A basis for analyzing test–retest reliability. *Psychometrika*, *10*, 255–282.

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.

Hsu, H.-Y., Kwok, O.-M, Jr., Lin, H., & Acosta, S. (2015). Detecting misspecified multilevel structural equation models with common fit indices: A Monte Carlo study. *Multivariate Behavioral Research*, *50*, 197–215.

Hunt, T. D., & Bentler, P. M. (2015). Quantile lower bounds to reliability based on locally optimal splits. *Psychometrika*, *80*, 182–195.

Jackson, P. H. (1979). A note on the relation between coefficient alpha and Guttman's "split-half" lower bounds. *Psychometrika*, *44*, 251–252.

Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I. Algebraic lower bounds. *Psychometrika*, *42*, 567–578.

Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling*, *20*, 265–282.

Jamshidian, M., & Bentler, P. M. (1998). A quasi-Newton method for minimum trace factor analysis. *Journal of Statistical Computation and Simulation*, *62*, 73–89.

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

Jöreskog, K. G., & Sörbom, D. G. (1996). *LISREL 8 user's guide*. Chicago: Scientific Software International.

Kaiser, H. F., & Caffrey, J. (1965). Alpha factor analysis. *Psychometrika*, *30*, 1–14.

Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods*, *21*, 69–92.

Labouvie, E., & Ruetsch, C. (1995). Testing for equivalence of measurement scales: Simple structure and metric invariance reconsidered. *Multivariate Behavioral Research*, *30*, 63–76.

Li, L., & Bentler, P. M. (2011). The greatest lower bound to reliability: Corrected and resampling estimators. *Modelling and Data Analysis*, *1*, 87–104.

Liang, J., & Bentler, P. M. (2004). An EM algorithm for fitting two-level structural equation models. *Psychometrika*, *69*, 101–122.

López-López, J. A., Botella, J., Sánchez-Meca, J., & Marín-Martínez, F. (2013). Alternatives for mixed-effects meta-regression models in the reliability generalization approach: A simulation study. *Journal of Educational and Behavioral Statistics*, *38*, 443–469.

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.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568–592.

Merkle, E. C., & Zeileis, A. (2013). Tests of measurement invariance without subgroups: A generalization of classical methods. *Psychometrika*, *78*, 59–82.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

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.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York: Routledge.

Raykov, T., & Marcoulides, G. A. (2013). Meta-analysis of scale reliability using latent variable modeling. *Structural Equation Modeling*, *20*, 338–353.

Raykov, T., Marcoulides, G. A., & Millsap, R. E. (2012). Factorial invariance in multiple populations: A multiple testing procedure. *Educational and Psychological Measurement*, *73*, 713–727.

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*, *16*, 583–601.

Sawilowsky, S. S. (2000). Psychometrics versus datametrics: Comment on Vacha-Haase's "reliability generalization" method and some EPM editorial policies. *Educational and Psychological Measurement*, *60*, 157–173.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of generalization. *Journal of Applied Psychology*, *62*, 529–540.

Schweig, J. (2014). Multilevel factor analysis by model segregation: New applications for robust test statistics. *Journal of Educational and Behavioral Statistics*, *39*, 394–422.

Shavelson, R. J., & Webb, N. (1991). *Generalizability theory: A primer*. Thousand Oaks, CA: Sage.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 229–239.

Tarkkonen, L., & Vehkalahti, K. (2005). Measurement errors in multivariate measurement scales. *Journal of Multivariate Analysis*, *96*, 172–189.

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, *54*, 837–847.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement*, *60*, 174–195.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*, 6–20.

Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development*, *44*, 159–168.

Van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers of Psychology*, *4*, 770. doi:10.3389/fpsyg.2013.00770.

Warrens, M. J. (2014). On Cronbach's alpha as the mean of all possible k-split alphas. *Advances in Statistics*, 1–5. doi:10.1155/2014/742863.

Werts, C. E., Rock, D. R., Linn, R. L., & Jöreskog, K. G. (1978). A general method of estimating the reliability of a composite. *Educational and Psychological Measurement*, *38*, 933–938.

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.

Yuan, K.-H., & Bentler, P. M. (2003). Eight test statistics for multilevel structural equation models. *Computational Statistics & Data Analysis*, *44*, 89–107.

Yuan, K.-H., & Bentler, P. M. (2007). Multilevel covariance structure analysis by fitting multiple single-level models. *Sociological Methodology*, *37*, 53–82.

Manuscript Received: 8 JUN 2015
Final Version Received: 25 JUN 2016
Published Online Date: 12 OCT 2016