Paper D04-2009

Using PROC CALIS and PROC CORR to Compare Structural Equation Modeling Based Reliability Estimates and Coefficient Alpha When Assumptions are Violated

Fei Gu, University of Kansas, Lawrence, KS
Todd Little, University of Kansas, Lawrence, KS
Neal M. Kingston, University of Kansas, Lawrence, KS

ABSTRACT

PROC CORR is widely used to calculate the Cronbach's α, and it has been described as a lower bound for test reliability. However, previous research has shown that when certain assumptions are violated coefficient α can overestimate or underestimate reliability. Raykov has shown that structural equation modeling can be used to estimate reliability. This research illustrates Raykov's SEM method in PROC CALIS and shows that under certain violations of assumptions coefficient α estimates can show a substantial positive bias in the some extreme circumstances, and the magnitude of bias of coefficient α estimates are larger than that of structural equation modeling based reliability estimates.

INTRODUCTION

Precision of measurement is one of the major concerns in social and behavioral sciences. In classical test theory (CTT), reliability is defined as the ratio of the true score variance to observed score variance (e.g., Lord & Novick, 1968). In practice, however, this ratio cannot be directly calculated as true score variance is not directly estimable. Cronbach (1951) named coefficient α and recognized its generality in estimating test reliability. Cronbach's coefficient α is one of the most frequently used estimators in social and behavioral sciences, and for more than 50 years has been the subject of considerable study (e.g., Guttman, 1953; Novick & Lewis, 1967; Lord & Novick, 1968; Zimmerman, Zumbo, & Lalonde, 1993; Miller, 1995; Komaroff, 1997; Raykov, 1997a, 1997b, 2001).

For a long time Cronbach's α has been described as a lower bound of test reliability (Zimmerman et al., 1993). Novick & Lewis (1967) showed that essential τ-equivalence and linear experimental independence are necessary and sufficient conditions for coefficient α to be equal to the test reliability. Raykov (1997a) proposed a Structural Equation Modeling (SEM) method to estimate the composite reliability for the general congeneric tests. Further, Raykov (1997b, 2001) discussed the effect of violations of essential τ-equivalence and uncorrelated errors on Cronbach's coefficient α, and analytically expressed the bias of α in terms of item parameters thereafter. Previous studies have shown that Cronbach's coefficient α may either underestimate or overestimate the test reliability (e.g., Zimmerman et al., 1993; Komaroff, 1997). That is, the violation of essential τ-equivalence tends to deflate Cronbach's coefficient α; while the violation of uncorrelated errors tends to inflate Cronbach's coefficient α. However, the analytical expression of the bias of Cronbach's coefficient α is complicated. It is quite inconvenient to interpret the effect of each individual parameter on this bias. As pointed out by Raykov (1997b), "Due to the complex form of Equation 19 as a function of the above four types of parameters, specific statements about the magnitude of alpha's population discrepancy and related conditions seem currently possible only based on a comprehensive simulation study going beyond the scope and concerns of this discussion."

In this paper, we are going to illustrate the Raykov's SEM model in SAS PROC CALIS, and compare the SEM method in estimating the test reliability with Cronbach's coefficient α under violations of essential τ-equivalence and uncorrelated errors. Cronbach's α is much easier to estimate that Raykov's SEM reliability, thus it is important to determine under what circumstances the discrepancy between Cronbach's α and Raykov's SEM reliability is of practical significance.

BASIC EQUATIONS AND DEFINITIONS

In classical test theory an observed item score is defined as the sum of the true score and error score for this item: $X_i = T_i + E_i$. Also by definition the expected value of error scores for every true score is 0: $\varepsilon(E_i \mid T_i = \tau_i) = 0$. It is typically assumed that error scores are uncorrelated (linear experimental independence) with each other ($\rho(E_i, E_j) = 0$) and with true scores ($\rho(E_i, T_j) = 0$). However it should be noted that many tests have multiple blocks of items each related to a common stimulus material (for example items nested within reading passages) that might lead to correlated errors within each block.

A test score, *X*, is defined as the sum of each item score. That is,

$$\begin{aligned} X &= X_1 + X_2 + \cdots + X_k \\ &= (T_1 + T_2 + \cdots + T_k) + (E_1 + E_2 + \cdots + E_k) \end{aligned} \tag{1}$$

Test reliability is defined as

$$\rho_x = \frac{Var(T)}{Var(X)} = \frac{Var(T_1 + T_2 + \cdots + T_k)}{Var(X_1 + X_2 + \cdots + X_k)} = \frac{Var(T_1 + T_2 + \cdots + T_k)}{Var(T_1 + T_2 + \cdots + T_k + E_1 + E_2 + \cdots + E_k)} \tag{2}$$

Cronbach's coefficient α is defined as

$$\alpha = \frac{k}{(k-1)}\left[1 - \frac{\sum Var(X_i)}{Var(X)}\right] \tag{3}$$

As discussed by Raykov (1997a):

> Assume that the components $X_1, X_2, \ldots, X_k$ are congeneric. Formally, this means that for any pair of measures X_i and X_j there exists a linear relationship between their true scores T_i and T_j; that is, for any pair X_i and X_j there exist values a_{ij} and b_{ij}, such that $T_i = a_{ij} + b_{ij}T_j$ ($1 \le i, j \le k$; e.g., Jöreskog, 1971). In particular, without generality, for j = 1 and any $i > 1$ there exist a pair of values a_{i1} and b_{i1}, such that $T_i = a_{i1} + b_{i1}T_j$ (the second subindexes of *a* and *b* can therefore be dropped). The

well-known model of parallel tests is a special case of this, if the following assumptions are satisfied: (1) $a_1 = \ldots = a_k = 0$; (2) $b_1 = \ldots = b_k = 1$; and (3) $\mathrm{Var}(E_1) = \ldots = \mathrm{Var}(E_k)$. Similarly, τ-equivalent tests are obtained as another special case of congeneric tests when Assumption 3 is dropped; in this case error variances are permitted to differ. Also, essentially τ-equivalent measures result as a special case of congeneric tests when Assumption 1 and 3 are dropped; that is, when nonzero intercepts are allowed for τ-equivalent tests (e.g., Lord & Novick, 1968).

Therefore, the reliability for a congeneric test is herein further defined as

$$\rho_x = \frac{Var(T_1 + T_2 + \cdots + T_k)}{Var(T_1 + T_2 + \cdots + T_k + E_1 + E_2 + \cdots + E_k)} \tag{4}$$

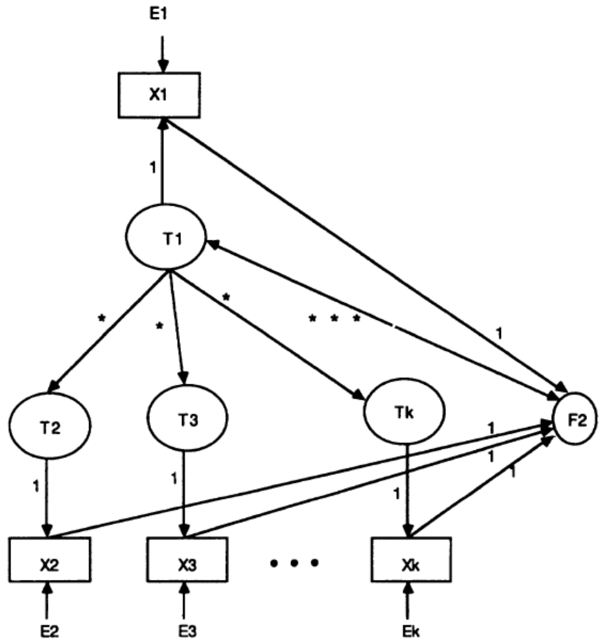
$$= \frac{\left(\sum b_i\right)^2 Var(T_1)}{\left(\sum b_i\right)^2 Var(T_1) + \sum Var(E_i) + \sum_{1 \le i \ne j \le k} Cov(E_i E_j)} \tag{5}$$

Though Raykov (1997a, p. 175, equation 6) did not include correlated error scores in the derivation of the SEM method, it is easy to generalize his equation.

$$\begin{aligned}
\rho(T_1, \sum X_i) &= \frac{Cov(T_1, \sum X_i)}{\sqrt{Var(T_1) Var(\sum X_i)}} \\
&= \frac{\left(\sum b_i\right) Var(T_1)}{\sqrt{Var(T_1) Var(\sum a_i + T_1 \sum b_i + \sum E_i)}} \\
&= \frac{\left(\sum b_i\right) \sqrt{Var(T_1)}}{\sqrt{\left(\sum b_i\right)^2 Var(T_1) + Var(\sum E_i)}} \\
&= \frac{\left(\sum b_i\right) \sqrt{Var(T_1)}}{\sqrt{\left(\sum b_i\right)^2 Var(T_1) + \sum Var(E_i) + \sum_{1 \le i \ne j \le k} Cov(E_i E_j)}} \\
&= \sqrt{\rho_x}
\end{aligned} \tag{6}$$

As Raykov (1997a) pointed out, if $Var(T_1) \neq 0$, the correlation of T_1 and the composite score X is the concluding reliability index, and $\rho(T_1, \sum X_i)$ can be obtained as an "external parameter" of the model presented in the path diagram shown in Figure 1.

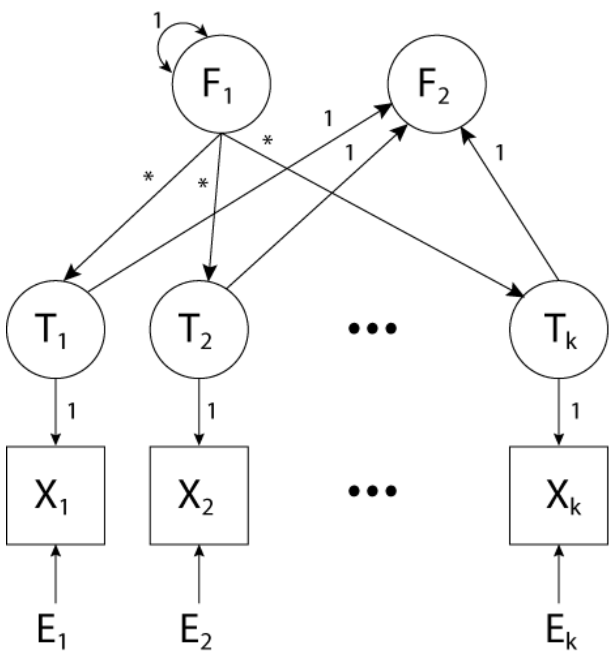
Figure 1
A Structural Equation Model for Estimation of Composite Reliability



Source: Raykov (1997a)

Figure 2 presents an alternative specification of the same path model and was used in this research. In this structural equation model F_1 is the common latent variable, and F_2 is the phantom variable.

Figure 2: Alternative SEM model to estimate reliability



METHODS

Five factors have been shown to influence the estimated reliability of a test: common true score variance, constructing loadings on the common true score, error score variances, error score covariance, and test length (Raykov, 2001). To model the context of error covariance and non-essential-τ-equivalence, this study simulates a 15 item test divided into three sets of five items each. These 15 items have correlated error within each set but not across each set. To avoid the confounding of error correlation and error variance, both are independently modeled in this study. Further, correlation is assumed equal for every pair of error scores, and error variances are also assumed equal.

SIMULATED FACTORS

For each simulated test the number of items violating the assumption of essential τ-equivalence is 6. This will be done evenly across the three blocks of items in each test. That is, there will be 2 items violating this assumption within each block. Item scores are generated under the constraint that each item will have unit variance. Simulated factors are described below:

First, the ratio of true score variance to error score variance is controlled on each item. There are 3 levels in this condition: 1:9, 5:5, 9:1. That is, the true score variance can be 0.1, 0.5, or 0.9 while the corresponding error score variance is 0.9, 0.5, and 0.1.

Second, the degree to which the essential τ-equivalence is violated through the loading shrinkage from true score to observed score on each item. There are 3 levels in this condition: no shrinkage, small shrinkage, and big shrinkage.

Third, the pairwise error correlation among regular items is controlled as well. The value of this correlation is 0.0, 0.2, and 0.4. When the correlation equals to 0.0, the uncorrelated error assumption is met.

These steps create 27 ($3\times3\times3$) conditions. Each condition is replicated 2000 times. For each replication 1000 examinees are simulated.

ANALYSIS

If the error correlation is simulated to be zero, the SEM model is represented by the path diagram in Figure 3. If the error correlation is simulated to be nonzero, parameters that represent error correlation should be added into Raykov's (1997a) SEM model. Instead of adding covariance/correlation between all pairwise true scores, we choose to use a method factor in each subset to capture the unwanted error correlation. The path diagram in Figure 4 represents the modified SEM model.

Figure 3: SEM model with uncorrelated errors

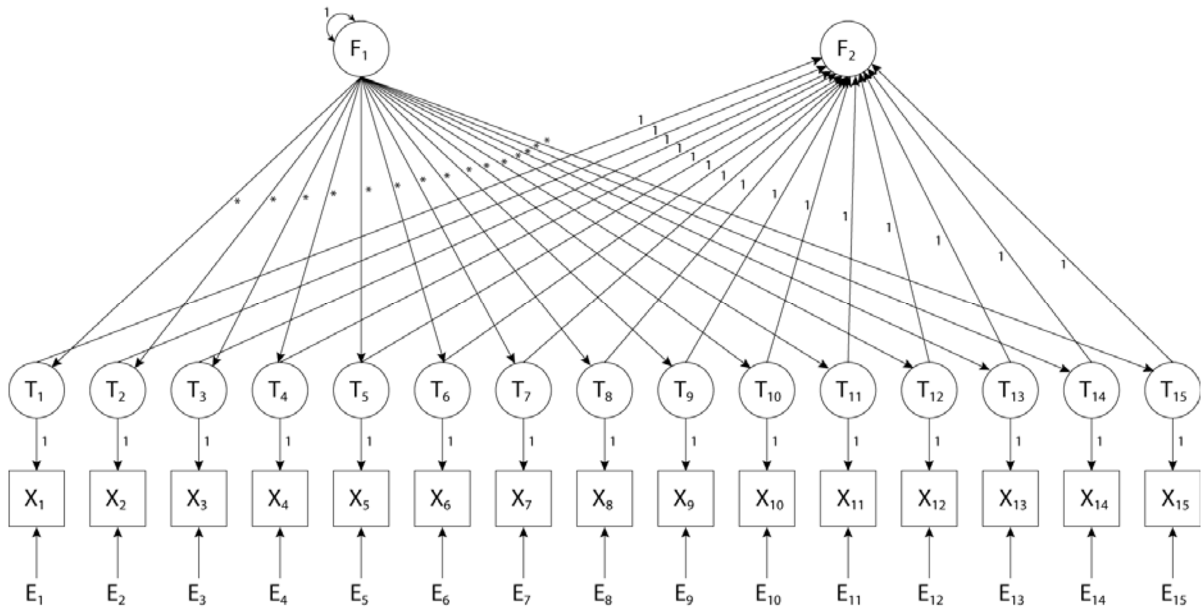
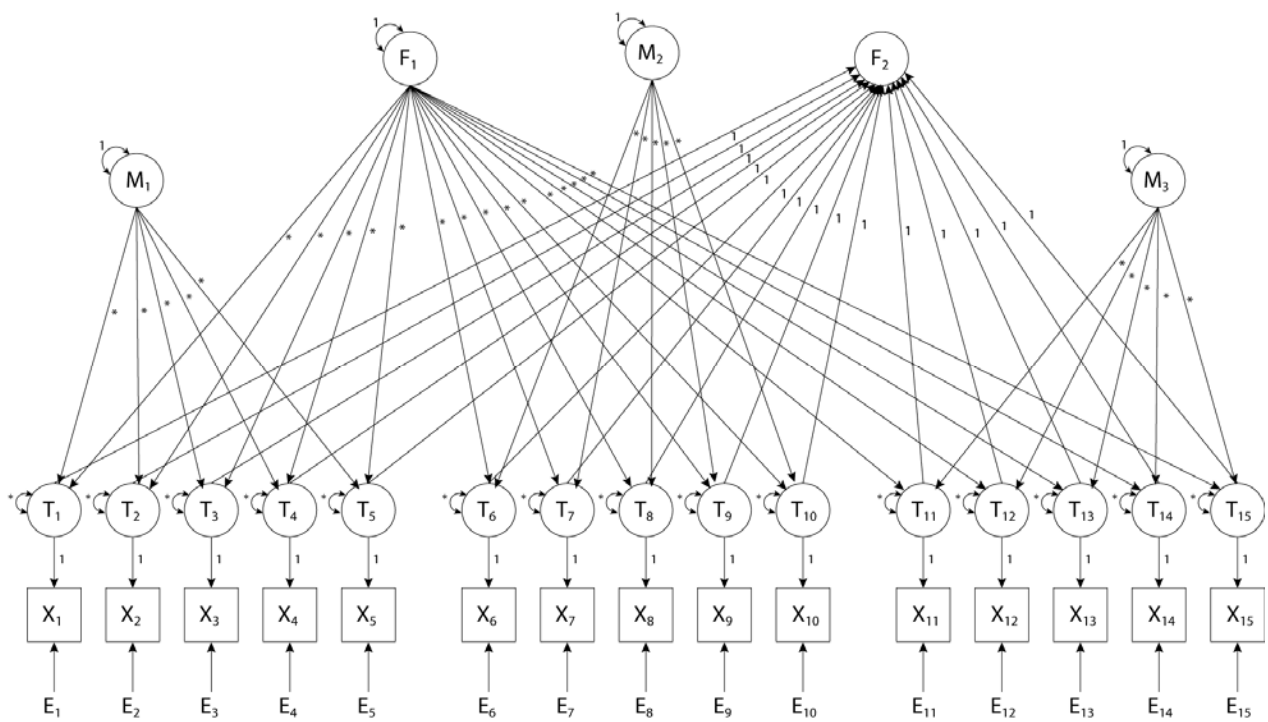


Figure 4: SEM model when correlated errors



All the work is done on a 64-bit cluster machine due to the extremely intensive computation required by this simulation. SAS (cluster version 9.2) PROC CORR is used to compute Cronbach's α. PROC CALIS is used to fit Raykov's SEM model. For certain trials, the optimization process required more iterations than the default setting. To this end the maximum number of iterations and maximum number of function calls were both set to 10000. The starting

value of 0.7 is provided to PROC CALIS for all parameter estimates.
Example SAS program for Figure 3 is such:

```
 PROC CALIS COV data=observed PLATCOV maxiter=10000 maxfunc=10000 inest=start;
LINEQS X1  =      F1    + E1,
         X2  =      F2    + E2,
         X3  =      F3    + E3,
         X4  =      F4    + E4,
         X5  =      F5    + E5,
         X6  =      F6    + E6,
         X7  =      F7    + E7,
         X8  =      F8    + E8,
         X9  =      F9    + E9,
         X10 =      F10   + E10,
         X11 =      F11   + E11,
         X12 =      F12   + E12,
         X13 =      F13   + E13,
         X14 =      F14   + E14,
         X15 =      F15   + E15,
         F1  = BE1  F_COM + D1,
         F2  = BE2  F_COM + D2,
         F3  = BE3  F_COM + D3,
         F4  = BE4  F_COM + D4,
         F5  = BE5  F_COM + D5,
         F6  = BE6  F_COM + D6,
         F7  = BE7  F_COM + D7,
         F8  = BE8  F_COM + D8,
         F9  = BE9  F_COM + D9,
         F10 = BE10 F_COM + D10,
         F11 = BE11 F_COM + D11,
         F12 = BE12 F_COM + D12,
         F13 = BE13 F_COM + D13,
         F14 = BE14 F_COM + D14,
         F15 = BE15 F_COM + D15,
      F_PHAN = F1  + F2  + F3  + F4  + F5
             + F6  + F7  + F8  + F9  + F10
             + F11 + F12 + F13 + F14 + F15;
   STD D1-D15 = PSI1-PSI15,
        F_COM = 1;
RUN;
```

Example SAS program for Figure 4 is such:

```
 PROC CALIS COV data=observed PLATCOV maxiter=10000 maxfunc=10000 inest=start;
  LINEQS X1  =      F1    + E1,
         X2  =      F2    + E2,
         X3  =      F3    + E3,
         X4  =      F4    + E4,
         X5  =      F5    + E5,
         X6  =      F6    + E6,
         X7  =      F7    + E7,
         X8  =      F8    + E8,
         X9  =      F9    + E9,
         X10 =      F10   + E10,
         X11 =      F11   + E11,
         X12 =      F12   + E12,
         X13 =      F13   + E13,
         X14 =      F14   + E14,
         X15 =      F15   + E15,
         F1  = BE1  F_COM + BE16 F_M1 + D1,
         F2  = BE2  F_COM + BE17 F_M1 + D2,
         F3  = BE3  F_COM + BE18 F_M1 + D3,
         F4  = BE4  F_COM + BE19 F_M1 + D4,
         F5  = BE5  F_COM + BE20 F_M1 + D5,
         F6  = BE6  F_COM + BE21 F_M2 + D6,
         F7  = BE7  F_COM + BE22 F_M2 + D7,
```

```
            F8  = BE8  F_COM + BE23 F_M2 + D8,
            F9  = BE9  F_COM + BE24 F_M2 + D9,
            F10 = BE10 F_COM + BE25 F_M2 + D10,
            F11 = BE11 F_COM + BE26 F_M3 + D11,
            F12 = BE12 F_COM + BE27 F_M3 + D12,
            F13 = BE13 F_COM + BE28 F_M3 + D13,
            F14 = BE14 F_COM + BE29 F_M3 + D14,
            F15 = BE15 F_COM + BE30 F_M3 + D15,
         F_PHAN = F1  + F2  + F3  + F4  + F5
                + F6  + F7  + F8  + F9  + F10
                + F11 + F12 + F13 + F14 + F15;
     STD D1-D15 = PSI1-PSI15,
          F_COM = 1,
          F_M1-F_M3 = 3*1;
RUN;
```

RESULTS

In order to compare the reliability estimates from Cronbach's α and SEM, the estimates are averaged across replications in every combination of conditions. Then, the averaged values are compared to the true values in every combination of conditions, which is calculated according to Equation 5.

Table 1 lists the bias of Cronbach's α and SEM estimates. First, SEM estimates and Cronbach's α give virtually identical results in conditions where both assumptions are met, which are consistent with the analytical proof. Second, SEM estimates and Cronbach's α show similar magnitude of bias when and only when the ratio of true score variance to error score variance is 9:1. Third, the mean of SEM estimates are all negatively biased, and it seems that only the ratio of true score variance to error score variance affects the bias of SEM estimates. Fourth, the bias of Cronbach's α tends to increase when the error correlation increases and when the degree of violation on the essential τ-equivalence increases; while it tends to decrease when the ratio of the true to error variance increases.

Table 1: 6 items violate the τ-equivalence

		Ratio of True to Error Variance					
		1:9		5:5		9:1	
loading change	**Corr. among Errors**	alpha-ρ	SEM-ρ	alpha-ρ	SEM-ρ	alpha-ρ	SEM-ρ
1.0	**0**	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.18	-0.06	0.03	-0.02	0.00	0.00
	0.4	0.31	-0.09	0.06	-0.04	0.01	-0.01
0.7	**0**	0.00	0.00	0.00	0.00	0.00	0.00
	0.2	0.20	-0.06	0.04	-0.03	0.01	-0.01
	0.4	0.34	-0.09	0.07	-0.05	0.01	-0.01
0.3	**0**	0.00	0.00	0.00	0.00	-0.01	0.00
	0.2	0.23	-0.06	0.05	-0.03	0.01	-0.01
	0.4	0.38	-0.08	0.10	-0.06	0.02	-0.02

DISCUSSION

When the assumptions of classical test theory are met coefficient α provides a good estimate of true reliability. When the assumptions are violated, in particular when errors are correlated in the way modeled in this study or τ-equivalence of items is violated with some items having relatively low loadings on true score, coefficient α can significantly *overestimate* reliability. SEM based reliability estimates also face bias problems, but to a much lesser degree. However, it is difficult and time-consuming to build an appropriate SEM model for the real data, and large sample size is often required for the maximum likelihood estimation method. It is suggested that researchers should calculate the reliability using both methods (if possible) and interpret the results with caution. Conditions in this study are very limited, and a more comprehensive simulation study is warranted in future.

REFERENCES

- Cronbach. L. J. (1951). Coefficient Alpha and the Internal Structure of a Test. *Psychometrika*, 16(3), 297-334.
- Kaiser, H. F. & Dickman, K. (1962). Sample and Population Score Matrices and Sample Correlation Matrices from an Arbitrary Population Correlation Matrix. *Psychometrika*, 27(2), 179-182.
- Komaroff, E. (1997). Effect of Simultaneous Violations of Essential τ-Equivalence and Uncorrelated Error on Coefficient α. *Applied Psychological Measurement*, 21(4), 337-348.
- Guttman, L. (1953). Reliability Formulas That Do Not Assume Experimental Independence. *Psychometrika*, 18(3), 225-239.
- Lord, F, M. & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading MA: Addison-Wesley.
- Miller, M. B. Coefficient Alpha: A Basic Introduction from the Perspectives of Classical Test Theory and Structural Equation Modeling. *Structural Equation Modeling*, 2(3), 255-273.
- Novick, M. R. & Lewis, C. (1967). Coefficient Alpha and the Reliability of Composite Measurements. *Psychometrika*, 32(1), 1-13.
- Raykov, T. (1997a). Estimation of Composite Reliability for Congeneric Measures. *Applied Psychological Measurement*, 21(2), 173-184.
- Raykov, T. (1997b). Scale Reliability, Cronbach's Coefficient Alpha, and Violations of Essential Tau-Equivalence with Fixed Congeneric Components. *Multivariate Behavioral Research*, 32(4), 329-353.
- Raykov, T. (2001). Bias of Coefficient α for Fixed Congeneric Measure with Correlated Errors. *Applied Psychological Measurement*, 25(1), 69-76.
- Raykov, T. & Shrout, P.E. (2002). Reliability of Scales With General Structure: Point and Interval Estimation Using a Structural Equation Modeling Approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 195-212.
- SAS Institute, Inc. (2002), SAS version 9, Online Help, Cary, NC.
- Zimmerman, D. W., Zumbo, B. D., & Lalonde, C. (1993). Coefficient Alpha as An Estimate of Test Reliability under Violation of Two Assumptions. *Educational and Psychological Measurement*, 53, 33-49.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Fei Gu
1122 W. Campus Rd.
Joseph R. Pearson Hall, Room 621
University of Kansas
Lawrence, KS 66045-3101
Phone: 785-864- 9705
Fax: 785-864-3820
E-mail: fgu@ku.edu