Use of SEM Programs to Precisely Measure Scale Reliability

Yutaka KANO[1] and Yukari AZUMA[2]

[1] School of Human Sciences, Osaka University, Suita, Osaka 565-0871, Japan
kano@hus.osaka-u.ac.jp
[2] School of Human Sciences, Osaka University, Suita, Osaka 565-0871, Japan
azuma@koko15.hus.osaka-u.ac.jp

Summary. It is first pointed out that most often used reliability coefficient α and one-factor model based reliability ρ are seriously biased when unique factors are covariated. In the case, the α is no longer a lower bound of the true reliability. Use of Bollen's formula (Bollen 1980) on reliability is highly recommended. A web-based program termed "STERA" is developed which can make *stepwise* reliability analysis very easily with the help of factor analysis and structural equation modeling.

Keywords: Cronbach's coefficient α, one-factor model based reliability, unique factor covariance, web-based program STERA, structural equation modeling, LM test.

1 Introduction

Reliability analysis has been discussed extensively since Cronbach (1951) proposed the famous reliability coefficient α, and the discussion recently has been made within the scope of the structural equation modeling (e.g., Raykov 2001; Hancock & Mueller, 2000; Green & Hershberger, 2000; Komaroff, 1997). A recent version of EQS (EQS6.0, Bentler 2002) can print a variety of scale reliability estimates. Although reliability analysis is an old topic, it is still given much attention.

Consider a one-factor model with possibly covariated[1] unique factors:

$$X_i = \mu_i + \lambda_i f + u_i \qquad (i = 1, \ldots, p), \tag{1}$$

where $\mu_i = E(X_i)$, λ_i being a factor loading parameter, $E(f) = E(u_i) = 0$, $V(f) = 1$, $\mathrm{Cov}(u_i, u_j) = \psi_{ij}$ and $\mathrm{Cov}(f, u_i) = 0$. Here f and u_i are called a (common) factor and a unique factor, respectively.

The scale score is defined as the total sum of X_i, i.e., $X = \sum_{i=1}^{p} X_i$. The scale reliability ρ' of X is then defined as the proportion of the true score variance to the total variance, that is,

[1] The term "correlated" is more often used in this context. Without a few exceptions, we use "covariated" in this paper since we mainly focus upon covariances between unique factors rather than correlations. Mathematically both of the terminologies are equivalent.

$$\rho' = \frac{V(\sum_{i=1}^{p}\lambda_i f)}{V(X)} = \frac{(\sum_{i=1}^{p}\lambda_i)^2}{(\sum_{i=1}^{p}\lambda_i)^2 + \sum_{i=1}^{p}\psi_{ii} + \sum_{i,j,i\neq j}^{p}\psi_{ij}}. \quad (2)$$

On the other hand the traditional reliability (test) theory assumes $\psi_{ij} = 0$ for $i \neq j$, so that the ρ' reduces to

$$\rho = \frac{(\sum_{i=1}^{p}\lambda_i)^2}{(\sum_{i=1}^{p}\lambda_i)^2 + \sum_{i=1}^{p}\psi_{ii}}. \quad (3)$$

The no-covariance assumption may not hold for many empirical data sets, and the recent literature focuses on effects of the unique factor covariances upon the traditional reliability measure Cronbach's α.

Bollen (1980) would be the first work that points out the importance of nonzero unique factor covariances in reliability analysis and derives the formula (2). He employed a confirmatory factor analysis model to develop a scale of political democracy, and found that the assumption of uncorrelatedness of the unique factors in the scale is inappropriate. He developed a confirmatory factor analysis model with covariated unique factors for the scale. The formula (2) is identical with Bollen's (Bollen 1980, formula (1) on page 378). Bollen's analysis (Bollen 1980, Figure A3) is reproduced in Fig.1.

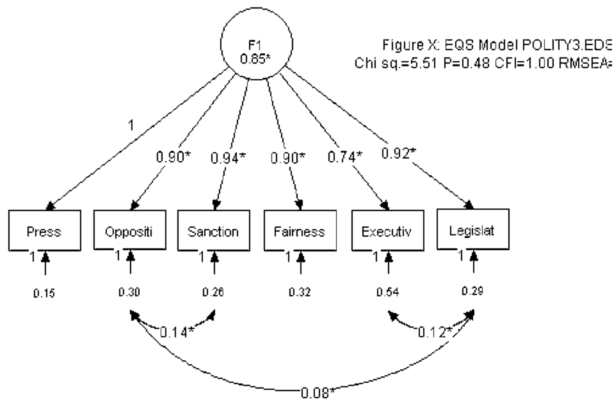


Fig. 1. Bollen's analysis: Political democracy scale

If there are many pairs of covariated unique factors, there may be additional common factors that can account for the covariances. Practitioners, then, consider that the scale is not unidimensional, extract multiple common factors and develop subscales. The problem is how to do it when there still remain covariances that can not be explained by the multiple factors, even after some subscales were developed. A typical case is where there is a common factor with only two indicators (Kano 1997) within a single (sub)scale and the factor is not interpretable or does not interest the practioners. They can then use the model (1) to estimate the reliability through the formula (2). The model could also be used when they are not interested in additional factors even if subscales are to be developed.

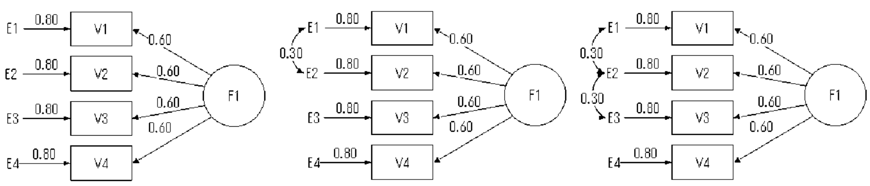
In this paper, we begin by illustrating how seriously covariated unique factors invalidate classical reliability coefficients α and ρ with $\psi_{ij} = 0$. In Section 3, we take Lagrange Multiplier (LM) tests to determine the pairs of unique factors to be covariated, and show a newly developed program called "STERA." We finally end with concluding remarks in Section 4.

2 Impacts of covariated unique factors

A most often used reliability coefficient in social sciences is neither ρ nor ρ' defined in (3) and (2) but Cronbach's coefficient α defined as

$$\alpha = \frac{p}{p-1}\left(\frac{\sum_{i\neq j}^{p}\mathrm{Cov}(X_i, X_j)}{V(X)}\right) \tag{4}$$

Under the assumption of a one-factor model without covariances between unique factors, it is known that $\alpha \leq \rho$, and that the equality holds when items are essentially τ-equivalent (or weakly parallel measurement), i.e., λ_i's are the same. Thus, the coefficient α is a conservative measure of the true reliability, and some researchers prefer the α because of the conservativeness. Here, a question arises as to what if the independence assumption on unique factors fails. What kind of influence is made on the ρ and α? An appealing example is provided in Fig.2.



	model 1	model 2	model 3
α :	0.69	0.74	0.77
ρ :	0.69	0.74	0.78
ρ' :	0.69	0.65	0.61

Fig. 2. Covariated unique factors and reliability coefficients

The model 1 in Fig.2 is a basic one-factor model with equal factor loadings of 0.6. The models 2 and 3 are those with unique factor covariances of 0.3

for one and two pairs of unique factors. We reported just below the path diagrams values of α, ρ and ρ' respectively defined in (4), (3) and (2). For computing ρ, we first calculated correlation matrices derived from the models and made one-factor analysis of the correlation matrices so that factor loading and unique factor variance estimates were obtained for each model.[2] The reliability was evaluated based on (3) using these estimates. Thus, the values of ρ are wrong reliability estimates that result from ignoring existing unique factor covariances.

The model 1 meets the essentially τ-equivalent assumption, and hence the α and ρ, identical with ρ', represent the true reliability as 0.69.

It is seen from the results in the models 1 and 2 that since ρ' gives the correct reliability value, α and ρ result in seriously biased estimates, and that more unique factor covariances result in more serious bias. The α coefficient is no longer a conservative estimate for the true reliability. This observation coincides with Zimmerman et. al. (1993) and Komaroff (1997). When negative covariances appear between unique factors, α and ρ underestimate the reliability. Thus, existence of unique factor covariances invalidates use of α and ρ.

It is easy to explain why such biases are made. Unique factor covariances contribute to the covariances between observed variables. The α is a monotonically increasing function of the total covariances, so positive unique factor covariances lead to a positive bias. A similar explanation can be made for ρ. When the unique factor covariances are ignored, factor loading estimates are seriously overestimated, leading to overestimation of ρ. On the other hand, when we use ρ', unique factor covariances contribute to the denominator of the formulae, and the covariances are regarded as an error. Usually the covariances of unique factors do not interest us, so the formula ρ' is reasonable.

One can examine the fit of a one-factor model employed to detect whether unique factor covariances should be introduced. A bad-fitted one-factor model may indicate existence of unique factor covariances. (Exploratory) factor analysis without model fit examination can cause an inflated or deflated reliability estimate, so it is not recommended.

3 LM approach and STERA

In the previous section, we pointed out the importance of introduction of unique factor covariances. Here, we discuss a pragmatic estimation method of the reliability based on the expression ρ' in (2). A drawback of the model (1) is nonidentifiability, so that one cannot estimate parameters. The ψ_{ij} has $p(p+1)/2$ parameters, which is the same as the number of variances and covariances of the observed variables.

[2] Obviously, the models are fitted very poorly.

F	Var	Factor Loading F1	FL@	ALPHA	RHO	Rho@	ALPHA 1-variable deletion	RHO 1-variable deletion	Test of Goodness-of-fit
F1	v3	0.901	0.901	0.934	0.935	0.935	0.917	0.916	X²=38.281 df=9 P=0.000
	v2	0.895	0.895				0.919	0.917	
	v1	0.885	0.885				0.915	0.918	
	v6	0.874	0.874				0.916	0.920	
	v4	0.789	0.789				0.927	0.929	
	v5	0.686	0.686				0.939	0.939	

Fig. 3. Preliminary exploratory factor analysis

One way to estimate ψ_{ij} is to use the residual covariance matrix, that is, $\hat{\psi}_{ij} = s_{ij} - \hat{\lambda}_i\hat{\lambda}_j \quad (i \neq j)$, where s_{ij} is the sample covariance between X_i and X_j, and $\hat{\lambda}_i$ is an estimate in the usual one-factor analysis model, i.e., the model with $\psi_{ij} = 0$ $(i \neq j)$. The approach, however, does not work because $\sum_{i,j,i\neq j}^{p} \hat{\psi}_{ij}$ is almost zero almost always. Estimation process tries to minimize the residuals and in many cases, this also minimizes the sum of residuals as in regression analysis.

Here we propose that one perform Lagrange Multiplier tests for unique factor covariances sequentially, as suggested by Raykov (2001). For this, the SEM program EQS is useful (Bentler 1995). The LM option of the EQS as `/LMTEST` with `SET= PEE;` gives a list of the pairs of unique factors to be covariated. A covariance parameter between a statistically significant pair is released to be a free parameter, and the model is reestimated.

Factor1		
Check the box if an error covariance is allowed.		
	Error Covariance	Predicted Chi-square
☑	E2 , E3	18.479
☐	E4 , E1	25.922
☐	E5 , E6	30.285
☐	E4 , E2	30.554
☐	E5 , E2	30.842
☐	E6 , E3	33.542
☐	E1 , E2	35.675
☐	E4 , E3	36.362
☐	E5 , E4	37.185
☐	E6 , E2	37.443

Fig. 4. Unique factor covariances to be allowed

It is a feasible process but very tedious. In addition, one has to calculate values of ρ' by him/herself. We have developed a web-based program that can easily implement such reliability analysis. The program is called STERA (STEpwise Reliability Analysis). Any one who can access internet can use this program.

You are requested to input a sample correlation matrix, the number of variables, the number of factors, and sample size when you access the top page of the STERA.

Factor1

Reliability analysis when error covariances may be allowed

Error Covariance Estimates

E2 , E3 = 0.123

Vars	Factor Loading	RHO'	Test of Goodness-of-fit
V1	0.900	0.918	χ²=21.590 df=8 P=0.006
V6	0.879		
V3	0.861		
V2	0.852		
V4	0.813		
V5	0.706		

LM test for error covariance: Predicted chi-square statistic for goodness-of-fit when an error covariance is allowed

Check the box if an error covariance is allowed.

	Error Covariance	Predicted Chi-square
☐	E6 , E2	13.642
☐	E4 , E1	14.757
☐	E6 , E1	15.201
☐	E5 , E6	16.126
☐	E5 , E2	18.235
☐	E4 , E2	18.849
☐	E6 , E3	19.853
☐	E4 , E6	20.428
☐	E5 , E1	20.662
☐	E1 , E3	20.891

Fig. 5. One unique covariance is introduced

We took Bollen's political democracy data (Bollen 1980, Table 2, $n = 113$) to demonstrate the program. When you submit a job after giving necessary information, you will receive an output webpage as in Fig.3, where results of exploratory factor analysis are presented including model fit information.

When you submit a job of multiple factor analysis, the program presents reordered factor loading estimates, and reliability analysis will proceed for each factor. Factor analysis with one common factor has been made for each set of observed variables with each factor, and reliability estimates are presented based on both one-factor and multiple factor analysis results. To distinguish, estimates based on one-factor analysis are attached with the symbol "@" as in Fig.3. Since this example is a one-factor analysis, both of the results are identical with each other. Values of α and ρ are estimated as 0.934 and 0.935. The goodness-of-fit statistic indicates rejection of the model ($\chi^2 = 38.281$ df$= 9$, p value$= 0.000$).

Below the table in Fig.3, you find anther table (Fig.4) which shows which pair of the unique factors should be covariated from the statistical point of view. The table indicates a predicted chi-square statistic of the model if the unique covariance of the pair is allowed. In the example, you check a box in the pair of `E2` and `E3` and submit the job to get model fit information of the revised one, reliability estimates and a further analysis of unique factor covariances. Those are presented in Fig.5. The covariance is estimated as 0.123 and the reliability estimate $\hat{\rho}'$ is given as 0.918, which is a bit lower than the original one. However, again the new model receives a poor fit (p value$= 0.006$), so you need to introduce one more covariance. For this, you can check the box of E6 and E2. You will be then noticed that you need to allow the covariance between E5 and E6. If you add the covariance, then you will reach a final accepted model, where the model is fitted satisfactorily (p value= 0.480). The model is identical with Bollen's (Fig.1).

The reliability coefficient ρ' estimated based on the model in Fig.1 is 0.907, which is lower than $\hat{\rho} = 0.935$ or $\hat{\alpha} = 0.934$.

4 Concluding remarks

Many text books (e.g., Allen and Yen, 1979) on measurement theory alert that the coefficient α is nothing but a lower bound for true reliability if essential τ equivalence assumption (i.e., item homogeneity; λ_i's are the same) is violated. Less attention on the independence assumption on unique factors has been paid, as noted by Green and Hershberger (2000). We feel, however, that the independence assumption is more important than the essential τ equivalence assumption because dependency among unique factors can cause overestimation of the true reliability.

It is important to examine goodness-of-fit of a factor analysis model to detect substantial unique factor covariances. When the model receives a poor fit and unique factor covariances are considered as a possible cause, an LM test is useful to determine which pair of the unique factor be covariated.

There are many situations, already studied in the literature, which cause unique factor covariances. According to Rozeboom (1966), speeded tests can introduce covariated unique factors so that the coefficient α cannot be used. He also noted that when items on a test are administrated on a single occasion, errors among items are likely to be positively correlated. Green and Hershberger (2000) made an attempt to model error covariances in true score models. There may be a method factor that influences some of the items, so that unique factors of the items influenced can be covariated. There may be a third variable (confounding variable) that is not noticed by the researcher but can influence some of the items.

There are also situations where the error covariances are a source of reliable and repeatable variance that has a similar interpretation to that of the true scores. In the case, one could consider that the error covariances are to be added to the variance of the true scores. We have not discussed this case fully in the paper.

It is said that in the context of structural equation modeling, one should not introduce error covariances (to improve a model fit) without substantial consideration, see, e.g., Browne (1982). The LM option of EQS as default does not print results on error covariances. Thus, error covariances are allowed only if adequate reasons are given to the introduction of the covariances. We feel that in reliability analysis one should be more optimistic for introducing unique factor covariances because inflated α or inflated one-factor based ρ can prove a wrong validation of the scale constructed.

Alternative approaches could be taken to estimate reliability for a case where unique factor covariances appear. Vautier (2001) proposed an alternative way of identifying pairs of nonzero unique factor covariances in which he calls it a heuristic shifting method. A stepwise variable selection in factor analysis (Kano & Harada, 2000) could be used to select a set of variables that perfectly conforms to a factor analysis model and use the coefficient ρ.

In this paper, we have discussed a feasible procedure of precisely estimating the true reliability ρ' defined in (2) of a scale and developed an easily accessible program STERA.[3] We hope many readers will access the program to correctly evaluate reliability of scales they use.

References

Allen, M. J. & Yen, W. M. (1979). Introduction to Measurement Theory. Brooks/Cole, Monterey, CA.

Bentler, P. M.(1995, 2002): EQS Structural Equations Program Manual. Multivariate Software, Los Angeles.

Bollen, K. A. (1980): Issues in the comparative measurement of political democracy. Amer. Sociol. Rev., **45**, 370–390.

Browne, M. W.(1982) Covariance structures. In: Hawkins, D. M. (ed.), Topics in applied multivariate analysis (pp.72-141). Cambridge University Press, Cambridge.

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

Green, S. B. & Hershberger, S. L. (2000): Correlated errors in true score models and their effect on coefficient alpha. Structural Equation Modeling, **7**, 251-270.

Hancock, G. & Mueller, R. (2000): Rethinking construct reliability within latent variable systems. In: Cukeck, R., Toit, D., & Söbom, D.(Eds.), Structural equation modeling: Present and future (pp.195-216). SSI, Chicago.

Kano, Y. (1997): Exploratory factor analysis with a common factor with two indicators. Behaviormetrika, **24**, 129–145.

Kano, Y. & Harada, A. (2000): Stepwise variable selection in factor analysis. Psychometrika, **65**, 7–22.

Komaroff, E. (1997): Effect of simultaneous violations of essential tau-equivalence and uncorrelated error on coefficient alpha. App. Psychol. Measurement, **21**, 337-348.

Raykov, T. (2001): Bias of Cronbach's coefficient alpha for fixed congeneric measures with correlated errors. App. Psychol. Measurement, **26**, 69–76.

Rozeboom, W. W. (1966): Foundation of the Theory of Prediction. The Dorsey Press, Homewood, IL.

Vautier, S. (2001). Assessing internal consistency reliability with the Jöreskog's Rho: New developments. (submitted)

Zimmerman, D. W., Zumbo, B. D. & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. Educ. Psychol. Measurement, **53**, 33-49.

[3] http://koko16.hus.osaka-u.ac.jp/stepwise/servlet/stera/