An Alternative Procedure for Assessing Convergent and Discriminant Validity

Donald R. Lehmann
Columbia University

This paper presents a simple procedure for establishing convergent and discriminant validity. The method uses ordinary least-squares regression (OLS) with the correlations between measures as the dependent variable. Illustrations indicate that the method provides reasonable answers. *Index terms: convergent validity, discriminant validity, multiple regression.*

Establishing the validity of measures is a major focus of research. Essential to establishing validity with multi-item measures are notions of convergent and discriminant validity (Anastasi, 1968; Bohrnstedt, 1970; Nunnally, 1978). Many approaches to convergent and discriminant validity assessment are derived from the multitrait-multimethod approach (Campbell & Fiske, 1959). This paper presents an alternative to the LISREL-based nested model procedure used by Bagozzi (1978) and Widaman (1985).

Background

The most common approach to establishing convergent and discriminant validity is to demonstrate that multiple measures of a construct are (1) related, and (2) more related to each other than to measures of other constructs, even when the two measurement methods are similar (Campbell & Fiske, 1959). The second criterion suggests that a construct should have more impact on a relation than does a common measurement method.

The assessment of validity typically has focused on the Pearson product-moment correlation coefficient. This requires both intervally scaled data and linear association, although a monotonic relation will often suffice. Although it is relatively easy to account for non-linearities in the relation (Lehmann, 1974), here the focus is on the linear correlation coefficient as the measure of association.

Convergent and discriminant validity require that multiple measures of a construct relate highly to each other and less highly to measures of other constructs. Consider the hypothetical two-construct, three-measure-per-construct model in Figure 1. In the case of independent constructs (Figure 1a), correlations of variable pairs (1,2), (1,3), and (2,3) should be high (Block I in Figure 2). Similarly, correlations of variable pairs (4,5), (4,6), and (5,6) should be high (Block III in Figure 2). By contrast, correlations of variable pairs in Block II such as (1,4) should be zero.
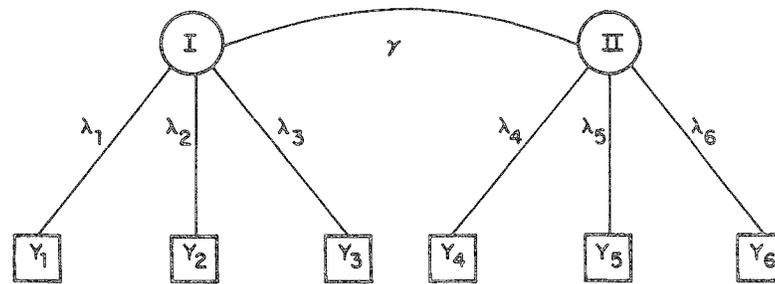
When some measures share a common measurement method (Figure 1b), then the correlation between them should be higher than among variables that do not use a common method. Thus, in Figure 1b if Y_1

Figure 1
Two Construct Models

(a) Independent Constructs
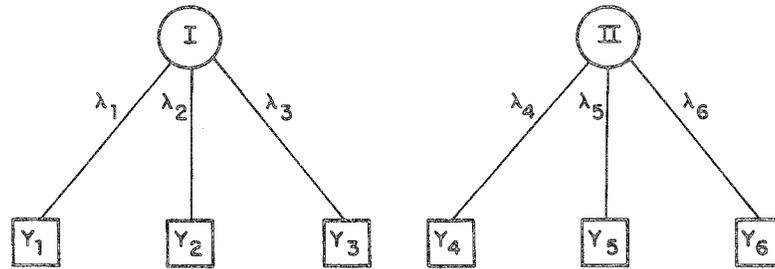

(b) Related Constructs

and Y_4 share a common measurement base, then r_{14} should be potentially higher than, say, r_{15}. However, even when variables 4 and 5 are maximally different measures, r_{45} should be greater than r_{14}. If a variable A shares more variance with its construct than another variable B, then the correlation between A and (1) all other indicators of the construct and (2) any indicator of a related construct should be larger.

Examinations of convergent and discriminant validity have followed three major approaches. The first approach involves examining some form of the multitrait-multimethod matrix (MTMM) approach to analyze the correlation matrix (i.e., Sullivan & Feldman, 1979) for internal and external consistency (Anderson & Gerbing, 1982).

Kalleberg and Klugel (1975) presented a path-analytic decomposition of the MTMM matrix, which they found inadequate due to assumptions that traits and methods are uncorrelated and that methods are minimally correlated with each other (Jackson, 1969). They also found the method "basically qualitative in nature" and turned to confirmatory factor analysis to assess validity. Kavanaugh, MacKinney, and Wolins (1971) used analysis of variance to decompose the observed data based on person, trait, and method variables.

A second approach to assessing validity is factor analysis. When exploratory factor analysis is used, the results are often largely—but not completely—consistent with prior theory. For example, if there are seven constructs with three indicators of each, a principal components analysis must produce seven eigenvalues greater than 1.0 (or possibly 10 if each method also had an effect) and the rotated factor matrix must have seven factors, each with three indicators loading on it. Because this almost never occurs, the researcher can only conclude that the results "generally support the prior theory." Thus, because exploratory factor analysis is imprecise and orthogonal factor analysis is not appropriate for correlated constructs, confirmatory factor analysis often is used. This approach (Anderson & Gerbing, 1982; Jöreskog,

Figure 2
Indicator Variable Correlation Matrix
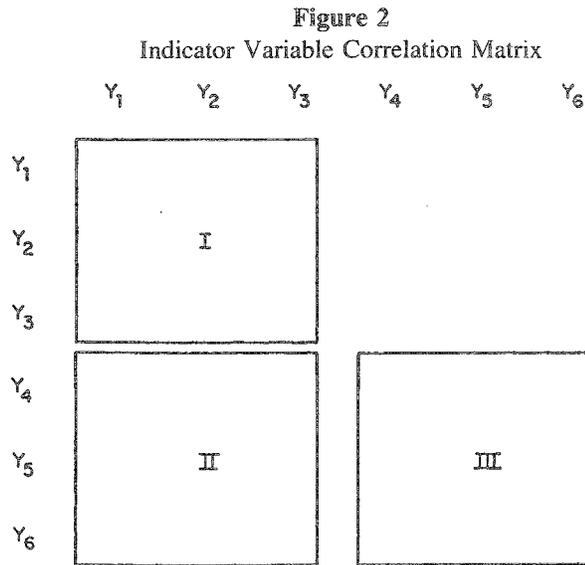
1971; Long, 1983) is reasonable and widely recommended (Schmitt & Stults, 1986) but requires an iterative procedure to arrive at a solution.

A number of factor-analytic procedures have been suggested for assessing convergent and discriminant validity. In a comparison of several of these, Schmitt, Coyle, and Saari (1977) suggested that the approach of Jackson (1969, 1975) has the advantage of requiring that the researcher explicitly state assumptions about the hypothesized structure.

The third approach used to assess convergent and discriminant validity jointly examines both the measurement model (convergent and discriminant validity) and the structural model (nomological validity), using LISREL (Bagozzi, 1980; Jöreskog & Sörbom, 1979) or PLS (Fornell & Larker, 1981; Wold, 1980). This approach requires some prior theory, so that the relations among the constructs must be specified exactly. Moreover, because the measurement and structural models are estimated simultaneously, errors in specification of the structural model may affect the results for the measurement model (Burt, 1976; Kumar & Dillon, 1986). Consequently, a preliminary analysis of convergent and discriminant validity seems desirable before attempting to estimate a full structural equations model.

Recent attempts to demonstrate discriminant validity have focused on using LISREL on a series of nested models, as recommended by Widaman (1985). The procedure begins with a parsimonious null model assuming that all measures are identical and gradually relaxes the assumption, allowing multiple traits (constructs) and multiple methods to exist. When the model that allows multiple traits fits better than the model that assumes a single trait, discriminant validity is said to exist.

Two possible weaknesses exist in this procedure. First, it requires the estimation of several LISREL models and hence is fairly computationally intense. Second, the procedure tests sets of coefficients simultaneously. Thus, if in a three-construct model, construct A is distinct from constructs B and C, and B and C are indistinguishable, the fit of the model allowing for separate traits will be significantly better than the model assuming a single trait. This will indicate that discriminant validity exists whereas it does not in the case of constructs B and C. Of course, adding separate traits one at a time eliminates this problem, but with a large number of traits this procedure becomes rather unwieldy.

The method proposed here differs from all three of these general approaches, yet incorporates elements of all. As in the first approach, it uses an analysis-of-variance procedure where the dependent variable

is the correlation between items, rather than the item score. In addition, it decomposes correlations based on trait and method variables. As in the second approach, this method requires assumptions, but only about (1) which items measure which construct and (2) which items share a common measurement method. In contrast to the third approach, this method, although easily adaptable to the nested-model approach, produces a single-stage least-squares estimate of a ''null'' model (separate traits and methods), which allows examination of discriminant validity from a single output.

Proposed Method

The approach followed here ''explains'' correlations using a dummy variable regression analysis in terms of the measure-to-construct relations, the construct-to-construct relations, and the impact of common measurement bases. Every correlation r_{ij}, where i is a measure of construct p and j is a measure of construct q, is assumed to be affected by several factors:

1.   The closeness of the relation between constructs p and q.
2.   Any common method that i and j might share.
3.   The relation of each measure to its construct.

The procedure estimates the impact of each of these on correlations, thus allowing an explicit test of the desired conditions.

Decomposing the correlations is directly related to the analysis-of-variance approach (see the Appendix). In essence, the correlations are decomposed as follows:

$$\text{Corr}(X_i, X_j) = B_0 + \sum_p \sum_q B_{pq} Z_{pq} + \sum_m \sum_n B_{mn} Z_{mn} + \sum B_k Z_k + e_{ij} \quad , \tag{1}$$

where $Z_{pq} = 1$ if p and q are the two constructs, 0 otherwise;
   $Z_{mn} = 1$ if m and n are the two methods, 0 otherwise; and
   $Z_k = 1$ if $i = k$ or $j = k$, 0 otherwise.

Due to collinearity among the dummy variables, it is necessary to estimate a reduced form of Equation 1. Two models are of special interest. The first assumes that all the B_ks equal 0 (which implies all are equally good measures of their respective constructs):

$$\text{Corr}(X_i, X_j) = B_0 + \sum_p \sum_q B_{pq} Z_{pq} + \sum_m \sum_n B_{mn} Z_{mn} + e_{ij} \quad . \tag{2}$$

This strong assumption allows estimation of, for example, correlated methods of measurement (B_{mn}s). A more reasonable initial assumption allows for the unequal quality of variables and deletes the variables relating to two measures of the same construct (Z_{pp}s) and the variables relating to measures by different methods (Z_{mn}s). This model becomes

$$\text{Corr}(X_i, X_j) = B_0 + \sum_k B_k Z_k + \sum_p \sum_q B_{pq} Z_{pq} + \sum_m B_{mm} Z_{mm} + e_{ij} \quad . \tag{3}$$

The interpretation of the coefficients in Equation 3 is as follows (assuming deletion of one of the Z_k variables).

Coefficients

Constant.   This represents the average correlation of the (dummy) variable omitted from the equation (in order to permit estimation) with other measures of the same construct.

Individual variable coefficients (B_is).   These indicate how the average correlation for a particular variable differs from the omitted (dummy) variable's average correlation. Thus, $B_i = .2$ means that correlations involving variable i average .2 larger than those involving the omitted variable. Ignoring

common method effects, the correlation among two variables which measure the same construct is predicted to be $B_0 + B_i + B_j$. Therefore, in order for a construct p to have convergent validity, all $B_0 + B_i + B_j$ must be greater than 0 for all (i,j) pairs, where i and j are measures of the particular construct p.

Cross-construct coefficients ($B_{pq}s$).   These coefficients indicate how the average correlation between a measure of construct p and a measure of construct q differs from the average correlation between two measures of the same construct (i.e., two measures of p, two measures of q, ...). In order to establish discriminant validity between constructs p and q, B_{pq} should be significantly less than 0. Further, correlated constructs exist if $B_0 + B_i + B_j + B_{pq}$ is greater than 0 for i (a measure of construct p) and j (a measure of construct q).

Common method coefficients ($B_{mm}s$).   These coefficients measure the difference between the average correlation of two measures using method m and two measures using different methods. When B_{mm} is significantly positive, this indicates a significant method effect for method m.

One useful feature of Equation 3 is that it provides an estimate of the magnitude of the common method effect. This magnitude is interesting both in its own right and in comparison to the magnitude of the effects of the traits. If the method effects are larger than the trait effects, then a problem may exist in isolating the effects of the traits, and the strength of the constructs is relatively weak.

Widaman (1985) suggested that the procedures of Campbell and Fiske (1959) have three main problems: (1) the correlations are not independent, (2) estimates of the trait-related and method-related variance are not obtainable, and (3) the procedures are distorted when the variables differ in reliability. Here the dependence of the correlations is explicitly modeled in Equation 3. The proposed procedure specifically separates trait-related and method-related sources of the correlations (the $B_{pq}s$ and $B_{mm}s$), which is one of the advantages of using LISREL for analyzing validity. Furthermore, Equation 3 explicitly allows for unequal reliabilities ($B_k s$) and provides a framework in which to test the constraint of equal reliabilities.

Estimation

The statistical estimation of the model depends on the distribution of e_{ij}. The variance of a correlation, and hence e_{ij}, is approximately $(1 - r_{ij}^2)/(n - 2)$. This suggests that a weighted least-squares (WLS) procedure will be more efficient than ordinary least-squares (OLS), and that WLS will tend to weight the larger correlations (typically between measures of the same construct) more heavily. The covariances of the $e_{ij}s$ are assumed to be 0. Note that this does not imply that the correlations themselves are unrelated. If all correlations come from the same dataset, constraints exist on r_{23} if r_{12} and r_{13} are known, although unless r_{12} and r_{13} are large the constraint is not very tight. Here it is assumed that the deviations in the correlations from some ''true'' value are uncorrelated.

An alternative approach is to use Fisher's r to z transformation on the correlations in Equation 3 to normalize the variances. This method is not recommended because the coefficients are difficult to interpret in terms of the magnitude of correlations, and because this procedure minimizes errors in predicting Z, which does not necessarily minimize error in predicting r.

Variable Coding

Coding the variables for the analysis of the model in Equation 3 is straightforward. For each correlation (i,j) the dummy variables D_i and D_j are coded as 1, and the rest of the dummy variables are coded as 0. If variable i is a measure of construct p and j is a measure of construct q, then D_{pq} is coded as 1 and

the other ''cross-construct'' dummy variables are coded as 0. Finally, if variables i and j share method m, the D_{mm} is coded as 1 and the rest of the common method dummy variables are coded as 0.

As an example of coding, consider the variables analyzed by Lawler (1967) which consisted of nine variables representing all combinations of ratings by supervisors, peers, and persons themselves (methods) of ability, quality, and effort on the job (traits):

1. Supervisor rating of quality (A1),
2. Peer rating of quality (A2),
3. Self rating of quality (A3),
4. Supervisor rating of ability (B1),
5. Peer rating of ability (B2),
6. Self rating of ability (B3),
7. Supervisor rating of effort (C1),
8. Peer rating of effort (C2), and
9. Self rating of effort (C3).

These nine variables produced 36 pairwise correlations, which are shown, along with the coding, in Table 1.

Example Applications

Assessing Discriminant Validity

This example examines discriminant validity using the superior, peer, and self ratings of performance on three criteria (quality, ability, and effort) presented by Lawler (1967). These correlations were analyzed using Equation 3 and OLS.

The data in Table 2 lead to some obvious conclusions:

1. $B_0 + B_i + B_j > 0$ for all i, j in the same construct, indicating convergent validity, though only weakly when self ratings are involved. (The negative values for the self-rating measures indicate that these correlations tend to be somewhat smaller.)
2. There is clear discriminant validity between ability and effort ($B = -.22$) and significant discriminant validity between quality and effort ($B = -.11$). There is no discriminant validity between quality and ability ($B = -.05$, not significantly different from 0).
3. The correlations involving supervisor or peer ratings are larger than those involving self ratings in general (B_is range from .27 to .35 vs. $-.07$ and $-.08$ for self ratings).
4. $B_0 + B_i + B_j + B_{pq}$ is generally positive, indicating correlated constructs, although again the self ratings tend to be different from the others.
5. There is massive method bias in the self ratings ($B = .64$), a smaller but significant one in the supervisor ratings ($B = .13$), and a nonsignificant B in peer ratings (.10). This suggests that self ratings are consistent across attributes (ability, quality, effort).

These results mirror those reported by Widaman (1985). His chosen model (Model 3C in his paper) showed a strong correlation between the supervisor and peer method factors (related to the higher correlations in their person variables reported here). He argued that effort is the most distinct trait (here effort was discriminantly different from both quality and ability), but that the traits were not easily distinguishable (here indicated by the relatively small coefficients of the trait variables). Thus, the one-step procedure recommended here produces essentially the same interpretation as the nested model testing procedure used by Widaman (1985).

Table 1

Bivariate Correlations (r) and Variable Coding for Lawler Data

(A = Quality, B = Ability, C = Effort, 1 = Supervisor, 2 = Peer, 3 = Self-Self)

Variables	r	Individual Variable Dummies (D_j)									Construct to Construct Dummies (D_{pq})			Common Method Dummies (D_{mm})		
		A1	B1	C1	A2	B2	C2	A3	B3	C3	A-B	A-C	B-C	1-1	2-2	3-3
A1-B1	.53	1	1								1			1		
C1	.56	1		1								1		1		
A2	.65	1			1											
B2	.42	1				1					1					
C2	.40	1					1					1				
A3	.01	1						1								
B3	.03	1							1		1					
C3	.06	1								1		1				
B1-C1	.44		1	1									1	1		
A2	.38		1		1						1					
B2	.52		1			1										
C2	.31		1				1						1			
A3	.01		1					1			1					
B3	.13		1						1							
C3	.01		1							1			1			
C1-A2	.40			1	1							1				
B2	.30			1		1							1			
C2	.53			1			1									
A3	.09			1				1				1				
B3	.03			1					1				1			
C3	.30			1						1						
A2-B2	.56				1	1					1				1	
C2	.56				1		1					1			1	
A3	.01				1			1								
B3	.04				1				1		1					
C3	.02				1					1		1				
B2-C2	.40					1	1						1		1	
A3	.17					1		1			1					
B3	.09					1			1							
C3	.01					1				1			1			
C2-A3	.10						1	1				1				
B3	.02						1		1				1			
C3	.30						1			1						
A3-B3	.43							1	1		1					1
C3	.40							1		1		1				1
B3-C3	.14								1	1			1			1

Table 2
B Weights from the Lawler
Data Analysis (R^2 = .92,
Adjusted R^2 = .87)

Variable	B
Individual Variables	
Supervisor Quality	.27*
Ability	.28*
Effort	.35*
Peer Quality	.27*
Ability	.30*
Effort	.35*
Self Quality	−.08
Ability	−.07
Effort	--
Trait Variables	
Quality-Ability	−.05
Ability-Effort	−.22*
Quality-Effort	−.11*
Method Variables	
Supervisor-Supervisor	.13*
Peer-Peer	.10
Self-Self	.64*
Constant	−.09

*t greater than 2.

Assessing the Three-Component Model of Attitude

As a more complex example of the procedure, an assessment of the convergent and discriminant validity of the three-component attitude model was made. Specifically, the data of Ostrom (1969) and Kothandapani (1971), which have been analyzed using covariance structure methods (Bagozzi, 1978; Widaman, 1985), were analyzed using a form of Equation 3.

Ostrom data.   The model used for the Ostrom data was

$$r = B_0 + \sum_{i=1}^{11} B_i Z_i + \sum_{j=1}^{4} B_{11+j} Z_{jj} + B_{16} Z_{12} + B_{17} Z_{13} + B_{18} Z_{23} \quad . \tag{4}$$

This model provides a direct test for method variance (if B_{12}, B_{13}, B_{14}, and B_{15} are positive and significant, this indicates common method variance). It also provides a direct test of discriminant validity (if B_{16}, B_{17}, and B_{18} are significantly negative, there is discriminant validity). Convergent validity requires that B_0 and all $B_0 + B_i$ be significantly positive; in other words, that correlations between measures of the same construct should be positive. A stronger test is possible if a certain minimum level of correlation of measures to their construct (e.g., rs above .6) is desired. In such situations, B_0 and $B_0 + B_i$ must be above that cutoff level.

The results from an OLS regression analysis of the model are shown in Table 3, which used 1-0 coding for the dummy variables. The conclusions that can be drawn from the Ostrom (1969) data are as follows:

1.   Measure-to-construct relations are remarkably similar for any pair of variables because B_is range only from −.08 to +.09.

2. Correlations involving Likert scales tend to be about .1 larger, suggesting that a stronger method component exists in them.
3. A significant common method effect is evident for self-rating scales ($B_{15} = .15$) and a small but significant effect exists for Likert scales ($B_{13} = .06$).
4. The behavioral construct is significantly different from both the affective and cognitive component, though only barely so. Moreover, the affective and cognitive components are not significantly different at the .05 level, and the sizes of the discriminant variables ($-.03$, $-.04$, $-.02$) suggest that the constructs are highly related.
5. Because all $B_0 + B_i + B_j \geqslant .46$, there is reasonable convergent validity.

Kothandapani data.   The Kothandapani (1971) data were analyzed using the same version of Equation 3 that was used for the Ostrom data, except that the three components were Feeling, Belief, and Intention instead of Affect, Behavior, and Cognition. The results (also shown in Table 3) differed in several ways:
1. Correlations involving Likert and Guilford scales tend to be higher (*B*s range from $-.07$ to $+.02$ for these vs. $-.19$ to $-.09$ for the Thurstone and Guttman scales).
2. A significant and substantial common method effect exists for all four methods. The *B*s = .23, .25, .29, and .31, indicating, for example, that two Thurstone scales on average are correlated .23 higher than correlations of variables which share no common method.
3. All the constructs are significantly distinct, with Feeling (Affect) and Intention (Behavior) the most distinct. Also, the coefficients of the discriminant variables are as large as the coefficients of the method factors (*B*s = $-.22$, $-.26$, $-.42$).

Table 3
B Weights and Significance Levels (*p*) From OLS
Decomposition of Ostrom and Kothandapani Correlation Matrices

	Ostrom Data				Kothandapani Data		
	Variable	B	p		Variable	B	p
Measure Variables							
Thurstone	Affective (A)	.04	.01	Thurstone	Feeling (F)	−.09	.05
	Behavior (B)	−.01	--		Belief (B)	−.19	.01
	Cognition (C)	.06	.01		Intention (I)	−.10	.05
Likert	Affective	.09	.01	Likert	Feeling	−.01	--
	Behavior	.12	.01		Belief	−.04	--
	Cognition	.11	.01		Intention	−.02	--
Guttman	Affective	−.08	.01	Guttman	Feeling	−.11	.05
	Behavior	−.01	--		Belief	−.13	.01
	Cognition	.04	.01		Intention	−.14	.01
Self-Rating	Affective	.09	.01	Guilford	Feeling	.02	--
	Behavior	.01	--		Belief	−.07	--
Method Variables							
Both Thurstone		.03	--	Both Thurstone		.23	.01
Both Likert		.06	.01	Both Likert		.25	.01
Both Guttman		−.01	--	Both Guttman		.29	.01
Both Self-Rating		.15	.01	Both Guilford		.31	.01
Discriminant Variables							
One A, Other B		−.03	.05	One F, Other B		−.22	.01
B, C		−.04	.01	B, I		−.26	.01
A, C		−.02	.10	F, I		−.42	.01
Intercept		.55	.01			.62	.01
R^2		.93				.87	
Adj. R^2		.90				.82	

4.   Because $B_0 = .62$ and all $B_0 + B_i + B_j \geq .33$, there is convergent validity.

Use of Fisher's r to z transformation is often recommended to normalize the errors. Even though the transformation is problematic for correlations above .7 (which often occur in MTMM data), the data were reanalyzed using the transformation. Specifically, the dependent variable used was $\frac{1}{2} \ln [(1 + r_{ij})/(1 - r_{ij})]$. All coefficients retained the same sign and the R^2 values were comparable.

Overall, the Kothandapani (1971) data give much stronger evidence of discriminant validity than the Ostrom (1969) data. Hence, this procedure seems to provide essentially the same interpretation as the nested model tests of Widaman (1985), in contrast to that of Bagozzi (1978). Notice also that formal nested model tests are also not only possible but relatively simple with the proposed method. However, running the complete model makes them unnecessary.

Limitations

The method as proposed has two major limitations. First, because the model in Equation 3 is only a linear approximation of the correlation generating process, the analysis contains some model (specification) error which can obscure differences. Based on initial investigations of well-behaved datasets, this appears to be a relatively minor problem.

Second, the tests do not explicitly take into account the sample size involved in the correlations. The model which is estimated is thus:

$$\text{Measured Correlation}_{i,j} = \text{True Correlation}_{i,j} + \text{Sampling Error}_{i,j} = \Sigma BD + \text{Model Error}_{i,j} \quad , \qquad (5)$$

where ΣBD is the model in Equation 3. Hence the model is essentially

$$\text{True Correlation}_{i,j} = \Sigma BD + e_{i,j} \quad , \qquad (6)$$

where $e_{i,j}$ is a combination of sampling and model error.

This suggests that for a smaller sample size, sampling error will be larger. Hence, assuming that sampling error and model error are uncorrelated, small samples will produce more error in the regression estimates, leading to fewer significant coefficients and hence less evidence of convergent and discriminant validity. Future developments may make it possible to explicitly include sample size in the analysis of the correlations.

Appendix

Relation of Decomposition of Correlations to the Analysis-of-Variance Decomposition of Item Scores

Consider a P construct model where each construct is measured by the same M methods. The observed score for person k on the ith measured variable which uses method m to assess construct p is

$$X_{ik} = B_0 + B_k + C_p + D_m + I_{pm} + f_{pk} + g_{mk} + h_{pmk} + e_{ik} \quad , \qquad (A1)$$

where B_0 is the average score on the items,

B_k is the average score of person k minus the average score of all persons on all items (person effect),

C_p is the average score of items on the pth construct minus the average score of all items (construct effect),

D_m is the average score on the mth method minus the average score on all methods (method effect),

I_{pm} is the interaction of the pth construct and the mth method,

f_{pk} is the interaction effect of construct p and person k,

g_{mk} is the interaction effect of method m and person k,

h_{pmk} is the interaction of the pth construct, the mth method, and the kth person, and

e_{ik} is the random component.

Rewriting, this gives

$$X_{ik} = (B_0 + C_p + D_m + I_{pm}) + B_k + (f_{pk} + g_{mk}) + h_{pmk} + e_{ik} \tag{A2}$$

or

$$X_{ik} = T_i + B_k + (f_{pk} + g_{mk}) + h_{pmk} + e_{ik} \quad , \tag{A3}$$

where T_i is the mean value on item i. Similarly,

$$X_{jk} = T_j + B_k + (f_{qk} + g_{nk}) + h_{qnk} + e_{jk} \quad . \tag{A4}$$

Now,

$$X_{ik} - \overline{\overline{X}}_i = B_k + f_{pk} + g_{mk} + h_{pmk} + e_{ik} \tag{A5}$$

and

$$X_{jk} - \overline{\overline{X}}_j = B_k + f_{qk} + g_{nk} + h_{qnk} + e_{jk} \quad . \tag{A6}$$

If two constructs p and q are correlated, f_{pk} and f_{qk} will be correlated. Similarly, if two methods are correlated, g_{mk} and g_{nk} will be correlated. Also,

$$\text{Cov}(X_k, X_j) = \text{Var } B_k + \text{Corr}(p,q)(\text{Var } f_p)^{1/2} (\text{Var } f_q)^{1/2} + \text{Corr}(m,n)(\text{Var } g_m)^{1/2} (\text{Var } g_n)^{1/2} +$$
$$\text{Corr}(pm,qn)(\text{Var } h_{pm})^{1/2} (\text{Var } h_{qn})^{1/2} \quad , \tag{A7}$$

$$\text{Var}(X_i) = \text{Var } B_k + \text{Var } f_p + \text{Var } g_m + \text{Var } h_{pm} + \text{Var } e_i \quad , \tag{A8}$$

$$\text{Var}(X_j) = \text{Var } B_k + \text{Var } f_q + \text{Var } g_n + \text{Var } h_{gn} + \text{Var } e_j \quad . \tag{A9}$$

Because

$$\text{Corr}(X_i, X_j) = \frac{\text{Cov}(X_i X_j)}{(\text{Var } X_i)^{1/2} (\text{Var } X_j)^{1/2}} \quad , \tag{A10}$$

the correlation between two measures is positively related to

1. Large construct correlation: $\text{Corr}(p,q)(\text{Var } f_p)^{1/2}(\text{Var } f_q)^{1/2}$
2. Large method correlation: $\text{Corr}(m,n)(\text{Var } g_m)^{1/2}(\text{Var } g_n)^{1/2}$
3. Large construct-method correlation: $\text{Corr}(pm,qn)(\text{Var } h_{pm})^{1/2}(\text{Var } h_{qn})^{1/2}$
4. Small variable variances: $\text{Var } e_i$, $\text{Var } e_j$

A linear approximation of this is given by

$$\text{Corr}(X_i, X_j) = B_0 + \sum B_{pq} Z_{pq} + \sum B_{mn} Z_{mn} + \sum B_{pmqn} Z_{pmqn} + \sum B_k Z_k + e_{ij} \quad , \tag{A11}$$

where $Z_{pq} = 1$ if p and q are the constructs, 0 otherwise;

$Z_{mn} = 1$ if m and n are the methods, 0 otherwise;

$Z_{pmqn} = 1$ if X_i is a measure of construct p by method m and X_j is a measure of construct q by method n, 0 otherwise; and

$Z_k = 1$ if $i = k$ or $j = k$, 0 otherwise.

The number of parameters to be estimated in Equation A11 is as follows:

$$B_{pq}: \frac{P(P-1)}{2} + P \quad , \tag{A12}$$

$$B_{mn}: \frac{M(M-1)}{2} + M \quad , \tag{A13}$$

$$B_{pmqm}: \left[\frac{P(P-1)}{2} + P\right]\left[\frac{M(M-1)}{2} + M\right] \quad , \tag{A14}$$

$$B_k: k \quad , \tag{A15}$$

For three constructs and three methods, this becomes

$$6 + 6 + 36 + 9 = 57 \quad . \tag{A16}$$

Because there are only $(9 \times 8)/2 = 36$ correlations, the full model is under-identified. In order to have an identified model, it is necessary to assume that the correlation of h_{pmk} and $h_{qnk} = 0$. This reduces the model to a manageable size. The assumption means that an individual's response to a combination of construct and method is explainable by a main-effects-only model. (Note that selected interactions could be estimated if a strong a priori reason existed for examining them.)

The model in Equation A11 thus reduces to

$$\text{Corr}(X_i, X_j) = B_0 + \sum B_{pq}Z_{pq} + \sum B_{mn}Z_{mn} + \sum B_k Z_k + e_{ij} \quad . \tag{A17}$$

Various nested versions of Equation A17 are also of interest. For example, if it is assumed that there is no correlation between the effects of methods m and n (except the common method effect when $m = n$), the number of parameters reduces. Comparison of the fit of this model to the fuller model in Equation A17 provides an approximate test of the reasonableness of this assumption.

References

Anastasi, A. (1968). *Psychological testing* (3rd ed.). New York: Macmillan.

Anderson, J. C., & Gerbing, D. W. (1982). Some methods for respecifying measurement models to obtain unidimensional construct measurement. *Journal of Marketing Research, 19,* 453–460.

Bagozzi, R. P. (1978). The construct validity of affective, behavioral, and cognitive components of attitude by analysis of covariance structures. *Multivariate Behavioral Research, 13,* 9–31.

Bagozzi, R. P. (1980). *Causal models in marketing.* New York: Wiley.

Bohrnstedt, G. W. (1970). Reliability and validity assessment in attitude measurement. In G. F. Summers (Ed.), *Attitude measurement.* Chicago: Rand-McNally.

Burt, R. S. (1976). Interpretation confounding of unobserved variables in structural equation models. *Sociological Methods and Research, 5,* 33–52.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Fornell, C., & Larker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18,* 39–50.

Jackson, D. N. (1969). Multimethod factor analysis in the evaluation of convergent and discriminant validity. *Psychological Bulletin, 72,* 30–49.

Jackson, D. N. (1975). Multimethod factor analysis: A reformulation. *Psychological Bulletin, 75,* 421–423.

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

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models.* Cambridge MA: Abe Books.

Kalleberg, A. L., & Klugel, J. R. (1975). Analysis of the multitrait-multimethod matrix: Some limitations and an alternative. *Journal of Applied Psychology, 60,* 1–9.

Kavanaugh, M. J., MacKinney, A. C., & Wolins, L. (1971). Issues in managerial performance: Multitrait-multimethod analyses of ratings. *Psychological Bulletin, 75,* 34–39.

Kothandapani, V. (1971). Validation of feeling, belief, and intention to act as three components of attitude and their contribution to prediction of contraceptive behavior. *Journal of Personality and Social Psychology, 19,* 321–333.

Kumar, A., & Dillon, W. (1986). *The interaction of measurement and structure in simultaneous equation models with unobservable variables.* Unpublished working paper.

Lawler, E. E. (1967). The multitrait-multirater approach to measuring managerial performance. *Journal of Applied Psychology, 51,* 369–381.

Lehmann, D. R. (1974). Some alternatives to linear factor analysis for variable grouping applied to buyer

behavior models. *Journal of Marketing Research, 11*, 206–213.

Long, J. S. (1983). *Confirmatory factor analysis*. Beverly Hills CA: Sage.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ostrom, T. M. (1969). The relationship between the affective, behavioral and cognitive components of attitude. *Journal of Experimental Social Psychology, 5*, 12–30.

Schmitt, N., Coyle, B. W., & Saari, B. B. (1977). A review and critique of analyses of multitrait-multimethod matrices. *Multivariate Behavioral Research, 12*, 447–478.

Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 10*, 1–22.

Sullivan, J. L., & Feldman, S. (1979). *Multiple indicators*. Beverly Hills CA: Sage.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1–26.

Wold, H. (1980). Soft modeling: Intermediate between traditional model building and data analysis. *Multivariate Statistics, 6*, 333–346.

Author's Address

Send requests for reprints or further information to Donald R. Lehmann, 507 Uris Hall, Columbia University, New York NY 10027, U.S.A.