Covariance and Regression Slope Models for Studying Validity Generalization

Nambury S. Raju
Illinois Institute of Technology

Rodney Fralicx
Stanard and Associates

Stephen D. Steinhaus
Illinois Institute of Technology

Two new models, the covariance and regression slope models, are proposed for assessing validity generalization. The new models are less restrictive in that they require only one hypothetical distribution (distribution of range restriction for the covariance model and distribution of predictor reliability for the regression slope model) for their implementation, in contrast to the correlation model which requires hypothetical distributions for criterion reliability, predictor reliability, and range restriction. The new models, however, are somewhat limited in their applicability since they both assume common metrics for predictors and criteria across validation studies. Several simulation (monte carlo) studies showed the new models to be quite accurate in estimating the mean and variance of population true covariances and regression slopes. The results also showed that the accuracy of the covariance, regression slope, and correlation models is affected by the degree to which hypothetical distributions of artifacts match their true distributions; the regression slope model appears to be slightly more robust than the other two models.

The currently available meta-analytic procedures for assessing validity generalization are based on the correlation model (Callender & Osburn, 1980; Pearlman, Schmidt, & Hunter, 1980; Raju & Burke, 1983; Schmidt, Gast-Rosenberg, & Hunter, 1980), which can be expressed as

$$r = \rho \sqrt{r_{xx}} \sqrt{r_{yy}} \frac{u}{\left[1 + (1 - u^2) \, \rho^2 r_{xx} r_{yy} \right]^{1/2}} + e \quad , \tag{1}$$

where r is the observed validity coefficient,

ρ is the unattenuated, unrestricted population validity,

r_{yy} and r_{xx} are the unrestricted population reliabilities of criterion y and predictor x, respectively,

u is the ratio of the restricted standard deviation to the unrestricted standard deviation of x in the population, and

e is the sampling error.

These procedures are designed to estimate how much of the variance in the observed validity (correlation) coefficients is due to variations in criterion reliabilities, predictor reliabilities, range restrictions, and sampling errors. In addition, they provide estimates of the mean (M_ρ) and variance (V_ρ) of ρ. Results from several monte carlo studies (Callender & Osburn, 1980; Callender, Osburn, Greener, & Ashworth,

195

1982; Raju & Burke, 1983) show these procedures to be quite accurate in estimating the parameters of interest.

One major concern with correlation-based procedures is that they rely on hypothetical distributions for criterion reliability, predictor reliability, and range restriction. These hypothetical distributions were originally proposed by Schmidt and Hunter (1977) as a solution to the serious practical problem surrounding the general unavailability of reliability and range restriction data for individual validity studies. The accuracy of the correlation-based procedures is, however, dependent upon how closely the hypothetical distributions match the true (but rarely obtainable) distributions of artifacts. A recent monte carlo study by Dribin (1981) showed that the accuracy of the independent multiplicative and non-interactive (correlation-based) procedures was somewhat affected by the degree of correspondence between the hypothetical and true distributions of artifacts. As Dribin's study was somewhat limited in scope, there is still a need for more comprehensive monte carlo studies to test the robustness of all of the correlation-based procedures. In addition, there is also a need to develop other models which could minimize or obviate the reliance on hypothetical distributions of artifacts. The purpose of the present study was to propose and assess two such models, the covariance model and the regression slope model, for studying validity generalization.

The Covariance Model

The covariance model can be written as

$$c_{xy} = \sigma_{xy}u^2 + e \quad , \tag{2}$$

where c_{xy} is the sample covariance,

σ_{xy} is the unattenuated and unrestricted population covariance, and

e is the sampling error associated with c_{xy}.

The proof of Equation 2 is given in Appendix A, and it is based on the assumptions that the measurement errors in x and y are uncorrelated with their true scores (x_t and y_t) and that the population regression coefficients in the restricted and unrestricted groups are equal. It should be noted that the correlation model, as expressed in Equation 1, is also subject to the same assumptions. Although both the correlation and covariance models are subject to the same set of assumptions and are both affected by predictor and criterion reliabilities, restriction of range, and sampling error, Equation 2 does not explicitly contain the r_{xx} and r_{yy} terms. Thus, the covariance model, unlike the correlation model, does not require information about the predictor and criterion reliabilities.

As opposed to the correlation model, the covariance model estimates the mean and variance of σ_{xy} instead of ρ_{xy}, using the data obtained from several validity studies. Ignoring subscripts x and y, estimates of the mean (M_σ) and variance (V_σ) of σ can be expressed as

$$\hat{M}_\sigma = \frac{M_c}{M_{u^2}} \tag{3}$$

and

$$\hat{V}_\sigma = \frac{V_c - V_e - M_\sigma^2 V_{u^2}}{M_{u^2} + V_{u^2}} \tag{4}$$

(see Appendix A for proofs). The quantities M_c and V_c in Equations 3 and 4 are easily obtainable; they are simply and respectively the mean and variance of observed covariances obtained from a set of different validity studies. The sample-size weighted M_c and V_c can be expressed as

$$M_c = \frac{\sum_{i=1}^{k} n_i c_i}{n} \tag{5}$$

and

$$V_c = \frac{\sum_{i=1}^{k} n_i c_i^2}{n} - M_c^2 \quad , \tag{6}$$

where

$$n = \sum_{i=1}^{k} n_i \quad , \tag{7}$$

    c_i is the observed covariance in Study i,

    n_i is the sample size in Study i, and

    k is the number of validity studies.

The formula for estimating the sampling variance (V_e) can be expressed (Jöreskog, 1969; Kendall & Stuart, 1977, p. 250) as

$$V_e = \frac{\sum_{i=1}^{k} \left(\frac{(C_i^2 + S_{x_i}^2 S_{y_i}^2) n_i}{n_i - 1} \right)}{n} \quad , \tag{8}$$

where $S_{x_i}^2$ and $S_{y_i}^2$ are the observed variances of the predictor and criterion, respectively, in Study i.

    In order to estimate $\hat{M}_\sigma$ and $\hat{V}_\sigma$, the mean and variance of u^2 are still required. As previously indicated in connection with the correlation model, the range restriction (u) values are not generally reported for individual validity studies. To estimate the mean and variance of u^2, the use of the hypothetical distribution of range restrictions proposed by Schmidt and Hunter (1977) is recommended, unless the distribution of range restrictions in an actual validity generalization study is available, or unless a different assumed distribution of range restrictions seems preferable. For Schmidt and Hunter's distribution of us, the mean and variance of u^2, as shown in Table 1, are .3690 and .0285, respectively.

    The parameters of interest can be estimated using Equations 3 and 4. While V_ρ is the essential criterion for inferring validity generalization in the correlation model, for the covariance model it may be inferred that the validity of a predictor under consideration is generalizable to similar predictor-job combinations whenever $\hat{V}_\sigma$ is zero or close to zero. This point is further elaborated below.

The Regression Slope Model

    The equation for the regression slope model can be written as

$$b_{yx} = B_{y_t x_t} r_{xx} + e \quad , \tag{9}$$

    where b_{yx} is the observed regression coefficient for predicting y from x,

        $B_{y_t x_t}$ is the unattenuated and unrestricted population regression coefficient, and

        e is the sampling error associated with b_{yx}.

The proof of Equation 9 is given in Appendix B; the assumptions underlying this equation are identical to the assumptions necessary for deriving the covariance and correlation models. Furthermore, the regression slope model, while reflecting the effects of predictor and criterion reliabilities, range restriction, and sampling error, does not explicitly utilize the r_{yy} and u terms. That is, the regression slope model does not require criterion reliability and range restriction data, whereas the correlation model does even though the two models require the same set of assumptions. Compared to the covariance model, the regression slope model does not require the range restriction data but does need the data on predictor reliability; criterion reliability information is not needed for either model.

    As with the other models, the objective is to estimate the mean and variance of the regression slope parameter ($B_{y_t x_t}$) for assessing validity generalization. Ignoring subscripts x and y, the mean (M_B) and

Table 1
Hypothetical Distributions of Criterion Reliability,
Predictor Reliability, and Range Restriction

Criterion Reliability (a^2)	f	Predictor Reliability (b^2)	f	SD of Predictor	u	f
.90	3	.90	15	10.00	1.000	5
.85	4	.85	30	7.01	.701	11
.80	6	.80	25	6.49	.649	16
.75	8	.75	20	6.03	.602	18
.70	10	.70	4	5.59	.559	18
.65	12	.60	4	5.15	.515	16
.60	14	.50	2	4.68	.468	11
.55	12			4.11	.411	5
.50	10					
.45	8					
.40	6					
.35	4					
.30	3					

$$M_a = .7686 \qquad M_b = .8943 \qquad M_u = .5945$$
$$V_a = .0093 \qquad V_b = .0022 \qquad V_u = .0146$$
$$M_{a^2} = .6000 \qquad M_{b^2} = .8020 \qquad M_{u^2} = .3690$$
$$V_{a^2} = .0214 \qquad V_{b^2} = .0066 \qquad V_{u^2} = .0285$$

variance (V_B) of B can be estimated with the following equations:

$$\hat{M}_B = \frac{M_b}{M_{r_{xx}}} \tag{10}$$

and

$$\hat{V}_B = \frac{V_b - V_e - \left(\frac{M_b}{M_{r_{xx}}}\right)^2 V_{rx}}{V_{r_{xx}} + M_{r_{xx}}^2} . \tag{11}$$

(See Appendix B for proofs.) The quantities M_b and V_b are the mean and variance, respectively, of the observed regression coefficients obtained from several validity studies. The sample-size weighted M_b and V_b can be obtained from

$$M_b = \frac{\sum_{i=1}^{k} n_i b_i}{n} \tag{12}$$

and

$$V_b = \frac{\sum_{i=1}^{k} n_i b_i^2}{n} - M_b^2 \quad , \tag{13}$$

where b_i is the observed regression coefficient obtained in Study i. The sampling variance (V_e) can be expressed (Draper & Smith, 1966) as

$$V_e = \frac{\sum_{i=1}^{k} n_i s_{y_i}^2 (1 - r_i^2)/[(n_i - 1)s_{x_i}^2]}{n} \quad , \tag{14}$$

where r_i is the correlation between x and y in Study i and the other terms are as previously defined.

Equations 12, 13, and 14 do not provide all the information needed in Equations 10 and 11; the $M_{r_{xx}}$ and $V_{r_{xx}}$ terms are still needed to estimate M_B and V_B. Because predictor reliabilities are not generally available for individual validity studies, use of the distribution of predictor reliabilities proposed by Pearlman et al. (1980) is recommended. The mean and variance of this distribution of predictor reliabilities are .8020 and .0066, respectively, as shown in Table 1. Using Equations 10 and 11, M_B and V_B can be estimated and then can be used to study the question of validity generalization. Whenever $\hat{V}_B$ is zero or close to zero, it may be concluded that the validity of a predictor under consideration is generalizable to similar job situations. This last point is further elaborated below.

The Question of Metric in the New Models

It is worth recalling that validity generalization as proposed by Schmidt and Hunter (1977) does not stipulate that the same predictor be used in all validity studies under consideration, provided the predictors from different studies measure the same set of constructs; the same also holds true for criterion evaluation instruments. Consequently, the correlation model, which does not depend on predictor and criterion standard deviations, is valid even when the predictor and criterion scores are expressed in different metrics across studies. In contrast to the correlation model, estimates of the parameters of interest in the covariance and regression slope models should theoretically be affected by scale differences in either or both of the predictor or criterion instruments used across the separate validity studies. Thus, using the models proposed above, validity generalization can be found when the true (unattenuated and unrestricted) covariances and regression slopes are equal, yielding V_σ or V_B equal to zero, and only when the predictor and criterion metrics are comparable across studies. Without the common metrics for the criterion and predictor variables, it is almost impossible to interpret credibility intervals of the type used with the correlation model.

The use of the new models for studying validity generalization, therefore, requires that the scales for the predictor and criterion variables be comparable across studies. Such a requirement will limit the applicability of the covariance and regression slope models. However, there are many practical situations in which this requirement is not difficult to satisfy. One example is the use of an aptitude test like the Scholastic Aptitude Test in various colleges where the grades are reported on a common scale ranging from 1 to 4. Another example is the use of the same predictor in different locations of a large corporation. Other examples in which the covariance and regression slope models are useful are given by Callender (1983) and Linn and Dunbar (1982). These and other non-trivial examples justify the study of the proposed models.

Method

Following Callender et al. (1982), monte carlo procedures were used to assess the comparative

accuracy of parameter estimation for the covariance, regression slope, and correlation models. Accuracy of estimation was determined by inspecting the differences between the estimated and true parameters connected with each of the models.

The estimates of covariance parameters were obtained using Equations 3 and 4. The population parameters in the regression slope model were estimated with Equations 10 and 11. Although five different procedures are currently available for estimating the relevant parameters in the correlation model, only one procedure, TSA 1, was used in the present study. The reason for this choice was that the results from a recent investigation (Raju & Burke, 1983) indicated that although all five procedures appeared to be comparable, TSA 1 was slightly more accurate than the others.

The generation of different populations for various phases of the present study was based on six different distributions of artifacts, two distributions of criterion and predictor standard deviations, and one distribution of unattenuated, unrestricted (true) correlation coefficients.

Distributions of Artifacts

Six different distributions were used in this study to assess the accuracy in the three models. Three of these distributions are shown in Table 1, and they are the same hypothetical distributions that Schmidt and Hunter (1977) and Pearlman et al. (1980) proposed for predictor reliability, criterion reliability, and range restriction. As previously indicated, some or all of these distributions are currently required in practice to study the question of validity generalization; the correlation model requires all three distri-

Table 2
Rectangular Distributions of Criterion Reliability,
Predictor Reliability, and Range Restriction

Criterion Reliability (a^2)	f	Predictor Reliability (b^2)	f	SD of Predictor	u	f
.90	10	.95	10	10.00	1.00	10
.80	10	.90	10	8.40	.84	10
.75	10	.85	10	7.60	.76	10
.70	10	.80	10	7.00	.70	10
.65	10	.75	10	6.50	.65	10
.60	10	.70	10	6.00	.60	10
.55	10	.65	10	5.60	.56	10
.50	10	.60	10	5.10	.51	10
.40	10	.55	10	4.70	.47	10
.30	10	.50	10	4.10	.41	10

$$M_a = .776 \qquad M_b = .847 \qquad M_u = .650$$
$$V_a = .014 \qquad V_b = .007 \qquad V_u = .029$$
$$M_{a^2} = .615 \qquad M_{b^2} = .725 \qquad M_{u^2} = .452$$
$$V_{a^2} = .0305 \qquad V_{b^2} = .021 \qquad V_{u^2} = .058$$

butions, whereas the covariance and regression slope models each only require one hypothetical distribution.

The next set of three distributions for predictor reliability, criterion reliability, and range restriction is shown in Table 2. They have the same ranges as their counterparts in Table 1, but are rectangular in shape. These distributions were used to determine the degree to which deviations from Schmidt and Hunter's hypothetical distributions of artifacts (see Table 1) would affect the accuracy of estimation obtained by the correlation, covariance, and regression slope models. It should be noted that there is nothing unique about the distributions in Table 2; in fact, this set is one of infinitely many sets of distributions available. These particular distributions were chosen, however, because they appeared to be different enough from the distributions of Schmidt and Hunter to provide a test of the robustness of the three models.

Distributions of Predictor and Criterion Standard Deviations

The rectangular distributions of criterion and predictor (attenuated) population standard deviations used in this study are given in Table 3. The ranges for the two distributions were chosen so as to reflect the range of values typically found in empirical validation studies. While the variation in standard deviations does not affect the correlation model, it was introduced into the design to evaluate the degree to which the covariance and regression slope models were affected by it.

Distribution of True Correlations

The distribution of unattenuated, unrestricted (true) correlation coefficients used in the present study was first proposed and used by Callender and Osburn (1980). This distribution is symmetrical and quasi-normal in shape with a mean of .500, a variance of .0308, and a range extending from .06 to .90. To generate various populations, unattenuated and unrestricted correlations were randomly selected from this distribution. The unattenuated and unrestricted covariances and regression slopes, however, were not

Table 3
Hypothetical Distributions of Criterion Standard
Deviations and Predictor Standard Deviations

Criterion Standard Deviation		Predictor Standard Deviation	
S.D.	f	S.D.	f
.60	10	1.00	10
.80	10	2.00	10
1.00	10	3.00	10
1.20	10	4.00	10
1.40	10	5.00	10
1.60	10	6.00	10
1.80	10	7.00	10
2.00	10	8.00	10
2.20	10	9.00	10
2.40	10	10.00	10

sampled from available pools of such coefficients; rather, they were obtained from the randomly generated unattenuated and unrestricted correlations ($\rho_{x_0 y_t}$), the population predictor and criterion reliabilities (r_{xx} and r_{yy}), and the attenuated predictor and criterion standard deviations (σ_x and σ_y). For example, the unattenuated and unrestricted covariance ($\sigma_{xy} = \sigma_{x_0 y_t}$) can be expressed as $\sigma_{xy} = \rho_{x_0 y_t} (r_{xx}^{1/2}\sigma_x)(r_{yy}^{1/2}\sigma_y)$, which will reduce to $\rho_{x_0 y_t} r_{xx}^{1/2} r_{yy}^{1/2}$ when $\sigma_x = \sigma_y = 1$. Similarly, the unattenuated and unrestricted regression slope ($B_{y_t x_t}$) can be written as $B_{y_t x_t} = \sigma_{xy}/(\sigma_x^2 r_{xx})$, which will reduce to $(\rho_{x_0 y_t} r_{yy}^{1/2})/r_{xx}^{1/2}$ when $\sigma_x = \sigma_y$.

Sample Sizes

Two different sample sizes were used in the present study to assess the effect of sample size on the three models. A sample size of 100 was chosen because it is at the upper end of the distribution of sample sizes typically used in validity generalization analysis. According to Callender et al. (1982), of the 129 validity generalization analyses published by June 1981, 33% had a mean sample size above 100, and only 6% had a mean sample size below 50. The sample size of 68 was also chosen for the present study because Lent, Aurbach, and Levin (1971) found it to be the average number of subjects in empirical validation studies in industry.

Procedure

Comparisons between the estimated and true parameters of interest were investigated by examining four different cases. A four-case design was used to manipulate the accuracy of estimation in the three models. This was accomplished by randomly selecting values of r_{xx}, r_{yy}, and u from either the Schmidt-Hunter distributions in Table 1 or the rectangular distributions in Table 2.

Case I–Part A. The parameters for each population were randomly generated from the appropriate hypothetical distributions. For Case IA, a true validity (ρ) was randomly selected from the distribution of true correlations; a range restriction value (u) was randomly selected from Table 1; and predictor reliability (r_{xx}) and criterion reliability (r_{yy}) were randomly selected from Table 2. The predictor and criterion standard deviations were set equal to 1.0. Each set of these six parameters defined a population; each set of population parameters (with the exception of predictor standard deviations) was drawn without replacement repeatedly to yield 100 populations. Two samples, one of 68 cases and the other of 100 cases, were randomly selected from each population, resulting in 100 samples for each sample size. (See Appendix C for the procedures used to generate random samples of appropriate size.) Within each sample size, three measures of validity (covariance, regression slope, and correlation) were computed for each of the 100 samples. The 100 values of each validity measure were then used to estimate the parameters of interest. Equations 3 and 4 were used to estimate M_σ and V_σ in the covariance model, Equations 10 and 11 were used to estimate M_B and V_B in the regression slope model, and the appropriate TSA 1 equations (Raju & Burke, 1983) were used to estimate M_ρ and V_ρ in the correlation model.

This procedure simulated one validity generalization study for each validity measure and each sample size. The procedure was repeated 50 times. Finally, the estimates of population parameters from the 50 validity generalization studies were averaged and compared with the true parameter values. This was done separately for the covariances, regression slopes and correlations.

The Case IA study was designed to favor the covariance model, inasmuch as the range restriction distribution used in the generation of 100 different populations was also used in Equations 3 and 4 to estimate the mean and variance of true population covariances. Furthermore, the predictor and criterion reliability values used in the 100 populations came from the rectangular distributions in Table 2, whereas the true distributions given in Table 1 for the same artifacts were used in estimating the parameters of interest for the regression slope and correlation models. This latter point should not have any bearing on

the covariance model because the predictor and criterion reliabilities are irrelevant for the covariance between two variables. The same, however, cannot be said of the regression slope and correlation models; their accuracy partly depends upon the degree to which the true and hypothetical distributions for predictor and criterion reliabilities match. This point further underscores the above statement that the Case IA study was designed to favor the covariance model.

Case I–Part B.   This study was similar to Case IA except that the predictor and criterion standard deviations (σ_x and σ_y) were allowed to vary. As with the selection of other parameters, for each population studied, one predictor standard deviation and one criterion standard deviation were randomly selected without replacement from the values shown in Table 3. Although the variation in standard deviations does not affect the correlation model, it was introduced into the design to evaluate the degree to which the covariance and regression slope models were affected by it.

Case II–Part A.   The ρ parameter was randomly selected (without replacement) from the distribution of true correlations. The r_{xx} parameter was selected without replacement from Table 1. The r_{yy} and u parameters, however, were selected from the rectangular distributions shown in Table 2. As in Case IA, the predictor and criterion standard deviations were set equal to 1.0. This case was designed to favor the regression slope model because the true and hypothetical distributions were the same for only r_{xx}.

Case II–Part B.   Same as Case IIA except that the predictor and criterion standard deviations were allowed to vary by randomly selecting them (without replacement) from Table 3.

Case III–Part A.   The ρ parameters were randomly selected from the distributions of true correlations. The other three parameters (r_{xx}, r_{yy}, and u) were randomly selected from Table 1, thereby making this a favorable case for all three models. The predictor and criterion standard deviations were set equal to 1.0.

Case III–Part B.   Same as Case IIIA except that the predictor and criterion standard deviations were randomly generated (without replacement) from Table 3.

Case IV–Part A.   The ρ parameters were selected as in the other cases. The parameter values for r_{xx}, r_{yy}, and u were randomly selected from the rectangular distributions in Table 2, thereby creating an unfavorable case for all three models; that is, the true and hypothetical distributions for each of the three artifacts did not match. As before, the predictor and criterion standard deviations were set equal to 1.0.

Case IV–Part B.   Same as Case IVA except that the predictor and criterion standard deviations were allowed to vary by randomly selecting them (without replacement) from Table 3.

Results

The results for all four cases ($N = 68$) are shown in Table 4. This table shows the estimated and true means and variances for the three validity measures (covariance, regression slope, and correlation). Table 4 also shows the percent difference, (estimated minus true)/true, separately for the mean and variance estimates, by validity measure. This percent difference is proposed as a means to compare the accuracy of the three models, because the scales for the three validity measures are different and a direct comparison of the estimates for the three models is therefore not advisable.

In Case IA, the covariance model was more accurate than the other two models, while the correlation model was the next most accurate. This is not a surprising result because the Case IA study was designed to favor the covariance model. The somewhat lower accuracy observed for the regression slope and correlation models is most likely due to the lack of congruence between the true and hypothetical distributions of some or all of the artifacts contained in these models. From a practical point of view, the accuracy appears to be quite good for all three models.

As expected, the data for Case IB also showed the covariance model to be more accurate than the other two models. The trends for Case IB were very similar to the trends for Case IA. The significant

finding in Case IB, as well as in Cases IIB, IIIB, and IVB, was the effect of variations in predictor and criterion standard deviations on the accuracy of the covariance and regression slope models. The accuracy of these two models in estimating the relevant parameters was not affected by the lack of metric comparability across studies. This does not imply, however, that the new models are useful for studying validity generalization when metrics are not comparable across studies. For generalizing validity, it must be shown that either $V_\sigma = 0$ or $V_B = 0$, which, for all practical purposes, is impossible unless the metrics are comparable.

Table 4
Estimates of True Mean and Variance
With Equal and Unequal Standard Deviations
for Cases I, II, III, and IV (N=68)

Model	Estimated Mean	True Mean	Percent* Difference	Estimated Variance	True Variance	Percent* Difference
Case IA: Equal SDs						
Covariance	.320	.329	-3	.015	.017	-12
Regression Slope	.406	.463	-12	.025	.035	-29
Correlation	.475	.500	-5	.036	.031	16
Case IB: Unequal SDs						
Covariance	2.695	2.731	-3	5.332	5.283	1
Regression Slope	.177	.200	-12	.039	.054	-28
Correlation	.477	.500	-5	.034	.031	10
Case IIA: Equal SDs						
Covariance	.420	.347	21	.036	.018	50
Regression Slope	.433	.434	0	.029	.028	4
Correlation	.545	.500	9	.046	.031	48
Case IIB: Unequal SDs						
Covariance	3.476	2.867	21	9.827	5.528	81
Regression Slope	.192	.191	1	.052	.050	4
Correlation	.544	.500	9	.050	.031	61
Case IIIA: Equal SDs						
Covariance	.332	.344	-3	.014	.017	-18
Regression Slope	.423	.431	-2	.022	.027	-19
Correlation	.493	.500	-1	.030	.031	-3
Case IIIB: Unequal SDs						
Covariance	2.780	2.836	-2	4.701	5.362	-12
Regression Slope	.187	.190	-2	.046	.050	-8
Correlation	.503	.500	1	.034	.031	10
Case IVA: Equal SDs						
Covariance	.393	.328	20	.038	.018	-111
Regression Slope	.402	.462	-13	.025	.035	-29
Correlation	.507	.500	1	.047	.031	52
Case IVB: Unequal SDs						
Covariance	3.285	2.700	22	8.596	4.989	72
Regression Slope	.181	.203	-11	.048	.058	-17
Correlation	.514	.500	3	.048	.031	55

*(Estimated - True)/True

The estimates in Case IIA and Case IIB for the regression slope model were extremely accurate. The accuracy of estimation was worst for the covariance model, with the correlation model doing somewhat better. The accuracy of the variance estimates was quite low for the covariance and correlation models; the percent difference was 48 or higher. Both of these models overestimated the parameters of interest. This lack of sufficient accuracy in the covariance and correlation models is most likely a reflection of the differences between the true and hypothetical distributions of artifacts relevant to these models. As before, the variation due to the differences in predictor and/or criterion standard deviations across studies was minimal.

All three models yielded accurate estimates for Case IIIA and Case IIIB, with no model showing a definite superiority over the other two. The fact that the true and hypothetical distributions were identical for the three artifacts in Case III explains the high and comparable degree of accuracy observed for all three models.

In Case IVA and Case IVB, where the true distribution did not match the hypothetical distribution for each of the three artifacts, the three models appeared to do an adequate job in estimating the means, with the correlation model providing more accurate estimates than the other two models. The regression slope model performed less well than the correlation model but better than the covariance model. With respect to estimating the true variances, the regression slope model was superior to the other two models. Considering the mean and variance estimates together, the regression slope model appears to have a slight edge over the correlation model. The Case IV data also confirmed the earlier observation that the accuracy of each model is affected by the degree to which the true distribution of a relevant artifact corresponds to its hypothetical distribution.[1]

Discussion and Conclusions

The results of the present study illustrate that the covariance and regression slope models provide very accurate estimates of the relevant parameters. Their accuracy, however, is contingent upon how well the hypothetical distribution of an artifact (u for the covariance model; r_{xx} for the regression slope model) matches its true distribution. The present study also shows that the accuracy of the correlation-based procedures is equally dependent on the degree of congruence between the hypothetical and true distributions of artifacts. Even though this study included only one correlation-based procedure (TSA 1), the results are probably quite generalizable to other correlation-based procedures in view of the monte carlo studies of Callender and Osburn (1980), Callender et al. (1982), and Raju and Burke (1983).

The three hypothetical distributions of artifacts, originally proposed by Schmidt and Hunter (1977) and Pearlman et al. (1980), are an essential part of all currently available models for studying validity generalization. Without these distributions, the correlation, covariance, and regression slope models can only be used to correct for sampling error. Yet the use of these distributions, while making it possible to correct for more artifacts, may also lead to inadequate estimates of parameters when the hypothetical distributions differ significantly from the true distributions. Unfortunately, it is difficult to know when and to what degree these two sets of distributions (true and hypothetical) differ from each other. At the present state of development, the results from any validity generalization study, therefore, should be interpreted cautiously.

One way to avoid this reliance on hypothetical distributions is to develop other validity generalization models which are independent of hypothetical distributions. As far as is known, no such models currently exist. The development of the covariance and regression slope models is a step in that direction. These

---

[1]The results for a sample size of 100 were very similar to the results for a sample size of 68 and are available from the first author.

two models are not totally free of artifact distributions, but their reliance on these distributions is substantially reduced. The application of the covariance model, for example, needs only a distribution for u, whereas the regression slope model requires only a distribution for r_{xx}. Despite this important practical advantage, the degree of accuracy associated with these two models, just as with the correlation model, still leaves something to be desired when a hypothetical distribution does not match the true distribution. The accuracy of estimation is especially poor for the covariance model when the hypothetical distribution differs from the true distribution of u. Yet the two new models are very accurate when the hypothetical and true distributions of artifacts are identical.

Although the covariance and regression slope models possess the above-mentioned advantages in relation to the correlation model, they are more restrictive in terms of their applicability. The use of the new models requires that the predictor and criterion metrics be comparable across validity studies. That is, the predictor and criterion instruments must measure the same constructs on metrics that are comparable across studies. In the context of the covariance and regression slope models, metric comparability is assured when the predictor (and criterion) scales are identical or differ by an additive constant across studies. In the correlation model, however, the predictor and criterion metrics can differ by a linear transformation across studies.

An important question for the practitioner is which model to use for studying validity generalization in situations where all three models are applicable. The choice among the competing models would be relatively simple if some or all distributions of artifacts are known. In view of the results of the present study, the covariance model is better than the other two models when the distribution of u is known; the regression slope model seems to have the advantage when only the distribution of r_{xx} is known. When all artifact distributions are known, any of the three models would accurately estimate the parameters of interest. When the artifact distributions are hypothetical, their goodness of fit may be assessed in relation to the simultaneous application of all three models. Thus, similar results from the different models would imply that the hypothetical distributions are similar to the true distributions and hence would confirm the reasonableness of the assumptions underlying the application of the correlation model.

Any rational choice among the competing models must at least be based on the following: (1) familiarity of the statistic used, (2) the number of hypothetical distributions needed, and (3) the accuracy of estimation under ideal as well as less than ideal conditions. From the point of view of familiarity, the correlation model has a lot to offer; it is probably the most commonly used statistic among personnel psychologists. The covariance and regression slope models, however, have an advantage over the correlation model in terms of the number of hypothetical distributions needed. The correlation and regression slope models appear to be more accurate than the covariance model, with a slight edge for the regression slope model. The results for Case IV, for which all the hypothetical distributions differed from the true distributions, show the correlation model to be the most accurate for estimating the mean, whereas the regression slope model is the most accurate for estimating the variance. The regression slope model appears to be slightly more accurate than the correlation model when the estimation of mean and variance is considered simultaneously.

The regression slope is probably not as well known as the correlation coefficient, but most personnel psychologists are reasonably familiar with it. The regression slope plays a very crucial role in Cleary's (1968) regression model for studying selection bias. Linn (1978) appears to prefer regression lines over correlation coefficients in the study of single-group and differential validity. Furthermore, when the error of measurement in the predictor is treated as a matter of fact in personnel selection (Hunter, Schmidt, & Pearlman, 1982, and Schmidt, Hunter, McKenzie, & Muldrow, 1979, seem to argue in favor of such a position), the regression slope model can be expressed without the r_{xx} term; the model then requires no hypothetical distribution for its implementation. In such a modified form, the regression slope model can be used to determine whether differences in the observed regression slopes are solely due to sampling. The procedures of Novick, Jackson, Thayer, and Cole (1972) and Rubin (1980) are also useful (and

probably more rigorous) in this context for testing the null hypothesis of no difference among the observed regression slopes. Unlike the procedures of Novick et al. and Rubin, the proposed procedure for the full regression model takes r_{xx} into account and provides an estimate of V_B.

Based on the results of the present study and the preceding discussion, it appears that the regression slope model offers an alternative approach for studying validity generalization when the predictor and criterion metrics are comparable across studies. This model appears to be somewhat more accurate than the correlation model, and it relies on only one hypothetical artifact distribution. An additional, minor theoretical advantage for the regression slope model over the correlation model is that its sampling error is uncorrelated with the population regression slope. In the correlation model, the sampling error is correlated with population correlation even though its effect on the results of a validity generalization study is considered to be very minimal (Hedges, 1982; Linn & Dunbar, 1982).

The preceding discussion suggests that both the regression slope model and the correlation model should receive psychometric and empirical scrutiny. There is a definite need for additional research to study the robustness of the correlation and regression slope models under various conditions. First, there is the need for determining the effect of artifact distributions (other than the rectangular distributions used in this study) on the accuracy of the various models. It is quite possible that the results of the present study may not be generalizable to other distributions of artifacts. Second, the two sample sizes used here (68 and 100 cases per validity study) appear to produce very similar results for the three models. Practitioners will need to search for the minimum sample size which would yield results that are as accurate as those obtained with sample sizes of 68 and 100 cases. Third, the important question of the minimum number of validity studies required for an accurate assessment of validity generalization must be thoroughly explored using all the available models. The present study used 100 validity studies for each validity generalization study; such a large number may be impractical for most empirically-based validity generalization studies. Finally and more importantly, the search must continue for new models which are free of artifact distributions and insensitive to metric differences across studies.

Appendix A

To develop the covariance model, it is necessary to first consider the effect of criterion unreliability, test unreliability, and range restriction on the covariance between a predictor and a criterion. According to classical test theory (Lord & Novick, 1968), the covariance between x and y is the same as the covariance between the true score of x (x_t) and the true score of y (y_t). Formally, this result can be expressed as

$$\sigma_{x_t y_t} = \sigma_{xy} \quad . \tag{A1}$$

To determine the effect of range restriction on the covariance between x and y, two assumptions are needed. The first assumption is that there is explicit selection on variable x; in the following, x' and y' denote the predictor and criterion, respectively, in the restricted group. The second assumption is that the regression coefficients are equal in the restricted and unrestricted groups. These conditions yield

$$\frac{\sigma_{xy}}{\sigma_x^2} = \frac{\sigma_{x'y'}}{\sigma_{x'}^2} \quad . \tag{A2}$$

By rearranging,

$$\sigma_{x'y'} = \sigma_{xy} \frac{\sigma_{x'}^2}{\sigma_x^2} = \sigma_{xy} u^2 \quad , \tag{A3}$$

where u is the ratio of restricted standard deviation to unrestricted standard deviation. The covariance model, therefore, can be written as

$$c_{xy} = \sigma_{xy}u^2 + e \quad , \tag{A4}$$

where c_{xy} is the sample covariance, and e is the sampling error associated with c_{xy}. Ignoring subscripts x and y, the above expression can be written as

$$c = \sigma u^2 + e \quad . \tag{A5}$$

The variance of Equation A5 can be expressed as

$$V_c = V(\sigma u^2 + e) \quad . \tag{A6}$$

Assuming that the sampling errors are uncorrelated with σu^2, the above equation can be expressed as

$$V_c = V(\sigma u^2) + V_e \quad . \tag{A7}$$

Equation A7 must still be simplified in order to estimate V_σ. One way to simplify Equation A7 is to assume that σ and u^2 are independent. Under this assumption, the above equation can be rewritten as

$$V_c = V_\sigma (M_{u^2}^2 + V_{u^2}) + M_\sigma^2 V_{u^2} + V_e \quad . \tag{A8}$$

Solving this equation for V_σ gives the expression

$$\hat{V}_\sigma = \frac{V_c - V_e - M_\sigma^2 V_{u^2}}{M_{u^2}^2 + V_{u^2}} \quad . \tag{A9}$$

Similarly, the mean of observed covariance can be written as

$$M_c = M(\sigma u^2) + M_e = M_\sigma M_{u^2} + M_e \quad . \tag{A10}$$

Assuming that the mean of sampling errors is zero, the resultant expression is

$$\hat{M}_\sigma = \frac{M_c}{M_{u^2}} \quad . \tag{A11}$$

Appendix B

As in the case of the description of the covariance model, the first consideration in developing the regression slope model is the effect of criterion and predictor reliabilities and range restrictions on the regression (slope) coefficient. It should be noted that this presentation will only deal with the slope, and not the intercept, of the regression line.

Let B_{yx_t} be the unattenuated and unrestricted regression coefficient in the population. This can also be written as

$$B_{yx_t} = \frac{\sigma_{x_t y_t}}{\sigma_{x_t}^2} \quad , \tag{B1}$$

where $\sigma_{x_t y_t}$ is the covariance between true scores on x (x_t) and true scores on y (y_t). Since the attenuated covariance is the same as the unattenuated covariance, the above equation can also be written as

$$B_{yx_t} = \frac{\sigma_{xy}}{\sigma_{x_t}^2} \quad . \tag{B2}$$

However, it is well known that $\sigma_x^2 r_{xx} = \sigma_{x_t}^2$ (Lord & Novick, 1968), where r_{xx} is the unrestricted reliability of x. Therefore,

$$B_{yx_t} = \frac{\sigma_{xy}}{\sigma_x^2 r_{xx}} = \frac{B_{yx}}{r_{xx}} \quad , \tag{B3}$$

which shows the effect of unreliability on the unattenuated regression coefficient. Since it is generally assumed that the restriction of range has no effect on the regression coefficient (Lord & Novick, 1968), the right side of the above equation also represents the attenuated and restricted regression coefficient in

the population. Now the regression coefficient (b_{yx}) obtained in a specific study can be written as

$$b_{yx} = B_{yx} + e = B_{y_t x_t} r_{xx} + e \quad ,$$ (B4)

where e is the sampling error. Ignoring subscripts y_t and x_t for the regression coefficient, Equation B4 can be rewritten as

$$b = B r_{xx} + e \quad ,$$ (B5)

the variance of which can be written as

$$V_b = V(B r_{xx}) + V_e \quad ,$$ (B6)

provided the sampling error is assumed uncorrelated with $B r_{xx}$. Since r_{xx} is not generally known for every prediction study, additional assumptions are necessary in order to compute the variance of $B r_{xx}$. If it is assumed that B and r_{xx} are uncorrelated, then Equation B6 can be written as

$$V_b = V_B (V_{r_{xx}} + M_{r_{xx}}^2) + M_B^2 V_{r_{xx}} + V_e \quad ,$$ (B7)

where $M_{r_{xx}}$ and $V_{r_{xx}}$ are the mean and variance of predictor reliability (r_{xx}), respectively. The same assumption also yields

$$M_b = M_B M_{r_{xx}} + M_e = M_B M_{r_{xx}} \quad ,$$ (B8)

since M_e is generally assumed to be zero. Solving Equation B8 for M_B gives

$$\hat{M}_B = \frac{M_b}{M_{r_{xx}}} \quad .$$ (B9)

Also, solving Equation B7 for V_B results in

$$\hat{V}_B = \frac{V_b - V_e - \left(\dfrac{M_b}{M_{r_{xx}}}\right)^2 V_{r_{xx}}}{V_{r_{xx}} + M_{r_{xx}}^2} \quad .$$ (B10)

Appendix C

Predictor and criterion scores for individual subjects were generated using the prespecified population parameters, a FORTRAN program for generating random numbers (IMSL, 1979), and a regression procedure to induce desired covariances and correlations. As previously noted, each population was defined by six parameters: ρ, r_{yy}, r_{xx}, u, σ_x, and σ_y. The random selection of samples was accomplished with the help of these six parameters, a standard normal cutting point Z_u corresponding to a given selection ratio (u), a sample size (n), and the following psychometric relationships for population data (Lord & Novick, 1968):

$$\sigma_{y_t} = \sigma_y (r_{yy})^{1/2} \quad ,$$ (C1)

$$\sigma_{x_t} = \sigma_x (r_{xx})^{1/2} \quad ,$$ (C2)

$$\hat{y}_t = \left(\frac{\rho_{x_t y_t} \sigma_{y_t}}{\sigma_{x_t}}\right) x_t \quad ,$$ (C3)

$$\sigma_{x_e} = (\sigma_x^2 - \sigma_{x_t}^2)^{1/2} \quad ,$$ (C4)

$$\sigma_{y_e} = (\sigma_y^2 - \sigma_{y_t}^2)^{1/2} \quad ,$$ (C5)

and

$$\sigma_{y_t | x_t} = \sigma_{y_t} (1 - \sigma_{x_t | y_t}^2)^{1/2} \quad ,$$ (C6)

where σ_{y_t} is the standard deviation of true criterion scores (y_t),

σ_y is the standard deviation of observed criterion scores (y),

σ_{x_t} is the standard deviation of true predictor scores (x_t),

σ_x is the standard deviation of observed predictor scores (x),

$\hat{y}_t$ is the predicted true score on the criterion,

$\rho_{x_t y_t}$ is the correlation between the true scores on the predictor (x_t) and the criterion (y_t),

σ_{x_e} is the standard deviation of error for the predictor,

σ_{y_e} is the standard deviation of error for the criterion, and

$\sigma_{y_t|x_t}$ is the standard deviation of y_t given x_t.

The remaining terms in the above equations have been previously defined.

From the psychometric relationships described above, the restricted, attenuated samples were generated in the following manner:

Generate x_t from normal $(0, \sigma_{x_t})$
Generate x_e from normal $(0, \sigma_{x_e})$
Generate $x = x_t + x_e$
Generate y_t from normal $(\hat{y}_t, \sigma_{y_t|x_t})$
Generate y_e from normal $(0, \sigma_{y_e})$
Generate $y = y_t + y_e$
If (x, y) is such that $Z_x > Z_u$, then keep the pair (x, y). Otherwise continue random generation of cases. Retain the first 68 (x, y) pairs for which $Z_x > Z_u$.

The same procedure was used for the sample size of 100 cases.

References

Callender, J. C. (1983, March). *Conducting validity generalization research based on correlations, regression slopes, and covariances*. Paper presented at the Industrial-Organizational and Organizational Behavior Graduate Student Convention, Chicago.

Callender, J. C., & Osburn, H. G. (1980). Development and test of a new model for validity generalization. *Journal of Applied Psychology, 65,* 543–558.

Callender, J. C., Osburn, H. G., Greener, J. M., & Ashworth, S. (1982). The multiplicative validity generalization model: Accuracy of estimates as a function of sample size, mean, variance and shape of the distribution of true validities. *Journal of Applied Psychology, 67,* 859–867.

Cleary, T. A. (1968). Test bias: Prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement, 5,* 115–124.

Draper, N. R., & Smith, H. (1966). *Applied regression analysis*. New York: Wiley.

Dribin, L. S. (1981). *A monte carlo study of the additive and multiplicative models of validity generalization*. Unpublished doctoral dissertation, Illinois Institute of Technology, Chicago.

Hedges, L. V. (1982). *Methodological problems in validity generalization*. Unpublished manuscript, The University of Chicago.

Hunter, J. E., Schmidt, F. L., & Pearlman, K. (1982). History and accuracy of validity generalization equations: A response to the Callender and Osburn reply. *Journal of Applied Psychology, 67,* 853–858.

IMSL Library. (1979). *International mathematical and statistical libraries*. Houston TX: Author.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics, Vol. 1* (4th ed.). New York: Macmillan.

Lent, R. H., Aurbach, H. A., & Levin, L. S. (1971). Predictors, criteria, and significant results. *Personnel Psychology, 24,* 519–533.

Linn, R. L. (1978). Single-group validity, differential validity, and differential prediction. *Journal of Applied Psychology, 63,* 507–512.

Linn, R. L., & Dunbar, S. B. (1982, November). *Validity generalization and predictive bias*. Paper presented at the Fourth Johns Hopkins University Na-

tional Symposium on Educational Research, ''Performance Assessment: The State of the Art,'' Washington DC.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.

Novick, M. R., Jackson, P. A., Thayer, D. T., & Cole, N. S. (1972). Estimating multiple regression in groups: A cross-validation study. *The British Journal of Mathematical and Statistical Psychology, 25*, 33–50.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology, 65*, 373–406.

Raju, N. S., & Burke, M. J. (1983). Two new procedures for studying validity generalization. *Journal of Applied Psychology, 68*, 382–395.

Rubin, D. B. (1980). Using empirical Bayes techniques in the law school validity studies. *Journal of the American Statistical Association, 75*, 801–816.

Schmidt, F. L., Gast-Rosenberg, I. E., & Hunter, J. E. (1980). Test of a new model of validity generalization: Results for computer programmers. *Journal of Applied*

Psychology, 65, 643–661.

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

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). The impact of valid selection procedures on work force productivity. *Journal of Applied Psychology, 64*, 609–626.

Author's Address

Send requests for reprints or further information to Nambury S. Raju, Department of Psychology, Illinois Institute of Technology, Chicago IL 60616.