A MONTE CARLO INVESTIGATION OF THE ACCURACY OF TWO MODELS FOR VALIDITY GENERALIZATION

A Dissertation

Presented To

the Faculty of the Department of Psychology

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

John C. Callender

December, 1978

ACKNOWLEDGMENTS

I would like to thank the members of my committee, Drs. H. G. Osburn (Chairman), James E. Campion, Scott Maxwell, and Roger Blakeney for their thoughtful attention to this project.

I would especially like to thank Dr. H. G. Osburn who contributed greatly to the development of this project long before it had grown into a dissertation. Without the benefit of Dr. Osburn's effective teaching throughout my graduate career, his clear thinking regarding the issues raised in this research, and his personal dedication through it all, I would surely not have had the ability or confidence to tackle the problems involved in this study.

I would also especially like to thank my wife, Glenda Jean Callender, for the help she has given and sacrifices she has made to make a successful graduate career possible. Her typing of the preliminary and final drafts of this dissertation is but one example of the help she has given.

Finally, I would like to acknowledge the help from Jack M. Greener in developing the simulation program which really launched the project and W. B. Bryant, Shell Oil Company, who provided encouragement as well as time and computer resources.

A MONTE CARLO INVESTIGATION OF THE ACCURACY OF TWO MODELS FOR VALIDITY GENERALIZATION

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Psychology

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

John C. Callender

December, 1978

ABSTRACT

A Multiplicative Model for the generalization of validity developed by Callender and Osburn (1978) was tested by Monte Carlo methods. The model provides a method for estimating the mean and variance of distributions of unattenuated unrestricted population validities by removing the effects of range restriction, criterion unreliability and chance sampling error. The model generally builds on concepts previously outlined by Schmidt and Hunter (1977).

The accuracy of the model was compared with the model originally proposed by Schmidt and Hunter on hypothetical infinite sample size cases and on small sample size cases (N= 30, 68, 200) by means of computer simulation of sample restricted attenuated data sets. The correctness of the computer simulation was verified analytically and empirically. Both models were than applied to estimate the known mean and variance of the population distributions of unrestricted unattenuated validities.

The Multiplicative Model was found to give reasonably accurate estimates of the mean and variance of the population correlations and of lower credibility values. The errors made by the model were generally conservative, tending to overestimate the variance of the population correlations. The Schmidt-Hunter model was less accurate and in some conditions made substantial nonconservative errors. It was concluded that the Multiplicative Model could be recommended for future use in validity generalization analysis.

A Monte Carlo study of the accuracy for a formula for the standard error of a correlation was also conducted. Results indicated that it was nearly as accurate as the more familiar Fisher's z formula.

The Multiplicative Model was applied to four distributions of validities summarized by Ghiselli (1966). It was found that the variation in true validities was greater in each case than that previously reported by Schmidt and Hunter (1977). The analysis indicated that validity generalization was supported for two of the four distributions.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

TABLE PAGE

CHAPTER I

INTRODUCTION

Determining the validity of psychological tests for predicting job performance has been a central concern for personnel psychologists for many years. A vast wealth of data in the form of criterion related validities has been amassed and at times summarized in various forms (Ghiselli, 1966; Lent, Aurbach, and Levin, 1971). A perplexing state of affairs has arisen when the observed correlations have been tallied into distributional form, namely that there typically is considerable variation in the magnitude of the observed correlations even for apparently similar jobs. Ghiselli (1966) pointed out that there are a number of explanations for how this could occur including the effects of chance sampling error in correlations, differences between studies in the reliability of the criterion, sample homogeneity differences as produced by restriction in range, variations in test administration procedures, the crudeness of systems for classifying jobs and tests, and differences in the nature of the requirements for nominally similar jobs. Unfortunately there has been no way to determine which of all these plausible explanations were truly most important.

Tremendous progress toward resolution of this question has recently been made by Schmidt and Hunter (1977), with the development of a model for validity generalization. Schmidt and Hunter have pointed out that three of the factors which produce variation in observed correlations are statistical artifacts. These factors are variation in criterion reliabilities, variation in restriction in range and chance sampling errors. Since the sampling error for correlations has been determined mathematically and the effects of criterion unreliability and restriction in range also follow known mathematical relationships, Schmidt and Hunter contend that it should be possible to determine how much variation is produced by these factors and how much is produced by other factors such as variation in the nature of jobs or other situational factors. The proposed analysis produces two outcomes. First there is an estimate of the variance over studies of the true unrestricted unattenuated validities. These are the validities of theoretical interest since they would govern the utility of the selection test for individual decision making. If the variance of these validities is low, it implies that the true validities are consistent from situation to situation and that the sources of variation other than the statistical artifacts are weak. If the variance of the true validities is large, then there is reason to believe that the sources of variance beyond the statistical artifacts are relatively strong. A second outcome that derives from the estimate of the variance of the true validities is a confidence interval for the range of true validities. Using the normal model, it is possible to determine above what minimum level of true validity the prepoderance of true validities fall. For example if 90 percent of the true validities are found to be above zero, then positive validity is to be expected in future studies. The center of this validity generalization confidence interval is the mean true unrestricted unattenuated validity and would be taken to be the most likely true validity to be found in future studies.

Schmidt and Hunter suggest that the validity generalization confidence interval, which is determined by the estimated mean and standard deviation of the true validities, should be thought of as a Bayesian "prior" distribution. This would provide a mechanism for summarizing existing studies, for projecting the likely range of results for future studies, and also for integrating the results of new studies with prior existing ones. As the number of studies entering into the prior increases, a more and more stable confidence interval would result. That is, the amount of change which could result in the confidence interval with the addition of each new study would decrease. Another way of stating the principle is that past validity studies could be taken into account, along with a current study, in reaching a decision about validity. The more studies which had been done previously, the greater the weight the prior would have relative to the new study. The result should be a better estimation and decision about validity because more information is taken into account than just the latest study.

Some comment should be made about the relationship between the two kinds of inferences which the model provides. The theoretical limit of the estimate of variance of the true validities is zero, which would imply perfect consistency of true validities across the job, test type, and situational factors represented in the analysis. It is not necessary that this variance be zero in order to conclude that substantial validity can be expected in other situations, however. If the average true validity is high, it is quite possible for most of the prior distribution to lie in the positive range even though there _is_ variation from these other sources.

The key idea that transforms the Bayesian validity generalization concept into a useable analysis is a mathematical procedure for determining the variance of the true unrestricted unattenuated validities and their mean. Schmidt and Hunter (1977) have proposed a complex but intuitively appealing procedure for doing this. Essentially the procedure involves the computation of variance due to sampling, variance due to differences in criterion reliabilities, and variance due to differences in degree of range restriction. These variances are then subtracted from the variance of the Fisher's z transformed values of the observed correlations. The residual variance is the estimate of the variance of the true validities in Fisher's z form. The reader should consult the appendices to Schmidt and Hunter (1977) for the details of the procedure.

The procedure was applied by Schmidt and Hunter to four distributions of validities summarized by Ghiselli (1966). The estimated standard deviations of the true validities for the four distributions range from .08 to .23 and it was indicated that most of the variance of each distribution of observed validities should be attributed to the statistical artifacts. Assumed distributions of criterion reliabilities and range restriction effects were used in arriving at these estimates. An estimated prior distribution and a 97.5% confidence limit was computed for each distribution based on an assumed average sample size of 68. This is the mean sample size reported in the review of validity studies by Lent, Aurbach, and Levin (1971). The results indicated that validity could be generalized for mechanical principles tests for mechanical repairman jobs and for intelligence tests for clerks. Validity generalization was not supported for finger dexterity tests for bench workers or for spatial relations tests for machine tenders.

One shortcoming of the paper by Schmidt and Hunter (1977) was a lack of discussion of the mathematical rationale and assumptions of the proposed procedures for estimating the mean and variance of the true validities. In order to put validity generalization on a firm and explicit mathematical basis, the following model and variance estimation procedures were developed by Callender and Osburn (1978).

Theoretically, there is a true (population) validity for each study. However, three primary factors operate to produce an observed correlation which deviates from the population true validity. These three factors are criterion unreliability, restriction in range, and chance sampling error. The first two factors reduce the observed validities below their true value. The effect of chance sampling error may be to either increase or decrease the observed correlation. The effects of criterion unreliability and range restriction can be specified mathematically. We begin with the true unattenuated unrestricted correlation, which is denoted by ρ_{xy_T} to show that it is a population correlation between predictor x and the true criterion scores y_T. Let $\rho_{yy'}$ denote the unrestricted reliability of the criterion used in the particular study. Then, if we had an infinite and unrestricted sample, the following equation would give the correlation between x and the observed criterion score, y:

$$\rho_{xy} = \rho_{xy_T} \sqrt{\rho_{yy'}} \qquad (1)$$

What this says is that ρ_{xy_T} is reduced by a factor of $\sqrt{\rho_{yy'}}$ in order to obtain the correlation between the predictor and the observed criterion scores.

The effect of restriction in range when there is explicit selection on the predictor is obtained by equation (2). In equation (2) $\rho_{x^*y^*}$ denotes the restricted correlation, U is the ratio of the restricted standard deviation of x to the unrestricted standard deviation x and ρ_{xy} is the unrestricted correlation of the predictor and criterion from equation (1).

$$\rho_{x^*y^*} = \rho_{xy} \frac{U}{\sqrt{\rho_{xy}^2 U^2 - \rho_{xy}^2 + 1}} \tag{2}$$

Equation (2) can easily be obtained from the usual formula for restriction in range due to explicit selection by solving for $\rho_{x^*y^*}$ instead of ρ_{xy}. In effect, equation (2) shows that ρ_{xy} is reduced by the factor which is a function of U and ρ_{xy}.

If we had an infinitely large sample, we would surely find that the attenuated restricted correlations would be $\rho_{x^*y^*}$. However, with actual samples, there will be some difference between $\rho_{x^*y^*}$ and the correlation $r_{x^*y^*}$ that is computed on the data. This is due to chance sampling error and we represent it by an additive variable in equation (3).

$$\rho_{x^*y^*} + e = r_{x^*y^*} \tag{3}$$

Equations (1) through (3) can now be combined into a single equation which we will call the Multiplicative Model for validity generalization. Some substitutions will make the subsequent derivations easier to follow. Let "a" be the factor by which ρ_{xy_T} was multiplied in equation (1) where:

$$a = \sqrt{\rho_{yy'}} \tag{4}$$

Let "c" be the factor by which ρ_{xy} was multiplied in equation (2) where:

$$c = \frac{U}{\sqrt{\rho_{xy}^2 U^2 - \rho_{xy}^2 + 1}} \tag{5}$$

Substituting in equations (1) and (2), we have:

$$\rho_{xy} = \rho_{xy_T} \cdot a \tag{1'}$$

and

$$\rho_{x*y*} = \rho_{xy} \cdot c \quad . \tag{2'}$$

Substituting ρ_{xy} from (1') into (2'), we have:

$$\rho_{x*y*} = \rho_{xy_T} \cdot a \cdot c \tag{6}$$

Substituting $\rho_{x^*y^*}$ from (6) into (3) and denoting ρ_{xy_T} simply as ρ and $r_{x^*y^*}$ as r, we have:

$$\rho \cdot a \cdot c + e = r \tag{7}$$

Equation (7) is the Multiplicative Model for validity generalization. All of the terms of the equation are assumed to be variable over validity studies. It is because they differ from study to study that different actual correlations, r, are obtained.

The objective in validity generalization is to infer, by pooling studies, what the mean and variance of the variable ρ is. The mean of ρ, m_ρ, gives the average level true validity. The variance of ρ, S_ρ^2, gives the amount of variance in true validity from study to study and determines the width of the validity generalization confidence interval. These quantities must be estimated from observed values in actual studies. We turn now to the problem of estimating these quantities when only the observed sample correlations, restricted criterion reliabilities, and range restriction standard deviation ratios are known. A thorough understanding of how the model works in the "forward" direction, working from true to observed validities, makes most of the steps needed to work backward from observed values self evident.

The true validity can be estimated by adjusting each r upward for restriction in range, using the usual formula (8).

$$\hat{\rho}_{xy} = \frac{r_{x^*y^*}}{\sqrt{r^2_{x^*y^*} + U^2 - U^2\, r^2_{x^*y^*}}} \tag{8}$$

Then $\hat{\rho}_{xy}$ is further adjusted upward for unrestricted criterion reliability by equation (9).

$$\hat{\rho}_{xy_T} = \frac{\hat{\rho}_{xy}}{\sqrt{\hat{\rho}_{yy'}}} \tag{9}$$

Since the unrestricted criterion reliability is used in (9), it is typically necessary to adjust an observed restricted criterion reliability upward for restriction first. This can be obtained using equation (10)[1] where $r_{y*y*'}$ is the observed restricted criterion reliability.

$$\hat{\rho}_{yy'} = \frac{U^2 \, r_{y*y*'} + r^2_{x*y*}(1 - U^2)}{U^2 + r^2_{x*y*}(1 - U^2)} \tag{10}$$

The other quantity which must be determined is S^2_{ρ}, the variance of the true validities. We begin by returning to equation (7) and noting that two terms are added together to obtain r. We assume that the chance sampling errors, e, are uncorrelated with the population restricted attenuated correlation, ρac. Consequently, the variances of each can be summed as in equation (11).

$$V(\rho \cdot a \cdot c) + V(e) = V(r) \tag{11}$$

In order to obtain an expression for the variance of ρ, S^2_{ρ}, it was necessary to derive a formula for the variance of the product of the three variables ρ, a, c, in terms of the means and variances of each. In obtaining the formula it was assumed that the variables are uncorrelated with each other. This is, the amount of attenuation, a, reduction from restriction, c, and true validity, ρ, change from study

1. This formula is based on equation (6.2.1) given by Lord and Novick (1968, P. 130). It is not equivalent to a formula previously given for this purpose by Schmidt, Hunter, and Urry (1976). The Schmidt, Hunter, and Urry formula can be shown to be erroneous.

to study independently of each other. The detailed steps of this derivation are given in Appendix A. The resulting expression for $S^2_{\rho ac}$ is given by equation (12):

$$S^2_{\rho ac} = S^2_{\rho}(S^2_a + M^2_a)\ (S^2_c + M^2_c) + M^2_{\rho}\ S^2_a(S^2_c + M^2_c) + M^2_{\rho}\ M^2_a\ S^2_c \quad . \tag{12}$$

This can be substituted for V(ρac) in equation (11). We also relabel V(r) as S^2_r and V (e) as S^2_e in equation (11) and solve for S^2_{ρ} . The result is equation (13):

$$S^2_{\rho} = \frac{S^2_r - S^2_e - M^2_{\rho}\ S^2_a(S^2_c + M^2_c) - M^2_{\rho}\ M^2_a\ S^2_c}{(S^2_a + M^2_a)\ (S^2_c + M^2_c)} \quad . \tag{13}$$

Equation (13) is the formula for the variance of the true validities, S^2_{ρ} , which is dictated by the Multiplicative Model and the mathematical assumption of independence.

The actual value of these variables, except r, is not known for each individual study. However, the necessary means and variances can still be reasonably estimated. For each study, an estimated a and c can be computed by the following sample equations, using the observed restriction ratio, U, the observed correlation r_{x*y*}, and the restricted observed criterion reliability r_{y*y*}'.

$$\hat{c} = \sqrt{U^2 + (1-U^2)\,r^2_{x*y*}} \tag{14}$$

$$\hat{a} = \sqrt{\frac{U^2 \; r_{y^*y^*} + (1 - U^2) \; r^2_{x^*y^*}}{U^2 + (1 - U^2) \; r^2_{x^*y^*}}} \tag{15}$$

$$\hat{M}_\rho = \frac{M_{r_{x^*y^*}}}{M_{\hat{a}} \cdot M_{\hat{c}}} \tag{16}$$

Once the value of $\hat{a}$ and $\hat{c}$ is computed for each study, the mean of each of the variance of $\hat{a}$ and $\hat{c}$ (across studies) can be computed and substituted into (13). The mean and variance of the observed correlations can be computed directly. The estimated mean unrestricted unattenuated validity can then be computed by equation (16). Finally an estimate of the variance of the sampling errors can be obtained by equation (17).

$$\hat{S}^2_e = \frac{(1 - r^2_{x^*y^*})^2}{N} \tag{17}$$

Equation (17) is equivalent to that given by Johnson and Kotz (1970, p.225) as a direct (as opposed to Fisher's z) estimate of the sampling variance of a correlation. It should be computed for each study, and then the mean of the individual S^2_e values can be taken as the estimate of sampling variance. All of these quantities should be substituted in equation (13) to obtain $\hat{S}^2_\rho$. The value of $\hat{M}_\rho$ is the center of the validity generalization confidence interval and $\hat{S}_\rho$ determines its width.

The estimation procedures dictated by the Multiplicative Model and those proposed by Schmidt and Hunter (1977) are different. Thus, it is to be expected that they would give different results when applied to the same data. Naturally, the question arises as to what those differences might be. An even more important question, however, is that of the actual accuracy of the two methods. In order to test the accuracy of the models, it is necessary to begin with hypothetical situations in which the true unattenuated unrestricted validities are known. These can then be modified by criterion unreliability, range restriction, and sampling error effects to produce observed validity coefficients. The question then is how accurately do the models estimate the variance of the true validities when working back from the observed validity distribution.

The accuracy tests which follow are based on the hypothetical case of infinite sample size for each validity study. This means that for both models, the variance component due to sampling error is assumed to be zero. Although this situation is not encountered in practice, it is one in which the models can be easily and instructively tested for accuracy. It is also a situation in which others can independently verify calculations by generating distributions of observed validities which are identical to those which we use.

In obtaining the results which follow, the procedures given by Schmidt and Hunter (1977) in their Appendix B for computing variance due to differences between studies in criterion reliability were followed precisely. The procedures for computing variance due to range restriction differences between studies given in Schmidt's and Hunter's (1977) Appendix C were followed precisely, except in Step 2. The following correct equation for an expected restricted correlation was used, instead of that given by Schmidt and Hunter:

$$r_i = \frac{U_i R}{\sqrt{U_i^2 R^2 - R^2 + 1}} \quad . \tag{18}$$

In equation (18), R is the unrestricted validity computed in Step 1 of Appendix C.

The first case which was examined was that of constant true validity but varying criterion unreliability and range restriction effects. The distribution of criterion reliabilities which was used is given in Table 1. The distribution of range restriction effects used is given in Table 2. These distributions are identical to those suggested by Schmidt and Hunter (1977).

Observed validity coefficients were generated by completely crossing the 100 values of criterion reliability with the 100 values of range restriction ratios. That is, each criterion reliability was paired with every range

Table 1

Hypothetical Distribution of Criterion Reliabilities

Reliability	Frequency
.90	3
.85	4
.80	6
.75	8
.70	10
.65	12
.60	14
.55	12
.50	10
.45	8
.40	6
.35	4
.30	3

Table 2

Hypothetical Distribution of Range Restriction Effects

Selection ratio (% selected)	Ratio of restricted to unrestricted predictor S.D.	Frequency
100	1.000	5
70	.701	11
60	.649	16
50	.603	18
40	.559	18
30	.515	16
20	.468	11
10	.411	5

restriction ratio, producting 10,000 pairs of criterion reliability and range restriction effects. Each of these was applied to a constant true validity of .1 to produce a distribution of 10,000 attenuated restricted coefficients. All variation in the resulting distribution of "observed" validities was produced by these factors alone. Both models were then applied to estimate the variance of the true validities. The correct result would be exactly zero, of course. This whole process was repeated for assumed true validities of .2 through .9, and the results are shown in Table 3.

In Table 3, it can be seen that the Multiplicative Model was closer to the correct value than the Schmidt-Hunter procedure in every case. The error made by the Multiplicative Model was always conservative, that is more variance was attributed to the true validities than actually existed. The Schmidt-Hunter procedure always made a nonconservative error by attributing more variance to the criterion reliability and range restriction effects than actually existed.

The last two columns of Table 3 show the effect of criterion reliability and range restriction on the true validities. The observed validities were generally about half as large as the true validity. The effects of criterion reliability and range restriction differences are reflected in the variance of the observed validities. It is interesting

Table 3

Estimated Variance of Distributions of 10,000 Identical True Validities

True validity	Estimated variance of true validities: Multiplicative model	Estimated variance of true validities: Schmidt-Hunter model	Mean of observed validities	Variance of observed validities
.1	.0000	-.0001	.046	.0001
.2	.0000	-.0002	.092	.0005
.3	.0001	-.0006	.139	.0010
.4	.0003	-.0011	.188	.0020
.5	.0007	-.0019	.240	.0031
.6	.0015	-.0031	.295	.0045
.7	.0029	-.0050	.355	.0064
.8	.0055	-.0084	.421	.0088
.9	.0100	-.0159	.497	.0123

Note. The correct variance of the true validities, either in raw or Fisher's *Z* form, is exactly 0. Negative values indicate model is accounting for too much variance.

to note that, even though the same criterion reliabilities and range restriction ratios were applied to each true validity, they produced greater variation in observed validities as the true validity increased. These values also help to put the model estimates of variance into perspective. Obviously, the absolute magnitude of the variance of observed correlations is a small number. This results from the fact that correlations happen to be on a scale of absolute values from 0 to 1. The model estimates of variance of the true correlations are also small numbers, not only because the correct value is exactly 0, but also because the variance which the models were to account for was quite small.

The next case which was studied was the reverse of the first. That is, what sort of answers do the two models give when it is known that all of the variation in the observed correlations resulted only from the variation in the true correlations, but not from attenuation or restriction? The following procedure was used for this test. A distribution of true validities was constructed by taking each of the correlation values from .00 to .99 at an interval of .01. Then, it was assumed that there was no restriction and no attenuation, i.e., $\rho_{yy}' = 1.0$ and $U = 1.0$ in every case. These values were applied to the 100 correlations. In this case, the observed attenuated restricted correlations were identical to the true correlations. Eight other distributions were similarly constructed. The nine cases were based on each possible combination of constant criterion reliabilities of 1.0, .60, and .30, and of constant range restriction effects based on selection percents of 100, 50, and 10. These values were chosen so the top, middle, and lowest degrees of attenuation and restriction would be represented, even though no variance was actually attributable to them in any of the observed correlation distributions.

The results appear in Table 4. When there was no range restriction (100 percent selected), the Multiplicative Model gave exactly the correct variance in each case. The Schmidt-Hunter model was correct only in the case of no range restriction and perfect criterion reliability. As the criterion reliability decreased, the amount of variance attributed by the Schmidt-Hunter method to the true correlations falls off rapidly. Most of the variance is being attributed by the Schmidt-Hunter model to range restriction and criterion reliability effects, when in fact, none of it should have been.

The introduction of range restriction produced inaccuracies in both models. The Multiplicative Model tended to overestimate the variance of the true validities when there was range restriction. Again, this is a conservative error. The nonconservative error of the Schmidt-Hunter model of attributing too little variance to

Table 4

Model Estimates of Variance of True Validities From .00 to .99 for Selected Combinations of Constant Criterion Reliability and Range Restriction

Model	Range restriction (% selected)	Criterion reliability 1.00	.60	.30
Multiplicative	100	.083	.083	.083
(actual variance = .083)	50	.133	.106	.093
	10	.176	.117	.097
Schmidt-Hunter	100	.302	.080	.030
(actual variance = .302)	50	.174	.034	.012
	10	.103	.016	.005

Note. The Multiplicative Model should be compared with the variance of the true validities, which is .083. The Schmidt-Hunter Model, because of its use of Fisher's z conversions, must be compared with the variance of the Fisher's z values of the true validities, which is .302.

the true correlations was extreme when there was both range restriction and imperfect criterion reliability. For example, when the criterion reliability was .60, that is, the average reliability in the distribution suggested by Schmidt and Hunter and the range restriction was 50 percent selected, also, the middle of the distribution suggested by Schmidt and Hunter, the Schmidt-Hunter model estimated variance was .034 which is only about one-tenth of the actual variance. The Multiplicative Model estimate of .106 was considerably closer to the correct value of .083 in this case. In general, we conclude from Table 4 that the Multiplicative Model was always more accurate than the Schmidt-Hunter procedure and that the errors were conservative for the Multiplicative and nonconservative for the Schmidt-Hunter model.

A comment should be made as to the reason for the inaccuracy of some of the Multiplicative Model results in Table 4. It will be recalled that one of the mathematical assumptions made in deriving the Multiplicative Model estimate of true correlation variance was that the three variables, ρ_{xy_T}, a, and c, were uncorrelated across studies. The value of c depends not only on U, the ratio of standard deviations, but also on the attenuated but unrestricted validity, ρ_{xy}. Because both "a" and U were constant, a perfect correlation was introduced between the true correlation ρ_{xy_T}, and the value of c, in each case involving other than 100 percent selection. Thus, Table 4 clearly shows that the effect of the violation of the assumption of independence of ρ_{xy_T} and c was to produce a moderate but conservative error in estimating the true correlation variance. Of course, under more realistic circumstances there would be much less than a perfect correlation between the true correlation and the restriction factor, c, so that the error introduced would be expected to be smaller than that found in Table 4.

The last situation that was investigated was the most realistic, in that all three factors, true validity, criterion reliability, and range restriction ratio were allowed to vary. An approximately normal distribution of true validities with a mean of .50 and ranging from .06 to .94 was constructed for this test. The distribution is the "wide" one in Table 5. (Both a wide and a narrow distribution appear in the table. Both were used in some of the subsequent tests.) A set of 100 observed correlations was generated by randomly selecting (without replacement) a true validity from the wide distribution of Table 5, a criterion reliability from Table 1 and a range restriction ratio from Table 2. The process was continued until all 100 values of each table were used up. The resulting distribution of observed correlations was then run through both models. To provide replications, the entire procedure was repeated ten times. Thus, for each

Table 5

Hypothetical Normal Distributions of
True Unattenuated Unrestricted Validities

Wide Distribution	Narrow Distribution	Frequency
.94	.61	1
.90	.60	1
.86	.59	1
.82	.58	2
.78	.57	2
.74	.56	3
.70	.55	5
.66	.54	6
.62	.53	7
.58	.52	8
.54	.51	9
.50	.50	10
.46	.49	9
.42	.48	8
.38	.47	7
.34	.46	6
.30	.45	5
.26	.44	3
.22	.43	2
.18	.42	2
.14	.41	1
.10	.40	1
.06	.39	1

replication, there were 100 observed correlations. The ten distributions of correlations were not identical because the way in which true validities, criterion reliabilities, and restriction ratios were matched differed randomly from replication to replication.

Table 6 shows the results of applying both models to these distributions. In every case, the Multiplicative Model overestimates the variance of the true correlations, with the amount of overestimation ranging from 10 to 58 percent. On the other hand, the Schmidt-Hunter procedure underestimates the actual variance in each case. These estimates are typically only 15 percent of what the actual variance was-rather extreme degree of underestimation. Again, we note that the error made by the Multiplicative Model was always a conservative one, that is, it could have produced a conclusion that validity could not be generalized when in fact validity could be generalized. The Schmidt-Hunter model always produced a nonconservative error.

Purpose

The superior accuracy of the Multiplicative Model on infinite and very large samples is rather well established by the analyses reported so far. However, there is ample reason for further tests of the models on smaller sample sizes. In the Multiplicative Model, an estimate of sampling error is used which is not as well known and could be less

Table 6

Model Estimates of Variance of Wide Distribution of True Validities From Table 5 to Which Criterion Reliabilities and Range Restriction Ratios From Tables 1 and 2 Were Randomly Assigned

Replication	Estimate of variance: Multiplicative model (actual variance is .031)	Estimate of variance: Schmidt-Hunter model (actual variance is .077)	Mean of observed validities	Variance of observed validities
1	.036	.010	.242	.012
2	.043	.011	.246	.014
3	.043	.013	.245	.014
4	.038	.010	.245	.012
5	.044	.011	.248	.014
6	.038	.009	.245	.012
7	.040	.010	.247	.013
8	.034	.008	.245	.011
9	.042	.011	.247	.014
10	.049	.014	.249	.015

Note. The Multiplicative Model should be compared with the variance of the true validities, which is .031. The Schmidt-Hunter Model, because of its use of Fisher's Z conversions, must be compared with the variance of the Fisher's Z values of the true validities, which is .077.

accurate than the Fisher's z used by Schmidt and Hunter. Since a numerical characterization of the accuracy of this sampling error formula is not available, an empirical Monte Carlo study of its accuracy should be made.

Secondly, an issue raised by Frank Schmidt (personal communication) is that the variance of the estimated restriction and attenuation factors would be greater than the variance of the population factors because of sampling error, and the direction of any bias introduced by this would be nonconservative in the Multiplicative Model equation (13). Whether or not overestimation of the variance of restriction and attenuation factors would have sufficient strength to overcome the already noted conservative bias in the Multiplicative Model estimate of the variance of the validities that is introduced by the positive correlation between the restriction factors and true validities remains to be seen. Thus there is need for small sample Monte Carlo tests of the accuracies of the two models.

Finally, there is some doubt about the substantive conclusions reached by Schmidt and Hunter (1977), in view of the poor performance of their model on the accuracy tests. Therefore, the distributions reported by Ghiselli (1966) and analyzed by Schmidt and Hunter (1977) should be reanalyzed by the Multiplicative Model, if it is found to be reasonably accurate.

CHAPTER II

METHOD

Simulation of Restricted Attenuated Sample Distributions

A FORTRAN computer program developed by Jack M. Greener and H. G. Osburn and modified by the author was used to generate sample bivariate distributions for studies of the accuracy of the correlation standard error formula and the accuracy of the two validity generalization models. The computer program utilizes the random normal generator N(0,1) from the IMSL statistical package to produce simulated data points which are accumulated into a distribution.

Each simulated data point produced had four components:

(1) a predictor score, x

(2) a true criterion score, y_T

(3) an error of criterion measurement, y_E

(4) an observed criterion score, $y = y_T + y_E$

A regression based strategy was used to generate the y_T component. First an x value was sampled by the random normal (0,1) generator. Then a predicted y value, Y_p, was generated from the x according to the population regression of y_T on x. The slope of this regression is:

$$B = R_{xy_T} \cdot \frac{S_{y_T}}{S_x} . \tag{19}$$

The R_{xy_T} value is the chosen true unrestricted unattenuated validity. The population standard deviations of x and of y (S_x , S_y) were fixed at 1.0. The value of S_{y_T}, the standard deviation of true criterion scores, was then obtained from the well known reliability relationship:

$$\rho_{yy'} = \frac{S^2_{y_T}}{S^2_y} \quad . \tag{20}$$

Solving for S_{y_T}, we have $S_{y_T} = S_y \sqrt{\rho yy'}$. Since S_y =1, the population standard deviation of the true criterion scores is simply the square root of the chosen population criterion reliability. The value of x was then multiplied by B, the slope of the regression of y_T on x, in order to get the predicted true criterion score, y_p. The expected population variance of the y_p values is:

$$B^2 \; S^2_x = R^2_{xy_T} \; \frac{S^2_{y_T}}{S^2_x} \; S^2_x = R^2_{xy_T} \quad S^2_{y_T} \quad . \tag{21}$$

since a constant, B, is multiplied times a random variable, x, whose sigma is 1.0.

Of course the true criterion scores, y_T, are not perfectly predictable from x. Therefore an error of predicting y_T from x had to be added to the y_p values to obtain y_T values. The error in predicting y_T from x was obtained from the well known formula for the conditional standard deviation of y_T given x:

$$s_{y_T} \cdot x = s_{y_T} \sqrt{1 - R^2_{xy_T}} \quad . \tag{22}$$

Another random normal value was sampled and multiplied by the above conditional standard deviation to obtain the error which, when added to y_p, produced a true criterion score, y_T. The expected variance of these error scores in the population is $s^2_{y_T \cdot x}$ since they are the product of the constant ($s_{y_T \cdot x}$) and the random normal (0,1) variable. The expected variance of the y_T values is simply the sum of the variance of the y_p values and the expected variance of the error in predicting y_T from x, because of their having been generated from independent normal variables. Thus we have that:

$$s^2_{y_T} = s^2_x B^2 + s^2_{y_T} \cdot x = s^2_x \; R^2_{xy_T} \; \frac{s^2_{y_T}}{s^2_x} + s^2_{y_T} \; (1-R^2_{xy_T}) \quad . \tag{23}$$

The right hand member of the above equation simplifies to $s^2_{y_T}$, thus confirming that the procedure for generating y_T values will produce the appropriate amount of variance of y_T's. The procedure also insures that the expected slope of the regression of y_T's on x is correct.

After an (x, y_T) pair was generated, direct range restriction on the predictor x was introduced. For the particular ratio of restricted to unrestricted predictor standard deviations which was chosen, the corresponding lower cutoff in the distribution of x was determined. The sampled (x, y_T) values in which x fell below this cutoff were discarded. This process was continued until the number of retained (x, y_T) values met the particular sample size requirement.

Finally an error of measurement on the criterion, y_E, was generated based on the well known formula for the standard error of measurement $S_E = S_y \sqrt{1 - \rho_{yy'}}$. Another random normal variable was sampled and multiplied by S_E to obtain y_E. The expected variance of the y_E values thus obtained is equal to S_E^2, since the constant S_E is multiplied times a random normal (0,1) variable.

Since the variables y_T and y_E were generated independently and the observed criterion score y is simply their sum we have that:

$$S_y^2 = S_{y_T}^2 + S_{y_E}^2 \quad . \tag{24}$$

We have already shown that the procedure would generate the correct $S_{y_T}^2$. Since

$$S_{y_E}^2 = S_E^2 = S_y^2 \, (1 - \rho_{yy'}) = S_y^2 \left[1 - \frac{S_{y_T}^2}{S_y^2} \right] = S_y^2 - S_{y_T}^2 \tag{25}$$

we have that

$$s_y^2 = s_{y_T}^2 + (s_y^2 - s_{y_T}^2) = s_y^2 = 1.0 \quad . \tag{26}$$

Thus the expected variance of the generated observed criterion scores, y, is the same as the fixed population variance of y.

In sum, the procedure for generating data maintains the correct relationships between x, y_T, y_E, and y and generates appropriate variances for each. It conforms with the requirement for independence of y_E from x and from y_T. It should also be apparent that the variation of y_T and y will be appropriately reduced when x is restricted because of the regression relationship between x and y_T.

Procedure for Study of Accuracy of Simulated Distributions

One way to confirm the accuracy of the computer generated distributions is to compare the population and sample values for very large samples. With sufficiently large sample size, sample values should converge very closely to population values, if the simulation is correct. A large sample (N=5000) test was conducted by generating 100 restricted attenuated distributions. A population unrestricted unattenuated correlation was randomly selected (without replacement) from the wide distribution of Table 5. An unrestricted criterion reliability was randomly selected (without replacement) from Table 1 and a ratio of restricted to unrestricted predictor standard deviations was similarly selected from Table 2. This process continued until all one hundred values in the tables were exhausted. The three parameters determine the theoretical expected restricted attenuated correlation between the predictor x and criterion y. This can be compared with the sample restricted attenuated correlation computed on the computer generated data points. This test is thorough in that a wide range of the three parameters was employed with a corresponding wide range in the expected restricted attenuated correlations.

Procedure for Study of the Accuracy of the Formula for the Standard Error of Correlations

The accuracy of the estimate of the standard error of a correlation, equation (17), was studied by generating sample data, computing the correlation, and then computing the standard deviation over sample distributions of the correlations. This should be close to the standard error computed from the formula, if it is accurate. Since the formula is to be used on restricted attenuated distributions, its accuracy was assessed in this context.

For each of the 45 combinations of population parameters resulting from the true validities of .1, .3, .5, .7, and .9, the criterion reliabilities of .6, .8, and 1.0, and the restriction ratios of U = .411, .603, and 1.0, a total of 400 restricted attenuated sample distributions of size 68 each was generated. The sample correlation and the Fisher's z transform of the sample correlation was computed for each distribution. Thus it was possible to compare the estimated standard error based on the expected restricted attenuated correlation with the actual standard deviation of the 400 r's and also to compare the standard deviation of the Fisher's z values of those r's with the well known standard error of Fisher's z converted correlations. In this case the Fisher's z standard error was $\frac{1}{\sqrt{N-3}}$ or .124.

This test of accuracy is thorough in the sense that a wide range of expected restricted attenuated correlations was studied on a small but reasonably representative sample size. Since the Multiplicative Model employs actual sample correlations (rather than population based expected restricted attenuated correlation) to estimate the standard error, a further comparison between these two standard error estimates was made by simulating 400 distributions each for the true validities of .1, .3, .5, .7, and .9 (without attenuation or restriction). A sample based standard error was computed from each distribution. The average sample estimated standard error, over the 400 estimates can then be compared with standard error computed from the actual population true validities, to determine the effect, if any, of substituting observed sample correlations for population parameters.

Design of Study of Accuracy of the Multiplicative and Schmidt-Hunter Models

The study of accuracy of the models included three levels of sample size (N = 30, 68, and 200) crossed with the two distributions of true validities from Table 5. For each of the six combinations of sample size and width of true validity distribution, 25 replications of a validity generalization analysis (by both models) was done. In each replication 100 simulated restricted attenuated distributions were generated by randomly selecting true validities from the relevant distribution in Table 5, criterion reliabilities from Table 1, and range restrictions from Table 2. Values were selected from the tables without replacement until the population parameters for 100 simulated distributions had been determined. The actual variance of the true validities and the actual lower 90% confidence value from the true validity distributions could then be compared with the estimates of these critical values by both models. The mean and standard deviation over the 25 replications of the model estimates can then be used to draw conclusions about both the accuracy (or bias) of each model and also the stability from replication to replication of the model estimates. Also, the accuracy and stability of each model can be compared across the different sample size and true validity distribution combinations.

Procedure for Application of the Multiplicative Model to the Ghiselli (1966) Distributions

The observed validity distributions reported by Ghiselli (1966, p. 29) are presented in bar graph form. The graphs were translated into frequency distributions, which are given in Table 7. Because of the crudeness of scaling of the bar graphs, the frequencies reported in Table 7 may differ somewhat from those which Schmidt and Hunter (1977) may have inferred. The possible occasional difference in interpolation from the graph would be expected to be too small to materially affect the mean and standard deviation of the distributions, however.

As done previously by Schmidt and Hunter (1977), the criterion reliabilities and range restrictions from Tables 1 and 2 were applied in the model. The assumed sample size was 68.

The components of equation (13) for the Multiplicative Model were computed as follows:

(1) The mean and variance of the distribution of observed correlations was computed.

(2) Attenuation factors were computed by taking the square root of each criterion reliability in Table 1. The mean and variance of the attenuation factors was computed.

(3) Restriction factors were computed by substituting each ratio of restricted to unrestricted predictor s.d. (U) from Table 2 into equation (14). The mean observed

Table 7

Distributions of Observed Validities From Ghiselli (1966, p. 29)

	Job/Test Type			
Validity	Mechanical repairmen/ Mechanical principles[a]	General clerks/ Intelligence[b]	Bench workers/ Finger dexterity[b]	Machine tenders/ Spatial relations[b]
-.575	-	-	1	1
-.525	-	-	-	-
-.475	-	-	-	-
-.425	-	-	-	-
-.375	-	1	1	3
-.325	-	-	1	-
-.275	2	-	3	3
-.225	-	-	5	2
-.175	-	-	5	6
-.125	-	2	5	-
-.075	2	-	5	19
-.025	-	-	1	9
.025	-	4	21	13
.075	2	2	15	6
.125	7	2	18	8
.175	4	6	11	7
.225	7	10	10	6
.275	12	4	23	4
.325	15	4	19	2
.375	10	2	9	4
.425	15	10	10	2
.475	16	7	9	2
.525	6	5	8	2
.575	4	4	3	-
.625	2	1	1	-
.675	2	4	-	-
.725	3	2	5	-
.775	2	1	1	-
.825	-	1	-	-
.875	-	-	1	-
Mean	.362	.350	.205	.048
S.D.	.186	.229	.239	.208
N	111	72	191	99

[a]Training criteria.
[b]Proficiency criteria.

correlation was taken as the value of $r_{x^*y^*}$ in the equation for each restriction factor. The mean and variance of the restriction factors was computed.

(4) The sampling variance estimate was computed by equation (17) for each value in the distribution of observed correlations using the assumed sample size of 68. Then the mean estimated sampling variance was computed for substitution in equation (13).

CHAPTER III

RESULTS

Accuracy of Simulation for Large Samples

Table 8 shows how the sample restricted attenuated correlation, estimated attenuation and estimated restriction factors compared with the corresponding population values. Only the first 20 of the 100 distributions are included, as they are quite adequate to show the extremely close correspondence of the sample and population values over a wide range of parameters. When averaged over all 100 distributions, there was only a difference of .001 between the population and sample values.

Accuracy of Formula for the Standard Error of Correlations

The results of this study are shown in Table 9. The standard errors are typically quite close to the computed standard deviation of the observed correlations. Also, the standard deviations of the Fisher's z values are typically quite close to the expected value of .124. The observed deviations from the expected value probably result from variations due to sampling of only 400 correlations. This interpretation is supported by the fact that the average difference between the standard errors and standard deviations was only .001 and the fact that the average Fisher's z value over all 18,000 simulated distributions was equal to its expected value of .124.

Table 8

Simulation Accuracy For Sample Size 5000

Distribution	Population unrestricted unattenuated validity	Population unrestricted criterion reliability	Ratio of restricted to unrestricted predictor S.D.	Expected restricted attenuated validity	Sample restricted attenuated validity	Population attenuation factor	Sample estimated attenuation factor	Population restriction factor	Sample estimated restriction factor
1	.58	.90	.603	.369	.364	.949	.948	.671	.663
2	.54	.75	.515	.263	.288	.866	.875	.562	.570
3	.58	.55	.559	.257	.251	.742	.736	.598	.592
4	.18	.85	.468	.079	.069	.922	.920	.473	.469
5	.34	.45	.468	.109	.095	.671	.676	.478	.471
6	.70	.35	.515	.228	.245	.592	.609	.551	.559
7	.66	.60	.559	.316	.322	.775	.781	.617	.613
8	.34	.70	1.000	.284	.280	.837	.831	1.000	1.006
9	.90	.40	.701	.437	.432	.632	.631	.767	.760
10	.38	.50	.603	.166	.158	.707	.702	.617	.613
11	.50	.55	.468	.184	.163	.742	.746	.495	.487
12	.30	.60	.649	.153	.137	.775	.776	.659	.657
13	.50	.50	.701	.256	.254	.707	.710	.724	.718
14	.22	.65	.559	.100	.075	.806	.806	.565	.559
15	.50	.65	1.000	.403	.424	.806	.813	1.000	1.012
16	.66	.60	.559	.316	.305	.775	.775	.617	.617
17	.74	.60	.649	.413	.401	.775	.768	.721	.718
18	.54	.65	.559	.261	.263	.806	.801	.599	.603
19	.42	.60	.649	.218	.209	.775	.773	.670	.669
20	.82	.80	.515	.486	.497	.894	.896	.662	.672
Mean	.500	.600	.595	.244	.245	.769	.768	.629	.630
S.D.	.175	.146	.121	.106	.110	.096	.099	.118	.120

Note. Mean and S.D. were computed on all 100 simulated distributions, only the first 20 of which are shown in the table.

Table 9

Accuracy of Estimate of Standard Error of r for N = 68

Population unrestricted unattenuated correlation	Population unrestricted criterion reliability	Ratio of restricted to unrestricted predictor S.D.	Expected restricted attenuated correlation	Standard error[a]	Standard deviation of sample correlations[b]	Standard deviation of Fisher's z conversion of sample correaltions[c]	Mean standard error[d]
.90	1.00	1.000	.900	.023	.024	.118	.023
		.603	.780	.048	.052	.130	
		.411	.647	.070	.081	.135	
	.80	1.000	.805	.043	.043	.119	
		.603	.633	.073	.076	.127	
		.411	.487	.093	.099	.130	
	.60	1.000	.697	.062	.062	.121	
		.603	.506	.090	.087	.116	
		.411	.371	.105	.107	.127	
.70	1.00	1.000	.700	.062	.064	.124	.062
		.603	.509	.090	.093	.127	
		.411	.374	.104	.109	.129	
	.80	1.000	.626	.074	.073	.119	
		.603	.436	.098	.098	.122	
		.411	.313	.109	.115	.128	
	.60	1.000	.542	.086	.091	.128	
		.603	.363	.105	.108	.125	
		.411	.256	.113	.123	.134	
.50	1.00	1.000	.500	.091	.094	.126	.089
		.603	.329	.108	.111	.125	
		.411	.231	.115	.117	.125	
	.80	1.000	.447	.097	.097	.123	
		.603	.289	.111	.116	.127	
		.411	.201	.116	.123	.131	
	.60	1.000	.387	.103	.100	.119	
		.603	.246	.114	.116	.124	
		.411	.170	.118	.122	.127	
.30	1.00	1.000	.300	.110	.115	.127	.109
		.603	.186	.117	.121	.127	
		.411	.128	.119	.114	.118	
	.80	1.000	.268	.113	.120	.132	
		.603	.166	.118	.120	.125	
		.411	.114	.120	.117	.121	
	.60	1.000	.232	.115	.114	.122	
		.603	.143	.119	.117	.120	
		.411	.098	.120	.116	.119	
.10	1.00	1.000	.100	.120	.118	.121	.118
		.603	.060	.121	.123	.125	
		.411	.041	.121	.123	.125	
	.80	1.000	.089	.120	.117	.119	
		.603	.054	.121	.113	.115	
		.411	.037	.121	.113	.114	
	.60	1.000	.077	.121	.120	.122	
		.603	.047	.121	.119	.121	
		.411	.032	.121	.118	.120	
Mean: .50	.80	.671	.331	.101	.102	.124	

[a]Computed with the expected restricted attenuated correlation and the standard error formula $SE_r = \frac{1-r^2}{\sqrt{N}}$.

[b]Based on 400 samples.

[c]The Fisher's z standard error is $\frac{1}{\sqrt{N-3}} = .124$

[d]Computed with actual sample correlations input to above SE_r formula. The mean is computed over the 400 such standard error estimates. Standard errors were computed only on unrestricted unattenuated correlations (.1, .3, .5, .7, .9) due to expense.

The same simulation procedure was used to generate the sample distributions for this study and for the validity generalization model tests. The results of this study further confirm the accuracy of the simulation. A faulty simulation would hardly be expected to produce distributions which agree so perfectly with the Fisher's z standard error.

<u>Accuracy of Simulation on Small Samples</u>

Previous results with large sample size (N = 5000) demonstrated the accuracy of the simulation program. In conducting the study of the accuracy of the validity generalization models on small samples, it was possible to collect further small sample information about the correspondence of population and small sample distributions. Tables 10, 11, and 12 report the results of some of the key variables for the sample sizes of 30, 68, and 200 respectively. Each table includes the results for both the narrow and the wide distribution of true validities. Looking first at Table 10 (sample size = 30) it is noted that, on the average over replications, the sample estimated means for the unrestricted unattenuated correlations, for the restricted attenuated correlations, for the unrestricted criterion reliabilities, for the attenuation factors, and for the restriction factors are extremely close to the population mean values for both the wide and narrow distributions.

Table 10

Mean Over 25 Replications of Mean and S.D. of Population and Sample Values for Sample Size 30

Distribution of true validities	Variable	Means computed on the 100 distributions in each replication: Population values	Means computed on the 100 distributions in each replication: Sample values	Standard deviations computed on the 100 distributions in each replication: Population values	Standard deviations computed on the 100 distributions in each replication: Sample values
Narrow[a]	Unrestricted unattenuated r	.500	.495[b]	.045	-[c]
	Restricted attenuated r	.240	.240	.060	.181
	Unrestricted criterion reliability	.600	.618	.146	.180
	Attenuation factor	.769	.776	.096	.124
	Restriction factor	.624	.623	.116	.144
Wide[a]	Unrestricted unattenuated r	.500	.495[b]	.175	-[c]
	Restricted attenuated r	.246	.241	.114	.206
	Unrestricted criterion reliability	.600	.615	.146	.177
	Attenuation factor	.769	.773	.096	.129
	Restriction factor	.629	.629	.118	.145

[a]From Table 5.

[b]For each replication the mean sample restricted attenuated r over the 100 distributions was divided by the product of the mean sample estimated restriction factor and the mean sample estimated attenuation factor. The entry is the mean of this estimated mean true validity over the 25 replications.

[c]There was no standard deviation as only a single value was produced for each replication as explained in "b" above.

Table 11

Mean Over 25 Replications of Mean and S.D. of Population and Sample Values for Sample Size 68

Distribution of true validities	Variable	Means computed on the 100 distributions in each replication: Population values	Sample values	Standard deviations computed on the 100 distributions in each replication: Population values	Sample values
Narrow[a]	Unrestricted unattenuated r	.500	.499[b]	.045	-[c]
	Restricted attenuated r	.240	.241	.059	.125
	Unrestricted criterion reliability	.600	.609	.146	.163
	Attenuation factor	.769	.772	.096	.111
	Restriction factor	.624	.624	.116	.131
Wide[a]	Unrestricted unattenuated r	.500	.505[b]	.175	-[c]
	Restricted attenuated r	.246	.243	.114	.159
	Unrestricted criterion reliability	.600	.605	.146	.165
	Attenuation factor	.769	.769	.096	.113
	Restriction factor	.629	.626	.118	.133

[a]From Table 5.

[b]For each replication the mean sample restricted attenuated r over the 100 distributions was divided by the product of the mean sample estimated restriction factor and the mean sample estimated attenuation factor. The entry is the mean of this estimated mean true validity over the 25 replications.

[c]There was no standard deviation as only a single value was produced for each replication as explained in b above.

Table 12

Mean Over 25 Replications of Mean and S.D. of Population and Sample Values for Sample Size 200

Distribution of true validities	Variable	Means computed on the 100 distributions in each replication: Population values	Sample values	Standard deviations computed on the 100 distributions in each replication: Population values	Sample values
Narrow[a]	Unrestricted unattenuated r	.500	.498[b]	.045	-[c]
	Restricted attenuated r	.240	.240	.060	.089
	Unrestricted criterion reliability	.600	.603	.146	.152
	Attenuation factor	.769	.769	.092	.113
	Restriction factor	.624	.624	.116	.121
Wide[a]	Unrestricted unattenuated r	.500	.508[b]	.175	-[c]
	Restricted attenuated r	.246	.246	.115	.132
	Unrestricted criterion reliability	.600	.604	.146	.153
	Attenuation factor	.769	.770	.096	.102
	Restriction factor	.629	.629	.118	.123

[a]From Table 5.
[b]For each replication the mean sample restricted attenuated r over the 100 distributions was divided by the product of the mean sample estimated restriction factor and the mean sample estimated attenuation factor. The entry is the mean of this estimated mean true validity over the 25 replications.
[c]There was no standard deviation as only a single value was produced for each replication as explained in b above.

Turning to the mean standard deviations, it should be noted that the standard deviations of the sample values are always greater than the standard deviations of the population values. This is a direct result of the effects of sampling error which enters into the sample values. The amount of increase in the sample value standard deviations of the restriction and attenuation factors over the population is quite similar for the narrow and wide distributions.

A similar pattern of results is noted in Tables 11 and 12 for sample sizes 68 and 200. The only difference is that, because of the reduction of sampling error with increased N, the sample standard deviations are not as discrepant from the population standard deviations. For sample size 200, they are remarkably similar.

Frank Schmidt (personal communication) has pointed out that the greater variance of sample estimated restriction and attenuation factors over the population values would tend to introduce a negative nonconservative bias when used in the Multiplicative Model equation (13) to estimate the variance of true validities. Since equation (13) is complex, it remains to be seen whether the increase in variance of sample attenuation and restriction factors is of sufficient magnitude to produce significantly nonconservative estimates of the variance of the true validities.

One result which was common to all six conditions was that the mean estimate of the unrestricted criterion reliability was larger than the mean population value. Further, the degree of discrepancy decreases with larger sample size. This appears to indicate a slight positive bias in sample restriction-adjusted reliability estimates. The degree of discrepancy, even for the very small samples of 30, is too minor to have any significant effect in actual adjustments of correlations for attenuation. Further it should be pointed out that any error introduced by this is conservative because an overestimate of reliability produces an under-adjustment for attenuation.

In sum, the results in Tables 10, 11, and 12 further support the accuracy of the simulation program under the same conditions in which the tests of model accuracy were conducted.

Accuracy of Model Estimates of Variance of True Validities

These results are summarized for all six conditions in Table 13. Looking first at the discrepancies between the actual variance and mean Multiplicative Model estimate, it can be seen that the model was very accurate for the narrow distribution, but somewhat overestimated the variance of the wide distribution for each sample size. The amount of overestimation is very similar to that previously reported for the wide distribution and infinite sample size in Table 6. This suggests that the bias introduced by small sample estimation of restriction and attenuation factors is negligible.

Table 13

Mean and Standard Deviation Over 25 Replications of Multiplicative and Schmidt-Hunter Model Estimates of Variance of True Validities

		Multiplicative model			Schmidt-Hunter model		
Distribution of true validities	Sample size	Actual variance	Mean of variance estimates	S.D. of variance estimates	Actual variance	Mean of variance estimates	S.D. of variance estimates
Narrow	30	.002	.001	.016	.003	-.007	.005
	68	.002	.002	.006	.003	-.003	.002
	200	.002	.002	.004	.003	-.002	.001
Wide	30	.031	.041	.021	.077	.005	.007
	68	.031	.036	.013	.077	.009	.005
	200	.031	.040	.009	.077	.010	.003

Note. The Multiplicative model mean estimated variance should be compared with the variance of the distribution of true validities. The Schmidt-Hunter model mean estimated variance must be compared with the variance of the Fisher's Z conversions of the true validities.

Comparison of the actual variance and mean variance estimate by the Schmidt-Hunter model shows that the actual variance was underestimated in every condition. There was a large difference between the actual variances of the wide and narrow distributions, yet the Schmidt-Hunter model gave very low to negative estimates for both. This indicates that the Schmidt-Hunter model is rather insensitive to the amount of variation in the true validities. No matter what that variation happens to be, the model estimates a near-zero variance. In every condition, the Multiplicative Model was more accurate (less biased) than the Schmidt-Hunter model.

The standard deviations in Table 13 provide a summary description of the consistency of the model estimates from replication to replication. It can be seen that as the sample size increases, the estimates made by both models become more consistent across the replications. Also, the variance estimates of both models were less variable on the narrow than the wide distribution. The Multiplicative Model estimates varied more than the Schmidt-Hunter estimates in each condition. The magnitude of the standard deviations of the Multiplicative Model variance estimate are disturbing particularly for sample size 30, where large over- or under-estimates were found on several replications.

In the introductory chapter, it was pointed out that the effect of positive correlation between the population c's, a's, and ρ's was to introduce a bias of overestimation by the Multiplicative Model of the variance of true validities. Overestimation bias was noted in Table 13 for the wide, but not for the narrow distribution. Table 14 shows the average correlations of ρ, a, and c for each condition. As should be expected, ρ and a were generally uncorrelated. However, c tended to be correlated with both ρ and a. This is to be expected because the value of c depends on both the value of ρ and the value of a. In the narrow distribution, where c has low correlations, overestimation bias was not found in the Multiplicative Model. With the wide distribution, where c was correlated about .21 with ρ and .08 with a, the overestimation bias was found.

Accuracy of Model Estimates of Credibility Interval Limits

The accuracy of model estimates of credibility interval limits is obviously tied to the results for the accuracy of the model estimates of variance. Nevertheless, it is interesting to note just how much error in credibility intervals is produced by errors in estimating variance. These results are shown in Table 15 for the one-tailed 90% credibility limit.

The Multiplicative Model credibility values were generally slightly too small and therefore conservative. The Schmidt-Hunter model credibility values were too high

Table 14

Mean Over 25 Replications of Correlations (Over 100 Distributions) Between Population True Validities (ρ) Attenuation Factors (a), and Restriction Factors (c)

Distribution of true validities:	Narrow			Wide		
Sample size:	30	68	200	30	68	200
Mean correlation of ρ with a	.03	.00	-.01	.00	.01	.02
Mean correlation of ρ with c	.08	.01	.04	.22	.21	.21
Mean correlation of a with c	.03	.05	.09	.07	.08	.09

Table 15

Mean and Standard Deviation Over 25 Replications of Multiplicative and Schmidt-Hunter Model Estimates of the One-Tailed 90% Credibility Value

Distribution of true validities	Sample size	Multiplicative model: Actual credibility value	Multiplicative model: Mean of credibility value estimates	Multiplicative model: S.D. of credibility value estimates	Schmidt-Hunter model: Actual credibility value	Schmidt-Hunter model: Mean of credibility value estimates	Schmidt-Hunter model: S.D. of credibility value estimates
Narrow	30	.445	.427	.097	.478	.508	.037
	68	.445	.444	.061	.478	.503	.023
	200	.445	.451	.047	.478	.498	.012
Wide	30	.280	.246	.095	.288	.436	.065
	68	.280	.264	.054	.288	.407	.044
	200	.280	.253	.032	.288	.388	.022

Note. Credibility estimates are computed from the estimate of variance of true validities by the respective model. Whenever the variance estimate was negative, it was taken to be zero for purposes of computing the credibility value. This frequently occurred with the Schmidt-Hunter model estimates for the narrow distribution.

(nonconservative) in each case and were grossly in error for the wide distribution.

The standard deviations of the credibility values are substantial, but are reduced with increasing sample size. The results for the Multiplicative Model show that credibility limit estimates varied about the same amount for the wide and narrow distributions.

Results of Application of Multiplicative Model to Ghiselli (1966) Distributions

The Multiplicative Model prior distribution and credibility limits are shown in Table 16. The mean unrestricted disattenuated validity from the Multiplicative Model is typically very close to that given by Schmidt and Hunter (1977). The mean is somewhat higher for machine tenders than that computed by Schmidt and Hunter, however.

The prior distribution standard deviations are all larger and the 97.5% credibility limits are all lower than those reported by Schmidt and Hunter (1977). Nevertheless, validity generalization is supported by the 90% credibility value in the case of mechanical principles tests for predicting training criteria for mechanical repairmen and for predicting the job proficiency of clerks with intelligence tests. It is not supported for predicting job proficiency of bench workers with finger dexterity tests nor for predicting job proficiency of machine tenders with spatial relations tests. These results are in general agreement with the previous conclusions of Schmidt and Hunter (1977).

Table 16

Results of Applying Multiplicative Model to Ghiselli (1966) Validity Distributions

Job/Test Type	Estimated prior distribution: Mean unrestricted disattenuated validity	S.D. of unrestricted disattenuated validities	One-tailed credibility value: 97.5%	90%
Mechanical repairmen/ Mechanical principles	.71 (1.03)	.27 (.08)	.19 (.70)	.36
General Clerks/ Intelligence	.69 (.81)	.37 (.20)	-.04 (.40)	.21
Bench workers/ Finger dexterity	.43 (.41)	.43 (.23)	-.41 (-.04)	-.12
Machine tenders/ Spatial relations	.10 (.05)	.36 (.18)	-.61 (-.30)	-.36

Note. The assumed sample size is 68. The numbers in parenthesis are the prior distribution values in Fisher's z form reported by Schmidt and Hunter (1977).

CHAPTER IV

DISCUSSION

Model Accuracy

The Multiplicative Model and its attendant estimation procedures was more accurate than the Schmidt-Hunter Model in every situation studied. It was found to be capable of distinguishing between distributions of zero, moderate, and large amounts of variation in the true validities. It is recommended for use in future validity generalization studies. One caution indicated by this study is that with small sample sizes of around 30, estimates of the variance of true validities and lower credibility values vary widely.

Need for Future Research

One parameter of the model which was never varied in the present research on accuracy was the number of validity studies. In each case this was fixed at 100. There is no reason to expect that there would be any noteworthy change in the bias of Multiplicative estimate of true validity variance with the introduction of more studies, but the stability of the estimate would be expected to increase. It would be interesting to know, for example, whether the same bias and stability would result from 100 studies each with sample size 200 as from 200 studies each with sample size 100. This could have implications for future preferences for small vs. large sample research and also provide some prior idea of how much stability of the estimate could be expected from the total N of all the available studies.

Possible Improvements

The degree of conservative bias in the Multiplicative Model is not so great as to be debilitating. However, it would be well if it could be removed somehow. It appears that taking into account the correlation of c with ρ and a would be the key to removing this bias. The development of a mathematical solution to the problem of finding the variance of the product of three variables which are not assumed to be uncorrelated may be a key step in solving this problem.

The present model could also be elaborated to allow for differences in the reliabilities of the predictors. Schmidt and Hunter suggest that this would be useful to further isolate the effects of job and situational factors on validity.

Situational Specificity Hypothesis

Schmidt and Hunter (1977) have referred to the idea that job and situational factors have an important effect on validity as the "situational specificity hypothesis" and speak in terms of accepting or rejecting it by means of validity generalization analysis. Of course, it could be "rejected" if the analysis indicates zero variance of the true validities. This may not be a common finding, however. The more likely result is that some variance will be left in the true validity prior. The result of the model makes it possible to make comparisons between different priors in terms of variance, and this may be beneficial in confirming or denying the importance of a job content or situational factor in the future. In the absence of a well accepted criterion below which the "situational specificity" hypothesis can be rejected, it may be more productive to think in terms of "how much" variance rather than "zero or not zero".

Use of Assumed Criterion Reliability and Range Restriction Distributions

Primarily because of lack of directly relevant data, assumed distributions of criterion reliability and range restriction values were used in both the accuracy research and the analysis of the Ghiselli distributions. Schmidt and Hunter (1977), have argued that these distributions are representative in the sense that their average values are likely to be encountered. In both validity generalization models, the dispersion of these values is also a factor in ascertaining the true validity variance. The same distributions were used for mechanical repairmen (training criteria) and for clerks (proficiency criteria), yet it is very plausible that there would be some differences in criterion reliability in these cases. Reasonable assumed distributions are better than none. However, more attention should be given to documenting and accumulating actual criterion reliability and range restriction distributions for future generalization studies. Criterion reliability and range restriction information from original reports should be preferred over the assumed distributions whenever available.

REFERENCES

Brunk, H. D. Mathematical Statistics, Waltham, Mass.: Blaisdell, 1965.

Callender, J. C. and Osburn, H. G. A model for the generalization of validity. Unpublished manuscript, University of Houston, 1978.

Ghiselli, E. E. The Validity of Occupational Aptitude Tests. New York: Wiley, 1966.

Johnson, N. L. and Kotz, S. Continuous Univariate Distributions Vol. II. Boston: Houghton-Mifflin, 1970.

Lent, R. H., Aurbach, H. A., and Levin, L. S. Predictors criteria, and significant results. Personnel Psychology, 1971, 24, 519-533.

Lord, F. M. and Novick, M. R. Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley, 1968.

Schmidt, F. L., Hunter, J. E., and Urry, V. W. Statistical power in criterion-related validity studies. Journal of Applied Psychology, 1976, 61, 473-485.

Schmidt, F. L. and Hunter, J. E. Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 1977, 62, 529-540.

APPENDIX A

The following derivation gives a general formula for the variance of the product of three uncorrelated variables. Brunk (1965) has shown that for independent random variables x and y:

$$E(xy) = E(x)\,E(y)\ . \tag{1}$$

Equation (2) is the usual formula for variance in terms of expected values.

$$V(x) = E(x^2) - \left[E(x)\right]^2 \tag{2}$$

Substituting xy for x in (2), we have:

$$V(xy) = E\left[(xy)^2\right] - \left[E(xy)\right]^2 \tag{3}$$

which can be written as:

$$V(xy) = E(xy\ xy) - \left[E(x)\ E(y)\right]^2 \tag{4}$$

which, in turn, can be written as:

$$V(xy) = E(x^2\ y^2) - E(x)\ E(y)^2\ . \tag{5}$$

If x and y are independent, then x^2 and y^2 are also independent and we have:

$$E(x^2y^2) = E(x^2)\ E(y^2)\ . \tag{6}$$

Substituting (6) into (5) gives:

$$V(xy) = E(x^2)\ E(y^2) - \left[E(x)\ E(y)\right]^2 \quad . \tag{7}$$

Equation (2) can be rearranged as:

$$E(x^2) = V(x) + \left[E(x)\right]^2 \quad . \tag{8}$$

Substituting (8) into (7) for both x and y gives:

$$V(xy) = \left[V(x) + E(x)^2\right] \left[V(y) + E(y)^2\right] - \left[E(x)\ E(y)\right]^2 \quad . \tag{9}$$

Expanding terms gives:

$$V(xy) = V(x)\ V(y) + V(x) \left[E(y)\right]^2 + V(y) \left[E(x)\right]^2 + \left[E(x)\right]^2 \left[E(y)\right]^2 - \left[E(x)\ E(y)\right]^2 \tag{10}$$

which simplifies to:

$$V(xy) = V(x)\ V(y) + V(x) \left[E(y)\right]^2 + V(y) \left[E(x)\right]^2 \quad . \tag{11}$$

By rewriting (11), using M for mean and S^2 for variance to be consistent with the notation of the main text, we have that:

$$S^2_{xy} = S^2_x\ S^2_y + S^2_x\ M^2_y + S^2_y\ M^2_x \quad . \tag{12}$$

The variance of the product of the variables is a function of the squares of the means and the variances of each of the variables. It can be shown that the extension of equation (12) to the product of three variables is:

$$S^2_{xyz} = S^2_x(S^2_y + M^2_y)\ (S^2_z + M^2_z) + M^2_x\ S^2_y(S^2_Z + M^2_Z) + M^2_x\ M^2_y\ S^2_z \quad . \tag{13}$$