PERSONNEL PSYCHOLOGY
2001, 54

ESTIMATION OF SAMPLING VARIANCE OF CORRELATIONS IN META-ANALYSIS

HERMAN AGUINIS
Graduate School of Business Administration
University of Colorado at Denver

Monte Carlo simulations were conducted to compare the performance of the traditional (Fisher, 1954) and mean (Hunter & Schmidt, 1990) estimators of the sampling variance of correlations in meta-analysis. The mean estimator differs from the traditional estimator in that it uses the mean observed correlation, averaged across studies, in the sampling variance formula. The simulations investigated the homogeneous (i.e., no true correlation variance across studies) and heterogeneous case (i.e., true correlation variance across studies). Results reveal that, compared to the traditional estimator, the mean estimator provides less negatively biased estimates of sampling variance in the homogeneous and heterogeneous cases and more positively biased estimates in the heterogenous case. Thus, results support the use of the mean estimator unless strong, theory-based hypotheses regarding moderating effects exist.

Meta-analysis constitutes a set of procedures used to quantitatively integrate a body of literature. Although several meta-analytic techniques are available (e.g., Hedges & Olkin, 1985; Rosenthal, 1991), validity generalization (VG) is one of the most commonly used approaches in applied psychology, management, and numerous other disciplines (Hunter & Schmidt, 1990; Schmidt, 1992). VG extends arguments from

I am grateful to Charles A. Pierce (Montana State University) and three *Personnel Psychology* anonymous reviewers for thoughtful and constructive comments regarding previous drafts. I also thank Roger E. Millsap (Arizona State University) for advice throughout the duration of this project and the derivation shown in the Appendix. The research reported herein was supported by grants from the Office of Research Administration and the Graduate School of Business Administration (University of Colorado at Denver).

A previous version of this article was presented at the Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA, April 2000. Portions of the research reported herein were conducted while the author was on sabbatical leave from the University of Colorado at Denver and holding visiting appointments at China Agricultural University–International College of Beijing (People's Republic of China), City University of Hong Kong (People's Republic of China), Nanyang Technological University (Singapore), University of Science Malaysia (Penang, Malaysia), and University of Santiago de Compostela (Spain).

Correspondence and requests for reprints should be addressed to Herman Aguinis, Graduate School of Business Administration, University of Colorado at Denver, Campus Box 165, P.O. Box 173364, Denver, CO 80217-3364; Herman.Aguinis@cudenver.edu; http://www.cudenver.edu/~haguinis.

psychometric theory to assert that a substantial portion of the variability observed in a predictor–criterion relationship across individual studies is the result of sources of variance not explicitly considered in a study design. Consequently, to estimate better a predictor–criterion relationship in the population, researchers should (a) attempt to control the impact of these extraneous sources of variance by implementing sound research designs, and/or (b) correct for the extraneous across-study variability by subtracting it from the total variance in study-level effect size estimates (Aguinis & Pierce, 1998b; Hunter & Schmidt, 1990).

Sampling error is a major source of extraneous across-study variability (Koslowsky & Sagie, 1994). Consequently, researchers are advised that it be corrected in conducting meta-analyses (Hunter & Schmidt, 1990). VG procedures include explicit steps to correct for sampling error, whereas other meta-analytic approaches include the estimation of sampling error implicitly (e.g., in chi-square tests of homogeneity; Hedges & Olkin, 1985). The traditional sampling variance estimator is:

$$S^2_{e_{rk}} = \frac{(1 - r_k^2)^2}{N_k - 1} \qquad [1]$$

where r is a study-level correlation coefficient (i.e., effect size estimate) based on a sample of size N for the kth study (Fisher, 1954). Estimates are calculated separately for each study included in a meta-analytic review and then are combined to yield an estimate of sampling variance across studies.

The estimator in Equation 1 has been studied analytically or through simulation by Callender and Osburn (1988), Millsap (1988, 1989), and Aguinis and Whitehead (1997). In general, the estimator has a negative bias, and this bias is larger in smaller samples. Stated differently, meta-analysts typically underestimate the variability of correlations across studies due to sampling error. Thus, researchers may overestimate the variability due to substantive moderating effects and, thus, incorrectly discover "false moderators."

In addition to sample size, other variables have been found to affect the accuracy of the traditional estimator. Millsap (1989) found that negative bias increases in the presence of measurement error and direct range restriction, and Aguinis and Whitehead (1997) showed that the bias is increased by as much as 8.5% in the presence of indirect range restriction. The measurement error findings are explained by the fact that introducing measurement error causes attenuation of the population uncorrected correlation coefficient, which causes the sampling error estimate to have more negative bias (Hunter & Schmidt, 1990, pp. 208–209). The range restriction effects are not fully explained by attenuation alone; they are also a function of the departure from normality in the restricted population.

A general conclusion to be drawn from the available studies of the sampling error estimator shown in Equation 1 is that the estimator is fairly accurate in large samples, but that some negative bias is expected in small samples ($N \leq 100$). In small samples, the negative bias is more substantial in the presence of measurement error and direct and indirect range restriction.

Given the bias in the estimator shown in Equation 1, Hunter and Schmidt (1990, pp. 208–209) proposed an alternative estimator that uses the average sample correlation in place of r in Equation 1. Let $\bar{r}$ be the average (weighted or unweighted) sample correlation across k independent studies. The new estimator is:

$$S^2_{e_{mk}} = \frac{(1 - \bar{r}^2)^2}{N_k - 1}, \qquad [2]$$

with N defined as in Equation 1. This estimator is calculated separately for each kth study using the appropriate N, and is then combined across studies. The estimator in Equation 2 is denoted the "mean estimator" in what follows, to distinguish it from the "traditional estimator" in Equation 1.

Hunter and Schmidt (1990, pp. 208–209) maintained that the mean estimator has less negative bias than the traditional estimator, particularly in small-to-moderate samples. More recently, Hunter and Schmidt (1994) showed analytically that the mean estimator would outperform the traditional estimator in the absence of true validity variance (i.e., no moderating effect) when the true validity does not exceed an upper limit related to the sample size of the validity studies. No such analytic demonstrations have been published for the case in which the true validity variance is nonzero and moderating effects exist.

The only published study that used simulations to examine the performance of the traditional and mean estimators in the heterogeneous case (i.e., presence of true validity variance) is Law, Schmidt, and Hunter (1994). In that study, data were simulated under conditions of heterogeneity in the true population correlations while also incorporating range restriction and attenuation artifacts. A limited range of heterogeneity conditions was examined, including six different combinations of mean true validity and true validity variance. The true validity variance ranged from .003 to .034, and these values were unevenly paired with mean true validity conditions.

Although Law et al.'s (1994) study provided some preliminary evidence illustrating the performance of the traditional and mean sampling variance estimators, the study has the following three limitations. First, no actual values for the two estimators in the simulations were reported. The study instead focused on estimates of the mean and variance of the true validity distributions and did not report results of a direct comparison of the performance of the two estimators. Second, the inclusion of range restriction and attenuation in the generated data makes it difficult to examine the performance of the estimators apart from these artifacts. Specifically, Aguinis and Whitehead (1997) and Millsap (1988, 1989) showed that range restriction and measurement error affect the performance of the traditional sampling error variance estimator. Similarly, it is likely that the mean estimator is also affected by these artifacts. Third, only a limited number of values for true mean validity (i.e., .3, .5, and .7) and true validity variance (i.e., .0030, .0122, .0218, and .0340) was examined. Thus, it is not clear that findings can be generalized to a broad set of situations encountered by researchers implementing meta-analytic methods in applied psychology, management, and other social sciences.

Given the claims and the possibility that the mean estimator provides an improvement over the traditional estimator, there is a need to conduct a systematic empirical comparison of the two estimators. Thus, the overall purpose of the present study is to investigate whether the mean estimator is less negatively biased than the traditional estimator.

Comparison of Traditional and Mean Estimators: Homogeneous and Heterogenous Cases

The performance of the traditional and mean estimators can be evaluated under either of two conditions. In the homogeneous case, a hypothesis that a moderating effect exists is false and the true validity ρ is constant across studies: There is no true validity variance. If sampling error is the only operative artifact, each sample correlation r is an estimate of the common true validity ρ. Simulation work has shown that the traditional estimator has a negative bias (Aguinis & Whitehead, 1997; Millsap, 1988, 1989). In addition, as noted above, Hunter and Schmidt (1994) showed analytically that the mean estimator would outperform the traditional estimator in the absence of true validity variance. Thus, the following research question is offered:

> *Research Question 1*: Is the mean estimator less negatively biased than the traditional estimator in the homogenous case?

In the heterogeneous case, a hypothesis that a moderating effect exists is true, and the true validity ρ does vary across studies, resulting in positive true validity variance. The sample study correlations rs are not estimates of a common true validity ρ, and the performance of both the traditional and the mean estimator is difficult to predict in general. Thus, the following research questions are offered:

> *Research Question 2*: Is the negative bias in the traditional estimator greater in the heterogenous as compared to the homogeneous case?
>
> *Research Question 3*: Is the mean estimator less negatively biased than the traditional estimator in the heterogenous case?

Method

Manipulated Parameters

The following parameters were manipulated in the simulation:

Sample size. Sample size N was set at values of 60, 100, and 140. These values cover a fairly typical range in several applied psychology specialties, especially in personnel selection research. For example, Lent, Aurbach, and Levin (1971) found that the median sample size in 1,500 validation studies was 68. More recently, Salgado (1998) examined all criterion-related validity studies published in *Journal of Applied Psychology, Journal of Occupational and Organizational Psychology*, and *Personnel Psychology* between 1983 and 1994, and reported that the median sample size across 96 studies was 113.

Population correlation. The correlation was set at values between .10 and .90 in increments of .10 so as to represent varying degrees of effect size. In the homogeneous case, rs were drawn from the same value of ρ. In the heterogenous case, 50% of the rs were drawn from one value of ρ and 50% of the rs were drawn from another value of ρ, including every possible pairing of ρ values from .10 to .90 in increments of .10. Stated differently, for the heterogeneous case, 50% of rs were sampled from $\rho_1 = .10$ and 50% of rs from $\rho_2 = .20$, 50% of rs from $\rho_1 = .10$ and 50% of rs from $\rho_2 = .30$, and so forth. This situation simulates a simple case of a moderator variable with two values (e.g., male–female, majority group–minority group).

The manipulation of the independent variables led to a design having a total of 135 cells, 27 of which simulated homogeneous cases and 108 simulated heterogenous cases.

Procedure and Dependent Variable

Computer programs. The simulation was performed using C++ programs based on algorithms described by Aguinis (1994). The programs used are available upon request.

Simulation procedure. The programs generated 5,000 samples for each of the cells (i.e., combination of sample size and population correlation values) of the design. The simulation involved the following two steps:

1. Bivariate (X, Y) arrays of size N were generated from multivariate normal populations with a mean of zero (i.e., $\mu_X = \mu_Y = 0$), unit variance (i.e., $\sigma_X = \sigma_Y = 1.0$) and correlation ρ.

2. Correlations (rs) were calculated from each of the 5,000 samples generated for each cell in the design. As noted above, rs were calculated from one value of ρ for the homogeneous case and from two different values of ρ for the heterogeneous case.

Dependent variables. To compare the traditional and the mean estimator in the homogeneous and heterogenous cases, the following quantities were computed: (a) the total variance S_r^2 from empirically generated sampling distributions of rs for each cell in the design, (b) the traditional estimator $S_{e_r}^2$ (shown in Equation 1 and omitting subscript k for simplicity), and (c) the mean estimator $S_{e_m}^2$ (shown in Equation 2 and omitting subscript k for simplicity). Note that in the homogeneous case, rs are generated from one population value (i.e., $\rho_1 = \rho_2$). Therefore, the true validity variance $S_t^2 = 0$, and the total variance $\underline{S}_r^2$ represents sampling error variance only (i.e., $S_r^2 = S_e^2$). Alternatively, in the heterogeneous case rs are generated from two population values (i.e., $\rho_1 \neq \rho_2$). Therefore, $S_t^2 > 0$ and includes both true variance and error variance. Given that (see the Appendix for the derivation of Equation 3)

$$S_t^2 = \frac{\rho_1^2 + \rho_2^2 - 2\rho_1\rho_2}{4}, \qquad [3]$$

in the heterogenous case sampling error variance is

$$S_e^2 = S_r^2 - S_t^2. \qquad [4]$$

In addition, the following differences were computed for each of the 135 cells in the design:

$$d_{e_r} = S_e^2 - S_{e_r}^2 \qquad [5]$$

$$d_{e_m} = S_e^2 - S_{e_m}^2. \qquad [6]$$

Finally, the following quantities were also computed: (a) the percentage by which the traditional estimator underestimates error variance (i.e., $\%\ S_{e_r}^2 = 100 - [(S_{e_r}^2\ /\ S_e^2) \times 100]$), (b) the percentage by which the mean estimator underestimates error variance (i.e., $\%\ S_{e_m}^2 = 100 -$

$[(S^2_{e_m} / S^2_e) \times 100]$), (c) and the difference between these two percentages (i.e., $\%S^2_{e_r} - \%S^2_{e_m}$).

Key Accuracy Checks

To assess the key accuracy of the computer programs, the results were compared to those reported by Millsap (1989) and Aguinis and Whitehead (1997) regarding S^2_r and $S^2_{e_r}$ in the homogeneous case and those computed using the present programs. (These simulations did not include the heterogeneous case.) All values were similar to the values reported in these articles.

Results and Discussion

Homogeneous Case

Table 1 shows that the traditional estimator has a negative bias. The traditional estimator underestimated the total variance in 100% of cases for $N = 60$, 100% of cases for $N = 100$, and 55.56% of cases for $N = 140$. In addition, Table 1 shows that the magnitude of this negative bias decreases as sample size increases. More precisely, the traditional estimator underestimated sampling variance by 3.93% for $N = 60$, 1.89% for $N = 100$, and .54% for $N = 140$.

Substantiating previous analytical work by Hunter and Schmidt (1994), results shown in Table 1 indicate that the mean estimator provides an improvement over the traditional estimator. More precisely, the mean estimator underestimated sampling variance by 3.68% for $N = 60$, 1.74% for $N = 100$, and .43% for $N = 140$. Collapsing across correlation values, these percentages represent improvements, albeit small, over the traditional estimator of .25%, .15%, and .11% for $N = 60$, $N = 100$, and $N = 140$, respectively.

The aforementioned improvement in the estimation of sampling variance by using the mean estimator as compared to the traditional estimator takes place in the lower effect size range (i.e., $\rho_1 = \rho_2 = .1$ to .5), which is the effect size range most typically observed in applied psychology and management research. For instance, for $N = 60$ the mean estimator outperformed the traditional estimator (i.e., $\%S^2_{e_r} - \%S^2_{e_m}$) by 3.21% for $\rho_1 = \rho_2 = .1$, 2.94% for $\rho_1 = \rho_2 = .2$, and 2.48% for $\rho_1 = \rho_2 = .3$. Collapsing across sample size values, when $\rho_1 = \rho_2 = .1$ the mean estimator outperformed the traditional estimator by 2.18%, whereas this percentage decreased to 1.99% for $\rho_1 = \rho_2 = .2$, 1.66% for $\rho_1 = \rho_2 = .3$, and 1.21% for $\rho_1 = \rho_2 = .4$. The percentage of improvement continues to decrease as effect size increases. (The appendix

TABLE 1

Underestimation of Sampling Error Variance Using the Traditional and Mean Estimators for the Homogenous Case

N	ρ_1	ρ_2	$\bar{r}$	$\%S^2_{e_r}$	$\%S^2_{e_m}$	$\%S^2_{e_r} - \%S^2_{e_m}$
60	0.1	0.1	0.10	5.24	2.03	3.21
60	0.2	0.2	0.20	4.70	1.76	2.94
60	0.3	0.3	0.30	4.87	2.39	2.48
60	0.4	0.4	0.40	5.94	4.13	1.81
60	0.5	0.5	0.50	3.66	2.71	0.95
60	0.6	0.6	0.60	2.82	2.91	−0.08
60	0.7	0.7	0.70	4.51	5.88	−1.36
60	0.8	0.8	0.80	2.41	5.33	−2.92
60	0.9	0.9	0.90	1.19	5.97	−4.78
100	0.1	0.1	0.10	2.85	0.92	1.93
100	0.2	0.2	0.20	0.07	−1.70	1.77
100	0.3	0.3	0.30	0.34	−1.11	1.45
100	0.4	0.4	0.40	3.32	2.26	1.07
100	0.5	0.5	0.50	0.75	0.23	0.53
100	0.6	0.6	0.60	4.90	4.99	−0.09
100	0.7	0.7	0.70	3.58	4.43	−0.86
100	0.8	0.8	0.80	0.95	2.79	−1.83
100	0.9	0.9	0.90	0.26	2.88	−2.62
140	0.1	0.1	0.10	3.44	2.06	1.39
140	0.2	0.2	0.20	−2.73	−3.98	1.25
140	0.3	0.3	0.30	2.84	1.78	1.05
140	0.4	0.4	0.40	0.16	−0.59	0.75
140	0.5	0.5	0.50	2.95	2.57	0.38
140	0.6	0.6	0.60	1.75	1.82	−0.07
140	0.7	0.7	0.70	−0.75	−0.11	−0.64
140	0.8	0.8	0.80	−2.47	−1.18	−1.29
140	0.9	0.9	0.90	−0.37	1.50	−1.87

Notes: $\bar{r}$ = mean observed correlation, $\%S^2_{e_r}$ = percentage by which the traditional estimator underestimates sampling variance (i.e., $100 - [(S^2_{e_r}/S^2_r) \times 100]$), $\%S^2_{e_m}$ = percentage by which the mean estimator underestimates sampling variance (i.e., $100 - [(S^2_{e_m}/S^2_r) \times 100]$). Tables including values for S^2_r (i.e., total sampling variance), $S^2_{e_r}$ (i.e., sampling variance computed using the traditional estimator shown in Equation 1), and $S^2_{e_m}$ (i.e., sampling variance computed using the mean estimator shown in Equation 2) are available from the author.

of Hunter & Schmidt, 1994, provides a possible explanation for this phenomenon.)

Heterogeneous Case

In the heterogeneous case, effect size estimates do originate from different populations (i.e., $\rho_1 \neq \rho_2$). Thus, this is a situation in which there actually exists a moderating effect in the population.

Traditional estimator. Tables 2–4 show that, overall, the traditional estimator suffers from a negative bias. However, as sample size increases,

TABLE 2

Underestimation of Sampling Error Variance Using the Traditional and Mean Estimators for the Heterogeneous Case (N = 60)

ρ_1	ρ_2	$\bar{r}$	$\%S^2_{e_r}$	$\%S^2_{e_m}$	$\%S^2_{e_r} - \%S^2_{e_m}$
.10	.20	.15	3.17	−0.41	3.58
.10	.30	.20	0.06	−4.82	4.88
.10	.40	.25	2.04	−4.90	6.94
.10	.50	.30	1.59	−8.19	9.78
.10	.60	.35	−5.42	−19.40	13.98
.10	.70	.40	5.23	−10.52	15.75
.10	.80	.45	−17.67	−37.99	20.33
.10	.90	.50	1.80	−12.23	14.04
.20	.30	.25	3.53	0.39	3.14
.20	.40	.30	4.83	0.63	4.20
.20	.50	.35	−7.88	−14.22	6.34
.20	.60	.40	−1.28	−9.56	8.28
.20	.70	.45	−1.72	−11.77	10.05
.20	.80	.50	1.36	−8.27	9.63
.20	.90	.55	−13.93	−18.88	4.95
.30	.40	.35	6.17	3.62	2.56
.30	.50	.40	−3.06	−6.49	3.43
.30	.60	.45	4.18	−0.13	4.31
.30	.70	.50	5.70	0.83	4.87
.30	.80	.55	10.49	7.24	3.25
.30	.90	.60	8.40	10.95	−2.55
.40	.50	.45	2.13	0.41	1.72
.40	.60	.50	−2.89	−4.87	1.98
.40	.70	.55	−1.30	−3.04	1.74
.40	.80	.60	3.22	3.93	−0.70
.40	.90	.65	5.83	13.75	−7.92
.50	.60	.55	5.33	4.69	0.64
.50	.70	.60	2.41	2.54	−0.14
.50	.80	.64	14.08	16.74	−2.66
.50	.90	.70	15.51	25.82	−10.31
.60	.70	.64	1.75	2.71	−0.96
.60	.80	.70	1.67	5.75	−4.09
.60	.90	.75	8.54	21.59	−13.05
.70	.80	.75	1.35	4.95	−3.60
.70	.90	.80	16.02	27.88	−11.86
.80	.90	.85	6.12	15.75	−9.63

Notes: $\bar{r}$ = mean observed correlation, $\%\underline{S}^2_{e_r}$ = percentage by which the traditional estimator underestimates sampling variance (i.e., $100 - [(\underline{S}^2_{e_r}/\underline{S}^2_e) \times 100]$), $\%\underline{S}^2_{e_m}$ = percentage by which the mean estimator underestimates sampling variance (i.e., $100 - [(\underline{S}^2_{e_m}/\underline{S}^2_e) \times 100]$). Tables including values for $\underline{S}^2_r$ (i.e., total sampling variance), $\underline{S}^2_{e_r}$ (i.e., sampling variance computed using the traditional estimator shown in Equation 1), and $\underline{S}^2_{e_m}$ (i.e., sampling variance computed using the mean estimator shown in Equation 2), $\underline{S}^2_t$ (i.e., true variance computed using Equation 3), and $\underline{S}^2_e$ (i.e., sampling error variance, $\underline{S}^2_r - \underline{S}^2_t$) are available from the author.

the number of conditions for which underestimation takes place decreases. More precisely, Table 2 shows that when $N = 60$, the underestimation occurred in 75% of cases, Table 3 shows that when $N = 100$, the underestimation took place in 61.11% of cases, whereas Table 4 shows that when $N = 140$, the underestimation occurred in only 44.44% of cases.

Table 2 shows that when $N = 60$, the traditional estimator underestimated the total sampling error variance by as much as 16.02% for condition $\rho_1 = .7, \rho_2 = .9$. Table 3 shows that when $N = 100$, this underestimation was as high as 25.01% for condition $\rho_1 = .1, \rho_2 = .9$. Table 4 shows that when $N = 140$, the underestimation was a high as 15.58% for condition $\rho_1 = .2, \rho_2 = .7$.

The conditions of most interest to applied psychology and management researchers are those including correlations in the .1 to .5 range. For these conditions, the mean magnitude of sampling error variance underestimation was 2.94% for $N = 60$, 3.67% for $N = 100$, and 3.70% for $N = 140$.

Tables 2–4 also show that the extent of underestimation is affected by the relative location of ρ_1 and ρ_2 (i.e., $\bar{r}$) and the absolute difference between ρ_1 and ρ_2 (i.e., $|\rho_1 - \rho_2|$). A regression analysis was conducted using data from Tables 2–4 to understand better these relationships. The criterion was the percent by which the traditional estimator underestimated sampling error variance, and the predictors were $\bar{r}$ and $|\rho_1 - \rho_2|$. Results summarized in Table 5 show that the underestimation increased as (a) $\bar{r}$ increased, and (b) $|\rho_1 - \rho_2|$ decreased. Note, however, that these results need to be qualified by the fact that, as can be seen by perusing Tables 2–4, these relationships are not completely linear.

Mean estimator. The mean estimator underestimated total sampling variance in 52 of the 108 heterogeneous conditions (i.e., 48.15%). As in the case with the traditional estimator, sample size affected the number of conditions for which underestimation took place such that sampling variance was underestimated in 19 of the 36 cases for $N = 60$ (i.e., 52.78%), 19 of the 36 cases for $N = 100$ (i.e., 52.78%), and 14 of the 36 cases for $N = 140$ (i.e., 38.89%).

Table 2 shows that when $N = 60$ the mean estimator underestimated the empirically derived sampling error variance by as much as 27.88% for $\rho_1 = .7$ and $\rho_2 = .9$. Table 3 shows that when $N = 100$ this underestimation was as high as 25.98% for $\rho_1 = .6$ and $\rho_2 = .9$. Table 4 shows that when $N = 140$ the underestimation was as high as 24.26% for $\rho_1 = .6$ and $\rho_2 = .9$.

The conditions of most interest to applied psychology and management researchers are those for which effect sizes range from .1 to .5.

TABLE 3

Underestimation of Sampling Error Variance Using the Traditional and Mean Estimators for the Heterogeneous Case (N = 100)

ρ_1	ρ_2	$\bar{r}$	$\%S^2_{e_r}$	$\%S^2_{e_m}$	$\%S^2_{e_r} - \%S^2_{e_m}$
.10	.20	.15	2.34	0.00	2.34
.10	.30	.20	3.92	0.30	3.62
.10	.40	.25	−2.40	−8.43	6.03
.10	.50	.30	−4.07	−13.29	9.21
.10	.60	.35	0.55	−11.68	12.23
.10	.70	.40	−22.09	−40.45	18.36
.10	.80	.45	−10.47	−28.58	18.11
.10	.90	.50	25.01	15.01	10.00
.20	.30	.25	4.18	2.09	2.09
.20	.40	.30	1.30	−1.94	3.24
.20	.50	.35	−1.20	−6.37	5.17
.20	.60	.40	5.25	−1.71	6.96
.20	.70	.45	−8.66	−18.31	9.65
.20	.80	.50	−1.39	−9.90	8.50
.20	.90	.55	−11.76	−15.37	3.61
.30	.40	.35	4.54	2.87	1.67
.30	.50	.40	−2.56	−5.23	2.67
.30	.60	.45	8.15	4.55	3.60
.30	.70	.50	−4.41	−9.05	4.64
.30	.80	.55	1.29	−1.46	2.75
.30	.90	.60	9.46	12.90	−3.44
.40	.50	.45	5.74	4.63	1.11
.40	.60	.50	−2.56	−4.12	1.56
.40	.70	.55	7.44	6.13	1.31
.40	.80	.60	−1.13	0.14	−1.28
.40	.90	.65	−19.36	−8.33	−11.03
.50	.60	.55	4.15	3.78	0.37
.50	.70	.60	4.45	4.75	−0.30
.50	.80	.65	12.26	15.22	−2.96
.50	.90	.70	6.13	18.00	−11.86
.60	.70	.65	6.15	6.89	−0.74
.60	.80	.70	2.03	5.90	−3.88
.60	.90	.75	13.76	25.98	−12.22
.70	.80	.75	3.92	6.74	−2.82
.70	.90	.80	−2.63	11.06	−13.69
.80	.90	.85	11.84	19.84	−8.00

Notes: $\bar{r}$ = mean observed correlation, $\%S^2_{e_r}$ = percentage by which the traditional estimator underestimates sampling variance (i.e., 100 − $[(S^2_{e_r}/S^2_e) \times 100]$), $\%S^2_{e_m}$ = percentage by which the mean estimator underestimates sampling variance (i.e., 100 − $[(S^2_{e_m}/S^2_e) \times 100]$). Tables including values for S^2_r (i.e., total sampling variance), $S^2_{e_r}$ (i.e., sampling variance computed using the traditional estimator shown in Equation 1), $S^2_{e_m}$ (i.e., sampling variance computed using the mean estimator shown in Equation 2), S^2_t (i.e., true variance computed using Equation 3), and S^2_e (i.e., sampling error variance, $S^2_r - S^2_t$) are available from the author.

TABLE 4

Underestimation of Sampling Error Variance Using the Traditional and Mean Estimators for the Heterogeneous Case (N = 140)

ρ_1	ρ_2	$\bar{r}$	$\%S^2_{e_r}$	$\%S^2_{e_m}$	$\%S^2_{e_r} - \%S^2_{e_m}$
.10	.20	.15	1.82	0.00	1.82
.10	.30	.20	−5.12	−8.40	3.29
.10	.40	.25	−7.64	−13.42	5.78
.10	.50	.30	−2.01	−10.60	8.59
.10	.60	.35	−10.53	−23.61	13.07
.10	.70	.40	−11.93	−28.58	16.65
.10	.80	.45	5.54	−9.60	15.14
.10	.90	.50	−71.27	−91.60	20.32
.20	.30	.25	2.84	1.23	1.61
.20	.40	.30	−1.58	−4.44	2.86
.20	.50	.35	−8.83	−13.87	5.04
.20	.60	.40	−10.72	−18.30	7.58
.20	.70	.45	15.58	8.13	7.45
.20	.80	.50	.03	−8.14	8.17
.20	.90	.55	−4.83	−7.75	2.92
.30	.40	.35	5.26	3.96	1.30
.30	.50	.40	−.77	−3.07	2.31
.30	.60	.45	.25	−3.20	3.45
.30	.70	.50	7.13	3.34	3.79
.30	.80	.55	−.20	−2.58	2.37
.30	.90	.60	−21.41	−16.17	−5.24
.40	.50	.45	4.90	4.00	0.90
.40	.60	.50	−1.83	−3.18	1.35
.40	.70	.55	−1.93	−3.19	1.26
.40	.80	.60	5.53	6.85	−1.32
.40	.90	.65	−16.72	−5.68	−11.04
.50	.60	.55	−3.48	−3.75	0.27
.50	.70	.60	3.06	3.36	−0.29
.50	.80	.65	10.34	13.45	−3.11
.50	.90	.70	4.94	17.12	−12.17
.60	.70	.65	−0.29	0.41	−0.70
.60	.80	.70	6.85	10.46	−3.61
.60	.90	.75	11.72	24.26	−12.54
.70	.80	.75	−4.78	−2.13	−2.65
.70	.90	.80	10.69	22.37	−11.68
.80	.90	.85	−1.33	6.98	−8.31

Notes: $\bar{r}$ = mean observed correlation, $\%S^2_{e_r}$ = percentage by which the traditional estimator underestimates sampling variance (i.e., 100 − $[(S^2_{e_r}/S^2_r) \times 100]$), $\%S^2_{e_m}$ = percentage by which the mean estimator underestimates sampling variance (i.e., 100 − $[(S^2_{e_m}/S^2_e) \times 100]$). Tables including values for S^2_r (i.e., total sampling variance), $S^2_{e_r}$ (i.e., sampling variance computed using the traditional estimator shown in Equation 1), $S^2_{e_m}$ (i.e., sampling variance computed using the mean estimator shown in Equation 2), S^2_t (i.e., true variance computed using Equation 3), and S^2_e (i.e., sampling error variance, $S^2_r - S^2_t$) are available from the author.

TABLE 5

Regression Analysis of Percentage by Which the Traditional Estimator Underestimates Sampling Error Variance on Mean Observed Correlation and Absolute Difference Between ρ_1 and ρ_2

Predictor	B	$SE\ B$	β
$\bar{r}$	11.16	5.56	0.18*
$\rho_1 - \rho_2$	−18.17	4.81	−0.34*

Notes: B = unstandardized regression coefficient; $SE\ B$ = standard error of B; β = standardized regression coefficient. $R = .39$, $p < .001$. Constant = .919. N =108.
*$p < .05$.

TABLE 6

Regression Analysis of Percentage by Which the Mean Estimator Underestimates Sampling Error Variance on Mean Observed Correlation and Absolute Difference Between ρ_1 and ρ_2

Predictor	B	$SE\ B$	β
$\bar{r}$	39.77	6.64	0.45*
$\rho_1 - \rho_2$	−36.28	5.75	−0.47*

Notes: B = unstandardized regression coefficient; $SE\ B$ = standard error of B; β = standardized regression coefficient. $R = .65$, $p < .001$. Constant = −9.38. N =108.
*$p < .05$.

For these conditions, the magnitude of underestimation was 1.26% for $N = 60$, 2.47% for $N = 100$, and 3.06% for $N = 140$.

The same aforementioned factors that affected the extent of underestimation of the traditional estimator also affected the degree of underestimation of the mean estimator. Table 6 summarizes a regression analysis of the extent to which the mean estimator underestimates sampling variance on $\bar{r}$ and $|\rho_1 - \rho_2|$. As expected, and similar to the traditional estimator, variance underestimation using the mean estimator had a positive relationship with $\bar{r}$ and a negative relationship with $|\rho_1 - \rho_2|$.

Results also show that the mean estimator *over*estimated sampling variance for numerous conditions. The mean estimator overestimated sampling error in 17 (i.e., 47.22%) conditions for $N = 60$, 17 (i.e., 47.22%) conditions for $N = 100$, and 22 (i.e., 61.11%) conditions for $N = 140$. This overestimation was quite severe for some cases. For instance, Table 2 ($N = 60$) shows an overestimation of as high as 37.99% for $\rho_1 = .1$ and $\rho_2 = .8$, and Table 3 ($N = 100$) shows an overestimation of as high as 40.45% for $\rho_1 = .1$ and $\rho_2 = .7$. For effect size conditions ranging from .1 to .5 for which overestimation took place, the mean magnitude of overestimation was 6.51% for $N = 60$, 7.05% for $N = 100$, and 8.97% for $N = 140$.

Relative Performance of Traditional and Mean Estimators

Homogenous case. In the homogenous case, the mean estimator has less negative bias than the traditional estimator. More precisely, the mean estimator exhibited less bias in 11 of the 15 homogeneous conditions for which $\rho_1 = \rho_2 < .6$. For these 15 conditions, the mean difference between the percent by which the traditional estimator underestimated the empirically derived variance and the percent by the which the mean estimator underestimated the empirically derived variance was 2.28% for $N = 60$, 1.35% for $N = 100$, and .97% for $N = 140$.

Heterogenous case. In the heterogenous case, the mean estimator also has less negative bias than the traditional estimator. More precisely, the mean estimator exhibited negative bias in 11 of the 30 cases for which ρ_1 and $\rho_2 < .6$, and the traditional estimator showed negative bias in 18 of these 30 cases. The mean difference between the percent by which the traditional estimator underestimated the empirically derived variance and the percent by the which the mean estimator underestimated the empirically derived variance for conditions showing underestimation was 1.68% for $N = 60$, 1.20% for $N = 100$, and .64% for $N = 140$.

Also in the heterogeneous case, the mean estimator yielded more positive bias than the traditional estimator. More precisely, the mean estimator exhibited positive bias in 19 of the 30 cases for which ρ_1 and $\rho_2 < .6$, and the traditional estimator showed positive bias in only 12 of these cases. The mean difference between the percentage by which the mean estimator overestimated the empirically derived variance and the percentage by the which the traditional estimator overestimated the empirically-derived variance for conditions showing overestimation was 1.04% for $N = 60$, 4.49% for $N = 100$, and 4.65% for $N = 140$.

Conclusions

Primary-level as well as meta-analytic researchers are concerned with the estimation of moderating effects (Aguinis, 1995; Aguinis, Petersen, & Pierce, 1999; Aguinis & Pierce, 1998a; Aguinis & Stone-Romero, 1997). In fact, numerous researchers assert that the accurate estimation of moderating effects is a strong indicator of the level of advancement and maturity of a scientific field (Aguinis, Boik, & Pierce, in press; Hall & Rosenthal, 1991). The results of the present study lead to the following three conclusions regarding the estimation of sampling error variance and moderating effects using meta-analysis.

First, as advocated by Schmidt, Hunter, and their colleagues (e.g., Hunter & Schmidt, 1994; Law et al., 1994), sampling error variance of the correlation coefficient computed using the traditional estimator shown in Equation 1 is systematically underestimated. The present simulation results indicate that the traditional sampling variance estimator has a systematic negative bias both in the homogeneous (i.e., no moderating effect in the population) and heterogenous (i.e., moderating effect in the population) cases. This negative bias was found in 85.52% of the homogeneous conditions and 60.19% of the heterogeneous conditions. Although the negative bias of the traditional estimator was more frequent in the homogeneous conditions, the magnitude of the bias was greater in the heterogeneous conditions. For the conditions for which sampling error was underestimated, the mean percentage by which the traditional estimator underestimated the empirically derived sampling error was 2.76% (range: .07 to 5.94%) in the homogeneous situations and 5.89% (range: .03 to 25.01%) in the heterogeneous situations.

The underestimation yielded by the traditional estimator may be quite large under some conditions frequently encountered by applied psychology and management researchers. For instance, assume a likely research scenario in the personnel selection literature in which sample size is 100, the population correlation for one moderator-based subgroup (e.g., women) is .2, and the population correlation for a second moderator-based subgroup (e.g., men) is .3. Using the traditional estimator would underestimate the actual sampling error variance by 4.18% (i.e., Table 3, line 9). Given this result, a researcher may attribute sampling error to "false moderators" because of this remaining proportion of variance that is not explained. In actuality, this variance is not caused by a moderating effect, but by an underestimation of sampling error variance. Similarly, a researcher implementing meta-analytic procedures other than VG may find that homogeneity statistics (e.g., Q, Hedges & Olkin, 1985; Q', Aguinis & Pierce, 1998b) are artificially inflated.

Second, results of the present study indicate that the proposed mean estimator shown in Equation 2 also has a negative bias in homogeneous situations (i.e., no moderating effect in the population). However, the mean estimator yields less negative bias (i.e., less underestimation of sampling variance) than the traditional estimator. This improvement regarding negative bias is most apparent for the effect size range most often encountered by applied psychology and management researchers (i.e., .1 to .5 range). For this range, the mean percent of improvement using the mean estimator is 1.61% (range: .38% to 3.21%).

Third, results regarding heterogeneous conditions indicate that (a) the mean estimator has less negative bias than the traditional estimator, and (b) the mean estimator has more positive bias than the traditional estimator. For the .1 to .5 effect size range, using the mean estimator yields a mean improvement regarding negative bias of 2.17% (range: .90% to

4.20%), and using the traditional estimator yields a mean improvement regarding positive bias of 5.06% (range: 2.31% to 9.21%).

Implications

The aforementioned results have implications for the assessment of moderating effects using meta-analysis because virtually all meta-analytic methods estimate the sampling variance of correlations. (The only exception seems to be the set of procedures proposed by Glass, McGaw, & Smith, 1981, which are rarely used at present.) For instance, the Hedges and Olkin (1985) approach incorporates sampling error in computing chi-square distributed homogeneity statistics.

The traditional estimator underestimates sampling error in both homogeneous and heterogeneous situations. Consequently, using the traditional estimator may lead researchers to conclude incorrectly that artifactual variance is due to potential moderator variables and, hence, false moderators may be "discovered." The use of the mean estimator provides an improvement over the traditional estimator regarding negative bias and the potential to commit Type I errors regarding moderator variable hypotheses. However, in heterogeneous situations the improvement regarding negative bias by using the mean estimator exists at the expense of positive bias. Thus, in heterogeneous situations, using the traditional estimator may lead to incorrectly attributing artifactual variance and finding "false moderators"; alternatively, using the mean estimator may lead to incorrectly attributing substantive moderator variance to sampling error and concluding that validity generalizes by failing to detect moderating effects.

As noted above, results indicate that the mean estimator outperforms the traditional estimator regarding negative bias in both homogeneous and heterogeneous conditions. However, are the percentage differences large enough to make a practical difference in the research community and change substantive conclusions of published meta-analyses that used the traditional estimator? A rule commonly used in the application of meta-analytic methods is to test for moderators when the total across-study variance accounted for by sampling error and other artifacts is less than a specific percentage of total variance. Usually, researchers follow Hunter and Schmidt's (1990; Hunter, Schmidt, & Jackson, 1982) 75% recommendation, but other cutoff percentages are also used (e.g., Organ & Ryan, 1995, used a 65% rule; Rasmussen & Loher, 1988, suggested various percentages depending on the number of studies in a meta-analytic database). As noted above, results for the more typical .1 to .5 effect size range in applied psychology and management indicate that using the mean estimator reduces negative bias by as much as

3.21% for homogeneous situations and 4.20% for heterogeneous situations. Thus, the conclusions of numerous published meta-analyses that used the traditional estimator and found that artifactual variance accounted for a percentage just under a specific cutoff (e.g., 70–75% for those using the 75% rule) may change had the mean estimator been used. In other words, in all these "near-miss" cases, using the mean estimator would have increased variance explained by artifacts over the cutoff percentage and, therefore, researchers would have concluded that tests for moderators were not warranted. Examples include meta-analyses on the relationship between personality and job performance (Hurtz & Donovan, 2000) and absence and turnover (Mitra, Jenkins, & Gupta, 1992). Results show that, given that the mean estimator reduces negative bias by as much as approximately 5% in many cases, the moderating effect of occupational category on the relationship between extraversion and job performance (cf. Hurtz & Donovan, 2000, Table 2, p. 874), the moderating effect of performance type on the relationship between openness to experience and job performance (cf. Hurtz & Donovan, 2000, Table 3, p. 875), the moderating effect of industry type on the relationship between absence and turnover (cf. Mitra, Jenkins, & Gupta, 1992, Table 2, p. 884), and the moderating effect of study duration on the relationship between absence and turnover (cf. Mitra et al., 1992, Table 2, p. 884) may be artifactual findings. In fact, it is likely that these moderating tests would not have been conducted had the mean estimator been used. It should be emphasized strongly that the above re-analysis is not intended in any way to devalue these excellent studies. These studies are merely used to illustrate how the present results shed new light on inferences drawn from previous meta-analyses that used the traditional estimator.

In the vast majority of published meta-analyses, researchers report the total across-study variance accounted for by artifacts (e.g., sampling error, measurement error, range restriction) without specifying the proportion due to sampling error. It is also likely that substantive conclusions about moderating effects of such meta-analyses whose overall artifactual variance fell just under a specific percentage need to be revisited. This is the case because results of Monte Carlo simulations indicate that sampling variance accounts for approximately 90% of artifactual variance in numerous meta-analyses conducted in applied psychology and management research (Koslowsky & Sagie, 1994). Thus, because sampling error is the main component of total artifactual variance, decisions about testing for moderating effects in near-miss cases in which total artifactual variance fell just under a specified percentage may have been also overturned had the mean estimator been used.

As noted above, the reduction in negative bias from use of the mean estimator is larger in the heterogenous case than in the homogeneous case. In the area of VG for aptitude and ability tests, there is evidence that the homogeneous case may hold in many jobs (Schmidt & Hunter, 1999). However, for all other meta-analyses (including VG meta-analyses on many other predictors), the heterogenous case is likely to be quite common—and perhaps nearly universal in the initial meta-analysis presented in a study (i.e., before there is any break out on potential moderator variables). This situation underscores the practical importance and relevance of the improvement in negative bias by using the mean estimator as compared to the traditional estimator.

Finally, as noted by an anonymous reviewer, beyond the question of meta-analysis or any other specific data analysis technique, the present results are also meaningful for statistics in general. That is, the most general implication is that in any statistical analysis in which $\bar{r}$ can be estimated, the reduction in the negative bias regarding sampling error estimation can take place by use of the mean estimation procedure. This is an implication for general statistics (or statistical analysis in general). Estimated $\bar{r}$s from a meta-analysis could be used to produce less negatively biased sampling error variance estimates in single individual studies that are conducted later (and are not part of the meta-analysis).

Limitations and Research Needs

First, the present Monte Carlo study used a multivariate random normal generator. Thus, although complying with the normality assumption is common practice in Monte Carlo investigations of meta-analytic methods (e.g., Aguinis & Whitehead, 1997; Callender & Osburn, 1988; Millsap, 1989), it should be acknowledged that the present study's results may not be generalizable to situations in which this assumption is not tenable (Oswald & Johnson, 1998). Thus, future research may address the extent of underestimation of sampling variance using the traditional and mean estimators when normality is violated.

Second, for the sake of simplicity, the choice was to create heterogeneous situations (i.e., true validity variance) including only two population correlations. This choice was made because it is the simplest and most parsimonious. It is unclear whether more moderator-based subgroups would lead to more or less underestimation of the sampling error variance. It is likely that the shape and location of the distribution of true validities will have some impact on the resulting bias in the estimator of the sampling error variance (Law et al., 1994). Future research could address how these features of the true validity distribution affect the performance of the traditional and mean sampling error variance estimators.

Third, the simulation did not incorporate corrections for methodological and statistical artifacts such as range restriction and measurement error. As noted in the Introduction, this is a strength of the present study because these artifacts also affect sampling variance (Aguinis & Whitehead, 1997; Millsap, 1989). By studying the relative performance of the two estimators in the absence of other artifacts, the performance difference between the two estimators was more clearly understood. However, an anonymous reviewer highlighted the fact that the present study did not investigate the relative performance of the estimators in the context of corrected correlation coefficients (cf. Aguinis & Pierce, 1998b; Raju, Burke, Normand, & Langlois, 1991). Aguinis and Pierce (1998b), and Raju et al. (1991) offered equations for estimating sampling variance of corrected correlations. These equations are different from the traditional estimator formula shown in Equation 1. Thus, the extent to which the present findings concerning sampling variance associated with uncorrected correlations generalize to sampling variance (and VG conclusions) associated with corrected correlations is an unanswered empirical question.

Finally, some might argue that the reduction in negative bias in sampling variance estimates resulting from the mean estimator is small in magnitude. However, as noted by an anonymous reviewer and described above, the difference is large enough to affect some meta-analytic conclusions. This reviewer noted that just as significant is the fact that an important hallmark of a science is continuous improvement in accuracy of measurements and estimations of theoretically important values. As stated by Hunter and Schmidt (1994), "given the vehemence of the debate concerning the hypothesis of situation-specific validity (e.g., Schmidt et al., 1993), any underestimation of sampling error variance has important theoretical implications because it can lead to non-zero (positive) estimates of standard deviation of true validities when, in fact, the true value of the standard deviation is zero" (pp. 173–174). Given that the present results showed that the underestimation of sampling variance can be quite substantial in some conditions, particularly in the heterogeneous case, future research could re-analyze previously published meta-analyses to assess (a) the extent to which sampling error variance may have been attributed to "false moderators," and (b) whether using the mean estimator instead of the traditional estimator may change substantive research conclusions. Issue (a) can be addressed by using information presented in Tables 1–4. More specifically, a researcher may estimate the extent to which sampling error variance has been underestimated in a specific meta-analysis by looking for a table entry matching a specific situation. Of course, researchers do not know a priori whether a specific meta-analytic database is homogeneous or heterogenous, and whether a moderator has two or more levels. Thus, researchers could estimate how much sampling error variance would be underestimated under different scenarios (i.e., sets of assumptions). Admittedly, information presented in Tables 1–4 do not address many situations (e.g., sample size larger than 140, more than two moderator-based subgroups). However, these tables may provide some useful benchmark information. Issue (b) can be addressed by re-analyzing meta-analytic databases using the estimator in Equation 2.

Closing Remarks and Recommendations

Meta-analytic procedures using the traditional estimator systematically underestimate the sampling error variance of correlations in both homogeneous (no population moderating effect) and heterogeneous (population moderating effect) situations. Because of this negative bias, using the traditional estimator artificially inflates Type I error rates regarding moderating effect hypotheses. Thus, using the traditional estimator may lead to the incorrect conclusion that moderators exist. Using the mean estimator instead of the traditional estimator yields a reduction in negative bias in both homogeneous and heterogeneous situations. However, the mean estimator shows positive bias in heterogeneous situations. Thus, using the mean estimator in heterogeneous situations may lead to a Type II error regarding moderating effect hypotheses and the incorrect conclusion that moderating effects do not exist. Researchers conducting meta-analyses do not know whether they are facing a homogeneous or heterogenous situation. Thus, given the trade-off between Type I and Type II error rates of using the traditional versus the mean estimator, the recommendation based on the present results is that researchers use the mean estimator shown in Equation 2 unless strong theory-based hypotheses regarding moderating effects exist.

REFERENCES

Aguinis H. (1994). A QuickBASIC program for generating correlated multivariate random normal scores. *Educational and Psychological Measurement, 54*, 687–689.

Aguinis H. (1995). Statistical power problems with moderated multiple regression in management reseach. *Journal of Management, 21*, 1141–1158.

Aguinis H, Boik RJ, Pierce CA. (in press). A generalized solution for approximating the power to detect effects of categorical moderator variables using multiple regression. *Organizational Research Methods*.

Aguinis H, Petersen SA, Pierce CA. (1999). Appraisal of the homogeneity of error variance assumption and alternatives to multiple regression for estimating moderating effects of categorical variables. *Organizational Research Methods, 2*, 315–339.

Aguinis H, Pierce CA. (1998a). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods, 1*, 296–314.

Aguinis H, Pierce CA. (1998b). Testing moderator variable hypotheses meta-analytically. *Journal of Management, 24*, 577–592.

Aguinis H, Stone-Romero EF. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology, 82*, 192–206.

Aguinis H, Whitehead R. (1997). Sampling variance in the correlation coefficient under indirect range restriction: Implications for validity generalization. *Journal of Applied Psychology, 82*, 528–538.

Callender JC, Osburn HG. (1988). Unbiased estimation of sampling variance of correlations. *Journal of Applied Psychology, 66*, 274–281.

Fisher RA. (1954). *Statistical methods for research workers*. Edinburgh, Scotland: Oliver and Boyd.

Glass GV, McGaw B, Smith ML. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Hall JA, Rosenthal R. (1991). Testing for moderator variables in meta-analysis: Issues and methods. *Communication Monographs, 58*, 437–448.

Hedges LV, Olkin I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hunter JE, Schmidt FL. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Hunter JE, Schmidt FL. (1994). Estimation of sampling error variance in the meta-analysis of correlations: Use of average correlation in the homogeneous case. *Journal of Applied Psychology, 79*, 171–177.

Hunter JE, Schmidt FL, Jackson GB. (1982). *Meta-analysis: Cumulating research findings across studies*. Newbury Park, CA: Sage.

Hurtz GM, Donovan JJ. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869–879.

Koslowsky M, Sagie A. (1994). Components of artifactual variance in meta-analytic research. PERSONNEL PSYCHOLOGY, *47*, 561–574.

Law KS, Schmidt FL, Hunter JE. (1994). A test of two refinements in procedures for meta-analysis. *Journal of Applied Psychology, 79*, 978–986.

Lent RH, Aurbach HA, Levin LS. (1971). Research design and validity assessment. PERSONNEL PSYCHOLOGY, *24*, 247–274.

Millsap RE. (1988). Sampling variance in attenuated correlation coefficients: A Monte Carlo study. *Journal of Applied Psychology, 73*, 316–319.

Millsap RE. (1989). Sampling variance in the correlation coefficient under range restriction: A Monte Carlo study. *Journal of Applied Psychology, 74*, 456–461.

Mitra A, Jenkins GD, Gupta N. (1992). A meta-analytic review of the relationship between absence and turnover. *Journal of Applied Psychology, 77*, 879–889.

Organ DW, Ryan K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behaviors. PERSONNEL PSYCHOLOGY, *48*, 775–802.

Oswald FL, Johnson JW. (1998). On the robustness, bias, and stability of statistics from meta-analysis of correlation coefficients: Some initial Monte Carlo findings. *Journal of Applied Psychology, 83*, 164–178.

Raju NS, Burke MJ, Normand J, Langlois GM. (1991). A new meta-analytic approach. *Journal of Applied Psychology, 76*, 432–446.

Rasmussen JL, Loher BT. (1988). Appropriate critical percentages for the Schmidt and Hunter meta-analysis procedure: Comparative evaluation of Type I error rate and power. *Journal of Applied Psychology, 73*, 683–687.

Rosenthal R. (1991). *Meta-analytic procedures for social research (revised ed.)*. Beverly Hills, CA: Sage.

Salgado JF. (1998). Sample size in validity studies of personnel selection. *Journal of Occupational and Organizational Psychology, 71*, 161–164.
Schmidt FL. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist, 47*, 1173–1181.
Schmidt FL, Hunter JE. (1999). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

APPENDIX

The formula shown in Equation 3 for the true validity variance in the case of two different values for the true validity can be derived as follows. Let ρ_1 and ρ_2 be the two different true validity values. Suppose that k replicates of each value exist, corresponding to k validity studies with each of these values. The average true validity across the $2k$ studies is then

$$\overline{\rho} = \frac{k\rho_1 + k\rho_2}{2k} = \frac{\rho_1 + \rho_2}{2}.$$

The variance in the true validities across the 2k studies is

$$S_t^2 = \frac{k(\rho_1 - \overline{\rho})^2 + k(\rho_2 - \overline{\rho})^2}{2k} = \frac{(\rho_1 - \overline{\rho})^2 + (\rho_2 - \overline{\rho})^2}{2}.$$

Noting that $\rho_1 - \overline{\rho} = (\rho_1 - \rho_2)/2$ and $\rho_2 - \overline{\rho} = (\rho_2 - \rho_1)/2$, and substituting these expressions into the above variance formula, results in

$$S_t^2 = \frac{2(\rho_1 - \rho_2)^2}{8} = \frac{\rho_1^2 + \rho_2^2 - 2\rho_1\rho_2}{4}.$$

This formula works for any value of k, the number of validity studies per validity value, as long as the number of studies corresponding to each validity value is the same (i.e., k is a constant).