Psychological Bulletin
1983, Vol. 93, No. 2, 388-395

0033-2909/83/9302-0388$00.75

A Random Effects Model for Effect Sizes

Larry V. Hedges
Department of Education, University of Chicago

Recent interest in quantitative research synthesis has led to the development of rigorous statistical theory for some of the methods used in meta-analysis. Statistical theory proposed previously has stressed the estimation of fixed but unknown population effect sizes (standardized mean differences). Theoretical considerations often suggest that treatment effects are not fixed but vary across different implementations of a treatment. Such considerations lead naturally to a random effects model (analogous to random effects analysis of variance) in which the population effect sizes are not fixed but instead are sample realizations from a distribution of possible population effect sizes. A large sample test that the effect-size distribution has zero variance is given. An analogy to variance component estimation is used to derive an unbiased estimator of the variance of the effect-size distribution. An example shows that these methods may suggest insights that are not available from inspection of means and standard deviations of effect-size estimates.

There has been intense interest in quantitative methods for research synthesis in the years since Glass (1976) proposed the use of statistical methods in research reviews. The most popular procedure seems to involve the use of the effect-size index proposed by Glass (1976). The method involves the calculation of an estimate of effect size (the standardized difference between experimental and control group means) from each study. The average of effect-size estimates across studies is used as an index of the overall effect size across studies. Substantive conclusions are usually drawn based on the magnitude of the average of the effect-size estimates. Some statistical theory for estimation of effect size was given by Hedges (1981, 1982), who derived the sampling distribution of effect-size estimators and showed how to construct confidence intervals for the effect size when a series of studies share a common population effect size.

One extension of analyses based on the average effect size used the idea that characteristics of a study may influence the magnitude of its effect size. Glass (1978) recommended a general strategy of coding the characteristics of studies as a vector of predictor variables and then regressing the effect-size estimates on these predictors to determine the relationship between characteristics of studies and their effect sizes. For example, Smith and Glass (1977) used ordinary linear regression to determine the relationship between characteristics of studies (e.g., type of therapy, duration of treatment, internal validity of the study) and effect size in their meta-analysis of psychotherapy outcome studies. The same methods have been used in quantitative syntheses of gender effects in decoding nonverbal cues (Hall, 1978), the relationship between motivation and academic achievement (Uguroglu & Walberg, 1979), and the effects of goal structures on achievement (Johnson, Maruyama, Johnson, Nelson, & Skon, 1981). In general these analyses have found few consistent relations between study characteristics and effect size.

One explanation for the elusiveness of these relations derives from a proposal by Cronbach (1980). He suggests that evaluation studies should consider a model in which each treatment site (or study) is a sample realization from a universe of related treatments. Thus when evaluators look at "replications" of a treatment across sites, they observe many different treatments, each sampled from some universe of possible treatments. If these variations in treatment are

This research was supported by the Spencer Foundation.

Requests for reprints should be sent to Larry V. Hedges, Department of Education, University of Chicago, 5835 S. Kimbark Avenue, Chicago, Illinois 60637.

more or less effective in producing the outcome, then variations in the true (population) effect of the treatment would be expected. Such variations might be expected to attenuate any relationship between a fixed characteristic (such as age or sex of subjects) and the outcome variable. Note that this model implies that there is no single true or population effect of the "treatment" across studies. Rather there is a distribution of true effects: Each treatment implementation (site) has its own unique true effect. One may speak of the average true effect of the treatment as an index of overall efficacy. The average true effect is not very meaningful though, without some measure of the variation in the true effect of the treatment. One might find, for example, that the average of the true effects was larger than zero, but the true effect of the treatment was negative in nearly half the implementations. The problem of estimating the variability in the true effects is complicated by the fact that the true effect in any treatment site (or study) is never known. We must estimate that true effect from sample data, and that estimate will itself be subject to sampling fluctuations.

The implication of this "random effects" model for quantitative research synthesis is that the underlying population effect sizes will not be constant across a series of studies that replicate the same treatment. Hedges (1982) developed a test of homogeneity of effect sizes. This procedure tests whether the observed estimates of effect size vary by more than would be expected if all studies shared a common population effect size. Applications of this test in one quantitative research synthesis (Giaconia & Hedges, in press) suggest that effect sizes of even carefully selected and apparently identical studies are often much more variable than would be expected if they shared a common underlying effect size.

The purpose of this article is to present an analogue to random effects analysis of variance for effect-size analyses. In this model we assume that the population values of the effect size are sampled from a distribution of effect-size parameters. Thus the observed variability in sample estimates of effect size is partly due to the variability in the underlying population parameters and partly due to the sampling error of the estimator about the parameter value. This model is appropriate when the studies used in the analysis are representative (if not a random sample) of a larger population and the researcher wants to generalize to that larger population. A statistical test is given to test the hypothesis that the variance in population effect sizes is zero. A method for estimating the variance of population effect sizes (the parameter variance component) is given, and the variance component estimates are shown to be unbiased.

Model and Notation

Cohen (1977) proposed a population measure d of effect size in connection with the t test for the difference between means. Glass (1976) proposed the quantitative synthesis of the results of a collection of experimental/control group studies by estimating d for each study and then combining the estimates across studies. The statistical analyses used in such studies typically involve the use of a t or F test to test for the difference in group means. In this article, we assume that the requirements for the validity of the two-sample t test are met in each study. We start by stating these assumptions explicitly.

Suppose that the data arise from a series of k independent studies, where each study compares an experimental group (E) with a control group (C). Let Y_{ij}^{E} and Y_{ij}^{C} be the jth observations on the ith experiment from the experimental and control groups, respectively. Assume that for fixed i, Y_{ij}^{E} and Y_{ij}^{C} are independently normally distributed with means μ_i^{E} and μ_i^{C} and common variance σ_i^2; that is,

$$Y_{ij}^{E} \sim \mathcal{N}(\mu_i^{E}, \sigma_i^2), \quad j = 1, \ldots, n_i^{E}, \quad i = 1, \ldots, k,$$

and

$$Y_{ij}^{C} \sim \mathcal{N}(\mu_i^{C}, \sigma_i^2), \quad j = 1, \ldots, n_i^{C}, \quad i = 1, \ldots, k,$$

where n_i^{E} and n_i^{C} are the experimental and control group sample sizes in the ith study. In this notation the effect size for the ith study (δ_i) is defined as

$$\delta_i = \frac{\mu_i^E - \mu_i^C}{\sigma_i}, \quad (1)$$

where we use the Greek letter δ (instead of d) to emphasize that this effect size is a population parameter.

Previous work on statistical theory for effect-size analyses has treated the δ_i, $i = 1, \ldots, k$ as fixed, but unknown constants. In this article we depart from previous practice and treat the δ_i, $i = 1, \ldots, k$ as realizations of a random variable Δ. Each δ_i considered by itself is indeed a population parameter for the ith study. The studies, however, are considered as a sample from a population of studies with a distribution of δ_i values. By sampling studies, we are, in effect, obtaining a sample of δ_i values. In principle, a replication of our research synthesis would result in a different sample of studies and therefore different δ_i values. Generalization to the population of δ_i values is possible from either sample, however. This conception of δ is analogous to the idea of the true score in classical test theory. In one sense, the true score for a particular individual is a population parameter describing the hypothetical distribution of observed scores for that individual. In another sense, the individual is sampled from a human population with a distribution of true scores. Thus the individual true score is both a parameter and a sample realization from a distribution of true scores.

The model proposed herein is also analogous to that of the random effects analysis of variance in which treatment levels are sampled from a population of possible treatments. In random effects analysis of variance the object of the statistical analysis is to test whether the treatment variance component is zero and to estimate that variance component. In the present context, the object of the statistical analysis is to estimate the mean $\bar{\Delta}$ of the distribution of δ_i values, test the hypothesis that the variance σ_Δ^2 of Δ is zero, and estimate σ_Δ^2.

Estimating Effect Size

The definition of effect size given in Equation 1 above defines a population parameter δ_i in terms of other population parameters μ_i^E, μ_i^C, and σ_i, and thus we will have to estimate δ_i from sample data. Glass (1976) proposed estimation of δ_i by essentially replacing μ_i^E, μ_i^C, and σ_i by their sample analogues in Equation 1. Hedges (1981) studied the properties of Glass's estimator and showed that it was biased. He derived a simple unbiased estimator that is always more precise than Glass's estimator. This unbiased estimator is given by

$$g_i = \frac{c(m_i)(\bar{Y}_i^E - \bar{Y}_i^C)}{S_i}, \quad i = 1, \ldots, k, \quad (2)$$

where $\bar{Y}_i^E$ and $\bar{Y}_i^C$ are the experimental and control group sample means, S_i^2 is the usual pooled within-group estimate of the sample variance

$$S_i^2 = \frac{(n_i^E - 1)(S_i^E)^2 + (n_i^C - 1)(S_i^C)^2}{n_i^E + n_i^C - 2},$$

$m_i = n_i^E + n_i^C - 2$, and

$$c(m) = \frac{\Gamma(m/2)}{\sqrt{m/2}\,\Gamma[(m-1)/2]}. \quad (3)$$

A table of exact values of $c(m)$ is given in Hedges (1981), but an excellent approximation (with maximum error less than .0005 when $m \geq 10$) is

$$c(m) \approx 1 - \frac{3}{4m - 1}. \quad (4)$$

The Variance of Estimates of Effect Size

Hedges (1981) showed that the sampling variance of g_i for fixed δ_i (i.e., the variance conditional on δ_i) is

$$\sigma_i^2(\delta_i) = \frac{a_i}{\tilde{n}_i}[1 + \tilde{n}_i\delta_i^2] - \delta_i^2, \quad (5)$$

where $\tilde{n}_i = n_i^E n_i^C/(n_i^E + n_i^C)$,

$$a_i = \frac{m_i[c(m_i)]^2}{(m_i - 2)}, \quad (6)$$

and $m_i = n_i^E + n_i^C - 2$. Since δ_i is a random variable, the unconditional sampling variance of g_i involves the variance σ_Δ^2 of the population effect sizes. Because g_i is an unbiased estimator of δ_i, it is easy to show (see the Appendix) that the unconditional sampling variance of g_i is given by

$$\sigma_\Delta^2 + \sigma_i^2(\delta_i). \quad (7)$$

Note that Expression 7 would not be true if a biased estimator (such as Glass's) of δ_i were used in place of g_i, because bias in the estimator may result in a nonzero covariance between sampling error and Δ.

Equation 5 for the conditional sampling variance of g_i shows that this variance depends only on sample size and on δ_i. This variance could be estimated by using g_i as an estimate of δ_i in Equation 5. Similarly, the unconditional sampling variance of the g_i could be obtained from the variance of a sample of g_i values. Therefore an estimate of σ_Δ^2 could be obtained by subtraction in a manner similar to that used to obtain estimates of variance components in random effects analysis of variance.

Estimation of Effect-Size Variance

In random effects analysis of variance, the expected values of the mean squares are expressed in terms of variance components. The expected values of the mean squares are then replaced with their sample values, and the equations are solved for the variance components. This process gives unbiased estimates of the variance components. The rationale for estimating σ_Δ^2 is the same as for the estimation of variance components in random effects analysis of variance. The expected value of the sample variance of the g_i is expressed as a function of variance components including the conditional sampling variances of the g_i, $i = 1, \ldots, k$. Unbiased estimators of conditional sampling variances will be given. Then the expected value of the unconditional variance of the g_i is replaced by the observed variance of the g_i, and the equations are solved for σ_Δ^2. This process results in an unbiased estimator of σ_Δ^2. It is worth noting that many quadratic functions of the observations can be used to estimate variance components in random effects analysis of variance (see, e.g., Searle, 1971). Similarly, many other estimates of σ_Δ^2 could be obtained by using quadratic functions of the g_i other than the sample variance.

The Expected Value Equations

Let $g_1, \ldots, g_k$ be the unbiased estimates (Equation 2) of effect size from k independent studies and define

$$S_g^2 = \frac{1}{k-1}\sum_{i=1}^{k}(g_i - \bar{g})^2,$$

where $\bar{g}$ is the unweighted mean of the g_i, $i = 1, \ldots, k$. A direct application of a theorem in Searle (1971, p. 55) yields the expected value of S_g^2 as

$$E[S_g^2] = \sigma_\Delta^2 + \frac{1}{k}\sum_{i=1}^{k}\sigma_i^2(\delta_i). \qquad (8)$$

A direct computation (see the Appendix) shows that an unbiased estimator of $\sigma_i^2(\delta_i)$ is

$$\hat{\sigma}_i^2 = \frac{1}{\tilde{n}_i} + (1 - 1/a_i)g_i^2 \qquad (9)$$

where a_i is given by Equation 6 and $\tilde{n}_i = n_i^E n_i^C/(n_i^E + n_i^C)$. Combining Equations 8 and 9 and solving for σ_Δ^2 yields

$$\hat{\sigma}_\Delta^2 = S_g^2 - \frac{1}{k}\sum_{i=1}^{k}\left[\frac{1}{\tilde{n}_i} + (1 - 1/a_i)g_i^2\right], \qquad (10)$$

which is an unbiased estimator of σ_Δ^2.

Testing that Parameter Variance is Zero

It is sometimes useful to test the hypothesis that $\sigma_\Delta^2 = 0$. This test is analogous to the F tests in random effects analysis of variance. Note that if $\sigma_\Delta^2 = 0$, then $\delta_1 = \delta_2 = \cdots = \delta_k = \delta$. Therefore a test of $\sigma_\Delta^2 = 0$ in this model corresponds to testing homogeneity of effect size in models with fixed parameters (Hedges, 1981, 1982). An asymptotic test that $\sigma_\Delta^2 = 0$ uses the test statistic

$$H = \sum_{i=1}^{k}\frac{(g_i - g_\cdot)^2}{\sigma_i^2(g_i)}, \qquad (11)$$

where $\sigma_i^2(g_i)$ is the conditional variance of g_i given in Equation 5 (using the sample estimate g_i for δ_i) and $g_\cdot$ is the weighted mean effect size given by

$$g_\cdot = \frac{\sum_{i=1}^{k}\frac{g_i}{\sigma_i^2(g_i)}}{\sum_{i=1}^{k}\frac{1}{\sigma_i^2(g_i)}}. \qquad (12)$$

When the null hypothesis that $\sigma_\Delta^2 = 0$ is true, the test statistic H has an asymptotic (as n_i^E and $n_i^C \to \infty$) chi-square distribution with $(k - 1)$ degrees of freedom. Therefore the test of $\sigma_\Delta^2 = 0$ involves comparing the obtained value of H with the $100(1 - \alpha)$ percent critical value of the chi-square distribution on $(k - 1)$ degrees of freedom. If the obtained value of H exceeds this critical value, then we reject the hypothesis that $\sigma_\Delta^2 = 0$ at the $100(1 - \alpha)\%$ level of significance. Although this is an asymptotic test, extensive simulations (Hedges, 1982) suggest that the nominal significance levels are quite accurate when the experimental and control group sample sizes exceed 10 and $|\delta| \leq 1.5$. Many bodies of research are likely to have sample sizes and effect sizes that meet these requirements.

Because H has an asymptotic chi-square distribution with $(k - 1)$ degrees of freedom when $\sigma_\Delta^2 = 0$, $E[H] = k - 1$ under this condition. When $\sigma_\Delta^2 > 0$ the statistic H does not have a chi-square distribution, but it is possible to show that $H \to \infty$ as n^E, $n^C \to \infty$ if $\sigma_\Delta^2 \neq 0$. Hence in the random effect size model of this article, the test statistic H will tend to be larger and will reject the null hypothesis more often when $\sigma_\Delta^2 > 0$ than when $\sigma_\Delta^2 = 0$.

Estimating the Mean Effect Size

Each of the estimators g_i, $i = 1, \ldots, k$ is an unbiased estimator of δ_i and therefore the unweighted average $\bar{g}$ is also an unbiased estimator of $\bar{\Delta}$, the mean of the distribution of Δ. However, the unweighted mean $\bar{g}$ is not the most precise estimator of $\bar{\Delta}$. If some of the k experiments have larger sample sizes they will yield more precise estimates of the corresponding δ_i than will experiments based on smaller samples. Thus it makes sense to give more weight to the estimators from the larger experiments than to those from the smaller ones. It is easy to show (Hedges, 1982) that the most precise estimate of $\bar{\Delta}$ is given by

$$g_\cdot = \frac{\sum_{i=1}^{k} \frac{g_i}{v_i}}{\sum_{i=1}^{k} \frac{1}{v_i}}, \qquad (13)$$

where v_i is the unconditional variance of g_i, $i = 1, \ldots, k$. A problem with Equation 13 is that the optimal weights depend on the unconditional variances of the g_i, which in turn depend on the unknown δ_i and the unknown variance component σ_Δ^2. One approach to the problem of weighting is to estimate the weights from sample data. Such an approach generally results in a biased estimator (even though the g_i are unbiased) although the bias is likely to be small if σ_Δ^2 is not large (Hedges, 1982). Another approach to the problem of weighting is to assign weights on some a priori basis, such as approximating v_i by $1/(n_i^E + n_i^C)$. This approach will give a simple unbiased estimator of $\bar{\Delta}$ that is slightly less efficient than the optimal weighted estimator.

The variance and distribution of the weighted estimator will generally depend on σ_Δ^2, therefore no parametric significance tests for $\bar{\Delta}$ are possible without placing restrictions on the distribution of Δ.

Computing Estimates and Test Statistics

The estimates and test statistics described in this article can easily be computed using any standard packaged computer program. Let each study be a separate case and define seven variables for each case:

$$\text{VAR1} = g_i,$$

$$\text{VAR2} = \frac{g_i}{\sigma_i^2(g_i)},$$

$$\text{VAR3} = \frac{g_i^2}{\sigma_i^2(g_i)},$$

$$\text{VAR4} = \frac{1}{\sigma_i^2(g_i)},$$

$$\text{VAR5} = \frac{n_i^E + n_i^C}{n_i^E n_i^C} + (1 - 1/a_i)g_i^2,$$

$$\text{VAR6} = w_i \text{(a weight assigned to the study, like } n_i^E + n_i^C),$$

$$\text{VAR7} = w_i g_i.$$

The homogeneity test statistic H can be expressed as

$$H = \sum_{i=1}^{k} \frac{g_i^2}{\sigma_i^2(g_i)} - \frac{\left(\sum_{i=1}^{k} \frac{g_i}{\sigma_i^2(g_i)}\right)^2}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2(g_i)}}$$

$$= \sum \text{VAR3} - \frac{(\sum \text{VAR2})^2}{\sum \text{VAR4}}. \quad (14)$$

Equation 14 is algebraically equivalent to Equation 11. The estimate (Equation 10) of σ_Δ^2 can be expressed as

$$\hat{\sigma}_\Delta^2 = S_g^2 - \frac{1}{k}\sum_{i=1}^{k} \frac{1}{\tilde{n}_i} + (1 - 1/a_i)g_i^2$$

$$= S_g^2 - \frac{1}{k}\sum \text{VAR5},$$

where $\tilde{n}_i = n_i^E n_i^C/(n_i^E + n_i^C)$ and S_g^2 is the sample variance of the g_i, $i = 1, \ldots, k$. The weighted estimator of $\bar{\Delta}$ is given by

$$\frac{\sum_{i=1}^{k} w_i g_i}{\sum_{i=1}^{k} w_i} = \frac{\sum \text{VAR7}}{\sum \text{VAR6}}.$$

Therefore the estimates described in this article can easily be computed from standard packages or calculators that can provide means and variances.

Example

The techniques described in this article were applied to 24 estimates of effect size obtained from a meta-analysis of the effects of open education (Hedges, Giaconia, & Gage, Note 1). The effect-size estimates reported in Table 1 are indices of the effect of open education on mathematics achievement. Twelve of the effect-size estimates were obtained from a sample of studies using random assignment of subjects to treatment groups. The other 12 effect-size estimates were obtained from a sample of studies that did not use random assignment. The average effect-size estimates for these two groups of studies are almost identical, with $\bar{g} = -.0197$ for the studies with random assignment and $\bar{g} = -.0120$ for the studies that did not use random assignment. The (unweighted) standard deviations of the effect-size estimates are also similar; $S_g = .278$ for the studies with random assignment and $S_g = .438$ for the studies without random assignment.

Investigators might be tempted to conclude that the studies with random assignment give essentially the same answer as the quasi experiments. Several meta-analyses (e.g., Smith & Glass, 1977, or Glass, 1978) have reported a similar finding; that is, that well controlled and poorly controlled studies yielded essentially the same average effect sizes.

Examination of the homogeneity test statistics (Equation 11) for the two groups of studies suggests that the variance σ_Δ^2 is not zero in either group. These values are $H = 23.09$ for the studies using random assignment of subjects to treatments and $H = 90.61$ for the studies without random assignment. Comparing these values with critical values for the chi-square distribution, we see that both are significant at the $\alpha = .02$ level, and the latter value is significant beyond the $\alpha = .001$ level. Thus in each case we reject the hypothesis that $\sigma_\Delta^2 = 0$. Calculation of the unbiased estimate (Equation 10) of σ_Δ^2 for each group yields $\hat{\sigma}_\Delta^2 = .036$ for the studies with random assignment and $\hat{\sigma}_\Delta^2 = .162$ for the studies without random assignment.

The difference in the magnitude of the estimates of σ_Δ^2 suggests that the population effect sizes are far more variable among the studies that did not use random assignment. Thus the statement that the two groups of studies "yield the same results" needs to be qualified. Whereas the average effect-size estimates are the same for the two groups of studies, the variation of the true (population) effect sizes is much less for the studies with random assignment. This finding makes conceptual sense because the preexisting differences between groups are not controlled in the quasi experiments. If the studies that did not have random assignment exhibited a distribution of real preexisting differences, then these differences would also be reflected in the distribution of (posttest) effect-size estimates. Unfortunately the studies from which these effect-size estimates were obtained did not provide information on pretest scores that might have helped to investigate this hypothesis. These findings do suggest that

Table 1
Effect Size Estimates From 24 Studies of the Effect of Open Education on Mathematics Achievement

Randomized experiments					Studies without random assignment				
Study	n^E	n^C	g	$\frac{1}{\tilde{n}} + (1 - 1/a)g^2$	Study	n^E	n^C	g	$\frac{1}{\tilde{n}} + (1 - 1/a)g^2$
1	57	112	.146	.027	1	42	100	.506	.035
2	48	86	.248	.033	2	74	44	.330	.037
3	180	180	.049	.011	3	89	425	.109	.014
4	80	61	−.313	.029	4	40	40	−.380	.051
5	131	138	−.267	.015	5	133	127	.260	.016
6	90	90	−.110	.022	6	76	105	−.488	.023
7	41	52	.124	.044	7	80	81	.584	.026
8	60	55	−.151	.035	8	72	72	.387	.028
9	10	10	.529	.209	9	120	150	−.345	.015
10	138	160	.190	.014	10	120	167	−.718	.015
11	156	50	−.362	.027	11	38	52	.006	.046
12	56	56	−.320	.036	12	40	40	−.395	.051

Note. E = experimental group; C = control group.

casual inspection of means and standard deviations of effect-size estimates is not sufficient to distinguish whether the studies in a series share a common effect size, nor does such inspection necessarily provide much evidence about the variability of the true (population) effect-size distribution.

Reference Note

1. Hedges, L. V., Giaconia, R. M., & Gage, N. L. *The empirical evidence on the effectiveness of open education* (Final report, Vol. 2). Stanford, Calif.: Stanford University Program on Teaching Effectiveness Meta-Analysis Project, 1981.

References

Cohen, J. *Statistical power analysis for the behavioral sciences.* New York: Academic Press, 1977.

Cronbach, L. J. *Toward reform of program evaluation.* San Francisco: Jossey-Bass, 1980.

Giaconia, R. M., & Hedges, L. V. Identifying features of effective open education. *Review of Educational Research,* in press.

Glass, G. V. Primary, secondary, and meta-analysis of research. *Educational Researcher,* 1976, *5,* 3–8.

Glass, G. V. Integrating findings: The meta-analysis of research. In L. S. Schulman (Ed.), *Review of Research in Education* (Vol. 5). Itasca, Ill.: Peacock, 1978.

Hall, J. A. Gender effects in decoding nonverbal cues. *Psychological Bulletin,* 1978, *85,* 845–857.

Hedges, L. V. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics,* 1981, *6,* 107–128.

Hedges, L. V. Estimation of effect size from a series of independent experiments. *Psychological Bulletin,* 1982, *92,* 490–499.

Johnson, N. L., & Kotz, S. *Distributions in statistics: Continuous univariate distributions—2.* New York: Wiley, 1970.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin,* 1981, *89,* 47–62.

Searle, S. R. *Linear models.* New York: Wiley, 1971.

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. *American Psychologist,* 1977, *32,* 752–760.

Uguroglu, M. E., & Walberg, H. J. Motivation and achievement: A quantitative synthesis. *American Educational Research Journal,* 1979, *16,* 375–390.

Appendix

The technical arguments used in this article all depend on an expression for the unconditional sampling variance of the g_i. Write the random variable g as $g = \Delta + e$, where $e = g - \Delta$ is the sampling error about the sample realization of Δ. Then

$$\mathrm{Var}(g) = \mathrm{Var}(\Delta) + \mathrm{Var}(e) + 2\mathrm{Cov}(\Delta, e),$$

and we need only show that Δ and e are uncorrelated to show that the unconditional variance of g is the variance of Δ plus the conditional variance of g, that is, the variance given in Equation 5. The covariance is $\text{Cov}(\Delta, e) = E[\Delta e] - E[\Delta]E[e]$. Since g is conditionally unbiased $E(e|\Delta) = 0$. Evaluate each of the expectations as follows:

$$E[\Delta e] = E\{E[\Delta e|\Delta]\} = E\{\Delta E[e|\Delta]\} = 0,$$

and

$$E[e] = E\{E[e|\Delta]\} = 0,$$

which imply that $\text{Cov}(\Delta, e) = 0$.

The unbiased estimator of the conditional sampling variance of g_i is obtained by finding an unbiased estimator of δ_i^2. Since $\sqrt{\tilde{n}_i} g_i$ has the noncentral t distribution with noncentrality parameter $\sqrt{\tilde{n}_i}\delta$ (Hedges, 1981), it follows that $\tilde{n}_i g_i^2$ has the noncentral F distribution with noncentrality parameter $\lambda = \tilde{n}_i\delta^2$. Using the expectation of the noncentral F distribution (Johnson & Kotz, 1970, p. 190), we obtain

$$E(g_i^2) = \frac{a_i}{\tilde{n}_i}(1 + \tilde{n}_i\delta_i^2),$$

where a_i is given in Equation 6. Solving this expression for δ_i^2 yields the unbiased estimator

$$\frac{g^2}{a_i} - \frac{1}{\tilde{n}_i}$$

for δ_i^2. Substitution of this estimator for δ_i^2 in Equation 5 for $\sigma_i^2(\delta_i)$ gives Equation 9.

Received May 11, 1982
Revision received June 24, 1982 ■

Editorial Consultants for This Issue: Quantitative Methods in Psychology

Bakan, David
Bock, R. D.
Bryant, Fred B.
Burt, R. S.
Camilli, Gregory
Cohen, Jacob
Cooper, Harris M.
Corballis, M. C.
Costanza, Michael C.
Doksum, Kjell
Elashoff, Janet
Frey, Peter J.
Greenwald, Anthony G.
Gross, Alan L.
Guthrie, Donald
Hamer, Robert M.
Hays, William L.
Herr, David G.
Hettmansperger, Thomas P.
Hocking, R. R.
Horn, John L.
Horst, Paul
Koehler, Kenneth J.
Lee, Howard B.
Losey, George S.
Mendoza, Jorge L.
O'Brien, Ralph G.
Olson, Craig A.
Osburn, Hobart G.
Phillip, Abraham
Pitz, Gordon F.
Rabkin, Judith Godwin
Rock, Donald A.
Saal, Frank E.
Shanteau, James
Simon, Gary A.
Stoline, Michael
Takane, Yoshio
Tatsuoko, Maurice
Wainer, Howard
Wattenburg, Franklin
Wickens, Thomas Dow
Wilson, Victor L.