DOCUMENT RESUME

ED 452 210		TM 032 495
AUTHOR	Hess, Brian; Olejnik, Stephen	
TITLE	Revisiting the Efficacy of Glass's Estimator of Effect Size for Program Impact Analysis.	
PUB DATE	2001-04-14	
NOTE	30p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).	
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC02 Plus Postage.	
DESCRIPTORS	*Effect Size; Estimation (Mathematics); Hypothesis Testing; *Meta Analysis; *Program Evaluation; Statistical Significance	
IDENTIFIERS	*Impact Evaluation; *Standardized Mean Difference	

ABSTRACT

Null hypothesis significance testing alone is not sufficient for program evaluation. To assess program impact adequately, effect sizes should be reported and interpreted in the context of similar or alternate programs. A popular effect size for the treatment-control group design has been the standardized mean difference, delta. Several estimators of delta (e.g., J. Cohen's "d") are known, and their efficacy under specific data conditions has been studied. In evaluation studies, the Glass estimator of delta (G. Glass, 1977) has been recommended, but its efficacy is known only under ideal data conditions. Using simulated data, this study assessed the efficacy of Glass's effect size when population variances were unequal, distributions were nonnormal, and group sizes were unequal. Implications for using Glass's effect size when conducting an impact analysis are highlighted. (Contains 6 tables, 6 figures, and 20 references.) (Author/SLD)

Running Head: EFFICACY OF GLASS'S EFFECT SIZE

Revisiting the Efficacy of Glass's Estimator of Effect Size for Program Impact Analysis

Brian Hess and Stephen Olejnik

The University of Georgia

Paper Presented at the Annual Meeting of the American Educational Research Association, Seattle, WA. April 14, 2001.

Correspondence concerning this paper may be sent to Brian Hess, 12711 Arbor Isle Drive, Temple Terrace, FL 33637. Email: bhess@coe.uga.edu.

Abstract

Null hypothesis significance testing alone is not sufficient for program evaluation. To adequately assess program impact, effect sizes should be reported and interpreted in the context of similar or alternate programs. A popular effect size for the treatment-control group design has been the standardized mean difference, δ. Several estimators of δ (e.g., Cohen's *d*) are known and their efficacy under specific data conditions have been studied. In evaluation studies, the Glass estimator of δ has been recommend, but its efficacy is known only under ideal data conditions. Using simulated data, this study assessed the efficacy of Glass's effect size when population variances were unequal, distributions were nonnormal, and group sizes were unequal. Implications for using Glass's effect size when conducting an impact analysis are highlighted.

Revisiting the Efficacy of Glass's Estimator of Effect Size for Program Impact Analysis

Not only has null hypothesis significance testing (NHST) been insufficient in substantive research, it has been insufficient in program evaluation as well. In substantive research, many have criticized NHST (e.g., Falk & Greenbaum 1995; Huberty & Pike 1999; Schmidt 1996). Subsequently, there has been a move toward measuring and reporting effect sizes in order to address the "practical significance" or meaningfulness of treatment effects or group differences (Greenwald, Gonzalez, Harris, & Guthrie 1996; Kirk 1996; Olejnik & Algina 2000; Richardson 1996; Strube 1988; Thompson 1999a, 1999b). Similarly, in program evaluation where the goal is to assess the strength or impact of a program, effect sizes have been recommended provided they are placed in the context of similar or alternate programs (Posavac 1998).

Historically, two different approaches have been considered when deriving effect size measures. One approach evaluates the proportion of the variance in the response variable that is explained by the independent or grouping variable, such as η^2 and ω^2 (Richardson 1996). The other approach is based on the comparison of two population means, known as the standardized mean difference, δ. Specifically, δ reflects the difference between the two population means divided by σ', a measure of population heterogeneity or individual differences, $\delta = |\mu_1 - \mu_2| / \sigma'$. When two samples are drawn from two populations, δ can be estimated by using Cohen's *d*, Hedges' *d'* or adjusted *d*, and Glass's *d*.

The difference among these three estimators of δ is the nature of the standardizing term (i.e., σ', the denominator value). Cohen's *d* standardizes the mean difference, $m_1 - m_2$, by pooling or taking the square root of the average of the two sample variances, $|m_1 - m_2| / s_{pooled}$. The pooled standard deviation is computed as: $\sqrt{[(n_1-1)\, s^2_1 + (n_2-1)\, s^2_2] / (n_1 + n_2 - 2)}$.

Cohen's *d* however is a biased estimator for δ. Specifically, Hedges (1981) showed that the expected value of *d* is equal to δ/c(m), where

$$c(m) = \frac{\Gamma(m/2)}{[(m/2)] \cdot \Gamma[(m-1)/2]},$$

$m = (n_1 + n_2 - 2)$ and Γ is the gamma function. Although it approaches unity when m is large, it is appreciably smaller than unity when m is small, indicating that *d* overestimates δ (Richardson, 1996). Hedges found that the bias inherent in *d* could be easily removed by defining a new estimator $d' = d \cdot c(m)$. Not only is this an unbiased estimator, but it also has a smaller variance than *d*. Asymptotically, the sampling distribution of *d'* is closely related to the noncentral *t* distribution and is normally distributed with a mean equal to δ and a variance equal to $[N / n_1n_2) + \delta^2/2N)]$, where *N* is the total sample size (Richardson 1996).

A major limitation of using either Cohen's *d* or Hedges' *d'* is when population variances are unequal, in which case the meaningfulness of these effect sizes is limited. In many intervention studies (e.g., determining program impact using a nonequivalent groups design), the variances of the treatment and control groups are different because the program affects *both* the mean and the variability of the response variable scores. Consequently, depending on the severity of this difference, the underlying homogeneity of variance assumption for the *t*- and *F*-tests may be severely violated, leading to improper Type-I error rates (unless an adjustment is made) and no meaningful or *pure* measure of effect size. However, one approach to effect size estimation under heterogeneous variances that has been suggested by Glass (1977) is to estimate δ using the standard deviation of the control group to standardize the difference in group means.

<u>The Glass Estimator of δ for impact analysis</u>

In intervention studies where a treatment group is compared to a control group, Glass (1977) recommended standardizing the difference between the group means against the standard deviation of the control group. Glass argued that the control group standard deviation was the best choice for standardizing the mean difference between groups. His rationale was that in most cases the researcher is primarily interested in how the intervention or program group scores compare on average with the control group scores. Furthermore, because a program might affect the variance of the response variable as well as the mean, using the standard deviation of the untreated group produces the best standardization. This is especially recommended for many program evaluations when the group variances and sample sizes are unequal and pooling the standard deviations is not reasonable / meaningful.

Deriving the Glass index using the noncentralized t-distribution, Hedges (1981) found, under optimal conditions (i.e., normality, equal variances and equal n), the Glass index was severely biased under small sample sizes. Another study by Hedges and Olkin (1985, p. 79) concluded that the bias and the precision of Cohen's d were smaller than that of Glass's d, and if the assumption of homogeneity of variance is tenable, Cohen's d is more precise then the Glass index.

Moreover, because the noncentralized t-distribution was used to derive the Glass index, Hedges did not determine it's sampling properties under unequal variances, unequal n, or nonnormality. This is because it is difficult to theoretically derive the distribution properties of the Glass estimator under these conditions.

Statement of the problem

Given the need for effect size estimation when conducting an impact analysis in program evaluation, evaluators must be cognizant of the specific data conditions in which popular measures like Cohen's *d* or Hedges adjusted *d'* can be used. For example, the problem with popular estimators of standardized mean difference like Cohen's *d* is that it is rendered meaningless under unequal population variances. However, Glass's index can be used in this context, and previous research assessing the efficacy of the Glass index by Hedges (1981) indicated only some of its distributional properties under optimal conditions (e.g., small sample bias). Therefore, the purpose of this study was to (1) empirically describe the sampling characteristics of the Glass index under a more comprehensive set of data conditions, such as variance heterogeneity and nonnormality; and (2) make recommendation for use when conducting an impact analysis under specific data conditions like unequal group variances and *n* sizes, and / or nonnormality.

Method

The following four data conditions were manipulated to study the sampling characteristics of the Glass index: (1) population separation (effect size), (2) variance pattern, (3) total sample size with equal and unequal *n*, and (4) distribution shape. Variations in these data conditions are commonly found in social science literature and in most practical situations; furthermore, previous simulation studies have found these factors to be critical determinants of understanding the sampling properties of the *F* and *t* statistics (Harwell, Rubenstein, Hays, & Olds 1992).

Three levels of population separation, or δ, were considered. These δ values were set so that population 2 had a mean which was .2, .5, or .8 standard deviations greater then population 1

($\sigma = 1$). These were chosen based on relative values of d outlined by Cohen (1988, pp. 24-27) as "small," "medium," and "large" effect sizes; these benchmarks are also embraced by some social scientists in practice.

Three population variance ratios were considered: 1:1, 1:4, and 1:8. These variance patterns reflect a consistent variance of 1 for population 1 while the variance of population 2 is incremented 1, 4, and 8. Previous researchers have used similar variance patterns and the 1:4 ratio has been found to a point of severity where the violation of variance homogeneity assumption seriously affects Type I error rates and effect size measures when sample sizes are unequal (see Carol & Nordholm 1975).

Three levels of total sample size were manipulated. Total sample size was initially varied at three levels, $N = 40$, $N = 100$, and $N = 600$. Based on the Cohen (1988, p. 30) power charts, these sample sizes were sufficient to test the null hypothesis of no population mean difference with power equaling .80 at alpha equaling .05 in a directional test when the populations differ by $.80\sigma$, $.50\sigma$, and $.20\sigma$, respectively. However, using an iterative procedure, the largest N needed was 300 because N sizes greater than 300 revealed no change in the sample estimates of δ. Thus the final three sample sizes used in this study were 40, 100, and 300.

For each level of N, three patterns of group or n sizes were used. For $N = 40$, sample size ratios of 20:20, 30:10 (where the larger n was associated with the smaller variance), and 10:30 (where the smaller n was associated with the smaller variance) were used. For $N = 100$, n ratios were 50:50, 75:25, and 25:75, and for $N = 300$, n ratios were 150:150, 225:75, and 75:225. Moreover, considering equal and unequal n ratios in combination with unequal variance patterns was viewed important to adequately describe the sampling characteristics of Glass's index. Previous researchers (e.g., Glass, Peckham, & Sanders 1972; Lix & Kesselman 1998) have considered this joint condition to assess the robustness of common test statistics such as the *t* and *F* statistics.

Finally, two levels of population shape were considered: a normal and a skewed-leptokurtic or peaked (1.75, 3.75) distribution. The distribution shapes were identical for the two populations being compared. A third level of nonnormality (skewed-mesokurtic .75, 0) was initially considered but based on preliminary results this level was dropped as unnecessary. Only two distribution shapes were considered sufficient to obtain a good picture of the sampling properties of the Glass index.

Data Generation

Data were generated to meet the above conditions using SAS IML (SAS Institute 1990). Within each of the two populations, observations Y_{ij} ($i = 1 \ldots n_j$ and $j = 1,2$) were generated using Equation 5:

$$Y_{ij} = \delta_j + X_{ij}\sqrt{\sigma_j^2}\ , \qquad (5)$$

where σ_j^2 is the population variance for group j, with, σ_1^2=1 and σ_2^2= 1, 4, or 8. X_{ij} is a standard random variable transformed to the desired target distribution shape using the Fleishman (1978) power transformation: $X_{ij} = a_{ij} + bZ_{ij} + cZ^2_{ij} + dZ^3_{ij}$, where Z_{ij} are independent standard normal variables generated using the SAS-RANNOR function. For normal distributions, a = 0, b = 1, c = 0, and d = 0. For the skewed-leptokurtic distributions, the constants were set to: a = - .399, b = .930, c = .399, and d = -.036. Finally, δ_j is a measure of group separation with $\delta_1 = 0$ and $\delta_2 = .2$, .5, or .8.

Data Analyses and Evaluation

The four data conditions were manipulated for the present investigation in a completely crossed design. A total of 162 conditions were investigated: 3 levels of population separation, 3 variance ratios, 9 sample size levels (i.e., including equal and two unequal n ratios under each of three total sample sizes), and 2 distribution shapes. For each of these conditions 5,000 replications were computed. In order to describe the distributional properties of the Glass index, means, standard deviations, and three quantiles (i.e., 25^{th}, 50^{th}, and 75^{th} percentiles, or Q_1, Q_2, and Q_3) were tabulated for each condition.

The accuracy or the degree of bias of the Glass index was computed as the difference between the sample mean of d and δ over 5,000 replications. Differences greater than +/- .30 (δ) (or, in other words, differences in excess of 30%) indicated severe bias. This 30% criterion was based on Bradley (1978) who recommended that a procedure might be considered robust to the violation of an assumption if the Type I error rate was within +/- $.50\alpha$. Bradley considered +/- $.50\alpha$ liberal and $.10\alpha$ conservative. Adopting Bradley's approach, .50 (δ) was considered to be too liberal and .10 (δ) to be too conservative; therefore it was concluded that .30 (δ) was a reasonable criterion for bias. Finally, precision was computed as the standard deviation of the sampling distribution of d under each condition. Box plots were also used to evaluate the precision of the Glass estimator. Specifically, the inclusion of the median of d at one level of population separation within the hinges (25^{th} and 75^{th} percentiles) of adjacent levels of population separation was viewed as unacceptable.

Results

Efficacy of Glass's Effect Size Under Optimal Conditions

The degree of accuracy (bias) and precision (variability) of the Glass index under the optimal conditions of variance homogeneity and data normality are shown in Table 1. Tabulated values under the column "*A*" reflect the accuracy of the estimate and are computed by taking the difference between the sample mean of d and the respective population value (e.g., d - δ).

Tabulated values under the column "*P*" reflect the precision or standard deviation of the sample estimate.

Under optimal conditions, the Glass index was severely biased (upward) when both population separation was small ($\delta = .20$) and when sample sizes were small ($N = 40$ and 100). For example, under equal *n*, when $\delta = .20$ (and the cutoff was +/- .06), the degree of bias was .110 when $N = 40$ and .073 when $N = 100$. Similarly, the variability of the Glass index was slightly larger when sample sizes were small. For example, when $N = 40$, the precision ranged from .239 to .367; however when $N = 300$, the precision improved and ranged from .098 to .107. Using a three-point summary, Figure 1 graphically depicts the effect of sample size on the precision of the Glass index under optimal conditions. Figure 1 shows when sample sizes were large ($N = 300$), the medians of larger effect size distributions were not captured between the hinges of smaller effect size distributions, thus indicating reasonably good precision.

When group or *n* sizes were unequal, the amount of bias and variability was also more severe under smaller population separations and sample sizes ($\delta = .20$ and $N = 40$). Unequal *n* did not present a problem when population separations and sample sizes were large. Moreover, under the optimal conditions of variance homogeneity and normality, the results indicated that the Glass index was a good estimator of δ, except when both population separations and sample sizes were jointly small (i.e., $\delta = .20$, $N = 40$ and 100) regardless of the *n* ratio.

<u>Under Variance Heterogeneity</u>

When population variances were heterogeneous and distributions were normal, the Glass index appeared to clearly overestimate δ when both population separations were smaller ($\delta = .20$ and .50) and sample sizes were smaller ($N = 40$ and 100), regardless of the *n* ratio. In fact, the bias and sampling variability was larger when the smaller *n* was associated with the larger variance. The effect of moderate (1:4) and extreme (1:8) variance heterogeneity on the Glass index are presented in Tables 2 and 3, respectively.

To elaborate, when variance heterogeneity was moderate (1:4), the Glass index was severely biased upward when population separations were small ($\delta = .20$) except when sample sizes were large ($N = 300$). For example, according to Table 2, $\delta = .20$ and $N = 40$ and 100, the bias ranged from .107 to .364 and the precision ranged from .354 to .514. Furthermore, the bias was larger when *n* sizes were unequal (ranging from .137 to .364), particularly when the smaller *n* was associated with the larger variance. However, when $N = 300$, the bias was not severe, ranging from trivial to .029 across all levels of δ and *n* ratios; similarly, the precision improved, ranging from .150 to .247. Figure 2 shows the improvement in precision across levels of δ when $N = 300$ (assuming equal *n*) under moderate variance heterogeneity (1:4).

When the variance heterogeneity was extreme (1:8), the Glass index was also biased upward, only this time including when $\delta = .50$ and $N = 40$. For example, according to Table 3, when $\delta = .20$ and $N = 40$ and 100, the bias ranged from .153 to .576; precision ranged from .267 to .592. Similarly, When $\delta = .50$ and $N = 40$, the bias ranged from .200 to .351; precision ranged from .539 to .656. However, as found under the moderate variance heterogeneity (1:4) condition, when $N = 300$, the amount of bias was not severe (i.e., trivial to .045) across all levels of δ and *n* ratio; similarly, the precision improved and ranged from .152 to .304. Furthermore, the improvement in precision of the Glass index when $N = 300$ under extreme variance heterogeneity (1:8) is shown in Figure 3 (assuming equal *n*).

Under Nonnormal Population Distributions

Table 4 shows the effect of extreme nonnormality when population variances were equal. Under nonnormality (1.75, 3.75), the Glass index was severely biased under both small population separations ($\delta = .20$) and small sample sizes ($N = 40$). The Glass index was also more precise under large sample sizes ($N = 300$). These results were consistent with the results found under data normality (see Table 1). For example, according to Table 1, when $N = 300$ the amount of bias was trivial (< .01) and the degree of precision ranged from .098 to .131 across all δ. Similarly under data nonnormality, Table 4 shows when $N = 300$, the amount of bias ranged from trivial (< .02) to .036 and precision ranged from .124 to .229.

As additional evidence, Figure 4 shows when population variances were equal, the sampling estimates were more precise under large sample sizes ($N = 300$) when data were nonnormal. This is similar to the degree of precision found under large sample sizes when data were normal, as demonstrated in Figure 1. That is, under large sample sizes, the medians of larger effect size distributions were not captured between the hinges of smaller effect size distributions. Finally, the severity of bias and variability was most apparent when *n* sizes were unequal, but similar to the optimal conditions, the estimates were not severely biased under large sample sizes ($N = 300$).

The influence of moderate (1:4) and extreme (1:8) variance heterogeneity on the Glass index under extreme nonnormality is shown in Tables 5 and 6, respectively. These tables show that regardless of the severity variance heterogeneity, the Glass index was biased upward when (a) population separations were small ($\delta = .20$) and sample sizes were smaller ($N = 40$ and 100) across all *n* ratios, and (b) population separations were moderate ($\delta = .50$) and sample sizes were small ($N = 40$) across all *n* ratios. The magnitude of the bias was acceptable when sample sizes were large ($N = 300$). Furthermore, the precision of the Glass index under large sample sizes ($N = 300$) when data were nonnormal and when variances were heterogeneous can be seen in

Figures 5 and 6, respectively (assuming equal n). Again, this is similar to what was found under data normality, as demonstrated in Figures 2 and 3, respectively.

Discussion

Based on the accuracy and precision of the Glass index, the following conclusions were drawn. When population variances are equal, the Glass effect size is recommended regardless of the distribution shape and n ratio provided sample sizes and population effect sizes are not jointly small ($\delta = .20$ and $N = 40$). In other words, the Glass index performs well when the total sample size is sufficient given the size of the population effect size (i.e., using Cohen's power tables, adequate N to maintain power at .80 in order to detect the effect at $\alpha = .05$). Similarly, when variances are heterogeneous, the Glass index can also be recommended regardless of the distribution shape and n ratio provided the sample size is sufficient given the size of the population effect size. Finally, for greater precision, the Glass index performs best under large sample sizes ($N = 300$ or greater) across all conditions.

The sampling properties of the Glass index were consistent with Hedges (1981) who found, under optimal conditions (i.e., normality, equal variances, and equal n), the Glass index was severely biased under small sample sizes. Results were also consistent with Hedges and Olkin (1985, p. 77) who noted that, in general, the bias and the precision of the Glass index were large across variance patterns (particularly larger than that of the Cohen d). Furthermore, the condition of extreme nonnormality (1.75, 3.75) had only a marginal influence on the accuracy and precision of the Glass index. This makes sense because t is robust to violations of the normality assumption.

In summary, if an evaluator conducts a power analysis when planing an impact analysis, he or she will find the Glass effect size to be a good measure of treatment impact. In other words, if an evaluator anticipates a particular effect size believed to be manifested by the treatment population and selects the proper sample size in order to maintain sufficient power (at .80) for detecting that effect at a particular significance level (say .05), then the Glass index may provide an accurate measure of effect size.

Implications and limitations

First, the present study only investigated the Glass estimator of δ and excluded other estimators like the Cohen *d* or Hedge's adjusted *d'*. Including these indexes in order to make comparisons with the Glass index would have been interesting. As pointed out, the Hedges adjusted *d* has been shown to be unbiased and therefore recommended under small sample sizes. Similarly, the Cohen *d* has been shown to be more precise than the Glass index under variance homogeneity (Hedges 1981).

Second, because only a limited number of data conditions were selected and manipulated, the findings can only be generalized to the specific data conditions and levels used in the present study. Specifically, only three levels of effect size were considered, and only a sample of distribution shapes, sample sizes, and variance patterns were studied. While the specific levels under each condition did provide sufficient information to adequately describe the sampling properties of Glass's effect size, additional levels may have provided a more thorough picture.

Conclusion

In general, program evaluators should recognize the need for estimating an effect size when conducting an impact analysis because traditional hypothesis testing has limitations (e.g., does not address practical significance of a program's impact). This study addressed the problem of effect size estimation, particularly using the Glass index for assessing program under a more comprehensive set of data conditions. Traditional measures of effect size measures have been limited to situations where population variances are equal, except in intervention studies wherein the Glass index has been suggested. Furthermore, this study is important because it provides greater understanding of the properties and limitations of the Glass estimator of the standardized mean difference used to estimate effect size for intervention studies / assessing program impact.

References

Carroll, R. M., and L. A. Nordholm. 1975. A sampling characteristic of Kelly's ε^2 and Hays'ω^2. *Educational and Psychological Measurement* 35:541-554.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. New York: Academic Press.

Falk, R., and C. W. Greenbaum. 1995. Significance tests die hard. *Theory & Psychology* 5: 75-98.

Glass, G. V. 1977. Integrating findings: The meta-analysis of research. *Review of Research in Education* 5:351-379.

Glass, G. V., P. D. Peckham, and J. R. Sanders. 1972. Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research* 42:237-288.

Greenwald, A. G., R. Gonzalez, R. L. Harris, and D. Guthrie. 1996. Effect sizes and p values: What should be reported and what should be replicated. *Psychophysiology* 33:175-183.

Harwell, M. R., E. N. Rubenstein, W. S. Hays, and C. C. Olds. 1992. Summarizing Monte Carlo research in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics* 17:315-339.

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

Hedges, L. V., and I. Olkin. 1985. *Statistical analysis for meta-analysis.* San Diego, CA: Academic Press.

Huberty, C. J, and C. J. Pike. 1999. On some history regarding statistical testing. In *Advances in social science methodology*, edited by B. Thompson, 1-22. Greenwich, CT: JAI Press.

Kirk, R. E. 1996. Practical significance: A concept whose time has come. *Educational and Psychological Measurement* 56:746-759.

Lix, L. M., and H. J. Keselman. 1998. To trim or not to trim: Test of location equality under Heteroscedasticity and nonnormality. *Educational and Psychological Measurement* 58:409-429.

Olejnik, S., and J. Algina. 2000. Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology* 25:241-286.

Posavac, E. J. 1998. Toward more informative uses of statistics: Alternatives for program evaluators. *Evaluation and Program Planning* 21:243-254.

Richardson, J. T. E. 1996. Measures of effect size. *Behavior Research Methods, Instruments, & Computers* 28:12-22.

SAS Institute Inc. 1990. *SAS / IML software: Usage and reference, version 6*. 1st ed. Cary, NC: Author.

Schmidt, F. L. 1996. Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods* 1:115-129.

Strube, M. J. 1988. Some comments on the use of magnitude-of-effect estimates. *Journal of Counseling Psychology* 35:342-345.

Thompson, B. 1999a. Statistical significant tests, effect size reporting, and the vain pursuit of pseudo-objectivity. *Theory & Psychology* 9:191-196.

-------. 1999b. Why "encouraging" effect size reporting is not working: The etiology of researcher resistance to changing practices. *Journal of Psychology* 133:133-140.

Table 1: Accuracy and Precision of the Glass Index under Equal Variances (1:1) and Normal Population Distributions (0,0)

		Population Separation					
N size	*n* ratio	$\delta = .20$		$\delta = .50$		$\delta = .80$	
		A	*P*	*A*	*P*	*A*	*P*
40	20:20	**.110**	.239	.033	.318	.039	.367
	30:10	**.146**	.262	.046	.345	.021	.381
	10:30	**.165**	.298	.080	.401	.083	.474
100	50:50	**.073**	.163	.010	.211	.014	.220
	75:25	.053	.180	.005	.234	.013	.246
	25:75	.059	.189	.018	.249	.022	.269
300	150:150	.003	.098	.000	.103	.004	.107
	225:75	.005	.109	.001	.118	.001	.120
	75:225	.004	.112	.006	.122	.006	.131

Note. *A* = accuracy, which is the difference between the sample mean of *d* and δ over 5,000 replications. *P* = precision, which is the standard deviation of *d*. Values in bold identify those conditions where the bias exceeded our criterion of .3(δ).

Table 2: Accuracy and Precision of the Glass Index under Moderate Variance Heterogeneity (1:4) and Normal Population Distributions (0,0)

		Population Separation					
N size	*n* ratio	δ = .20		δ = .50		δ = .80	
		A	*P*	*A*	*P*	*A*	*P*
40	20:20	**.248**	.354	.107	.433	.034	.502
	30:10	**.364**	.442	.144	.514	.100	.596
	10:30	**.258**	.383	.137	.469	.102	.563
100	50:50	**.107**	.228	.024	.306	.012	.335
	75:25	**.176**	.289	.053	.362	.018	.409
	25:75	**.104**	.232	.032	.306	.026	.339
300	150:150	.029	.151	.003	.190	.002	.191
	225:75	.052	.184	.006	.238	.007	.247
	75:225	.029	.154	.006	.186	.007	.150

Note. *A* = accuracy, which is the difference between the sample mean of *d* and δ over 5,000 replications. *P* = precision, which is the standard deviation of *d*. Values in bold identify those conditions where the bias exceeded our criterion of .3(δ).

Table 3: Accuracy and Precision of the Glass Index under Extreme Variance Heterogeneity (1:8) and Normal Population Distributions (0,0)

		Population Separation					
N size	*n* ratio	δ = .20		δ = .50		δ = .80	
		A	*P*	*A*	*P*	*A*	*P*
40	20:20	**.394**	.467	**.210**	.539	.101	.608
	30:10	**.576**	.592	**.351**	.656	.224	.756
	10:30	**.352**	.464	**.200**	.563	.142	.685
100	50:50	**.200**	.289	.054	.374	.039	.418
	75:25	**.288**	.370	.131	.453	.058	.523
	25:75	**.153**	.267	.037	.354	.036	.422
300	150:150	.042	.166	.005	.210	.005	.215
	225:75	.045	.180	.025	.304	.001	.182
	75:225	.031	.152	.013	.220	.001	.160

Note. *A* = accuracy, which is the difference between the sample mean of *d* and δ over 5,000 replications. *P* = precision, which is the standard deviation of *d*. Values in bold identify those conditions where the bias exceeded our criterion of .3(δ).

Table 4: Accuracy and Precision of the Glass Index under Equal Variances (1:1) and Nonnormal Population Distributions (1.75, 3.75)

		Population Separation					
N size	*n* ratio	δ = .20		δ = .50		δ = .80	
		A	*P*	*A*	*P*	*A*	*P*
40	20:20	**.156**	.334	.128	.498	.148	.580
	30:10	**.172**	.349	.081	.444	.091	.501
	10:30	**.285**	.563	**.268**	.737	.210	.950
100	50:50	.058	.206	.040	.275	.044	.319
	75:25	.057	.206	.031	.277	.034	.295
	25:75	.059	.264	.091	.380	.112	.454
300	150:150	.011	.124	.013	.151	.019	.175
	225:75	.008	.128	.007	.154	.011	.172
	75:225	.028	.151	.026	.194	.036	.229

Note. *A* = accuracy, which is the difference between the sample mean of *d* and δ over 5,000 replications. *P* = precision, which is the standard deviation of *d*. Values in bold identify those conditions where the bias exceeded our criterion of .3(δ).

Table 5: Accuracy and Precision of the Glass Index under Moderate Variance Heterogeneity (1:4) and Nonnormal Population Distributions (1.75, 3.75)

		Population Separation					
N size	*n* ratio	δ = .20		δ = .50		δ = .80	
		A	*P*	*A*	*P*	*A*	*P*
40	20:20	**.277**	.476	**.175**	.591	.158	.687
	30:10	**.380**	.530	**.187**	.620	.112	.183
	10:30	**.367**	.665	**.325**	.880	.213	1.043
100	50:50	**.113**	.268	.057	.356	.057	.415
	75:25	**.178**	.313	.061	.408	.046	.459
	25:75	**.140**	.304	.090	.423	.118	.510
300	150:150	.027	.165	.014	.207	.020	.190
	225:75	.057	.196	.012	.253	.018	.188
	75:225	.038	.176	.027	.227	.029	.261

Note. *A* = accuracy, which is the difference between the sample mean of *d* and δ over 5,000 replications. *P* = precision, which is the standard deviation of *d*. Values in bold identify those conditions where the bias exceeded our criterion of .3(δ).

Table 6: Accuracy and Precision of the Glass Index under Extreme Variance Heterogeneity (1:8) and Nonnormal Population Distributions (1.75, 3.75)

		Population Separations					
N size	*n* ratio	δ = .20		δ = .50		δ = .80	
		A	*P*	*A*	*P*	*A*	*P*
40	20:20	**.437**	.614	**.274**	.721	.193	.823
	30:10	**.585**	.670	**.364**	.781	.211	.930
	10:30	**.463**	.757	**.363**	.953	.218	1.124
100	50:50	**.199**	.335	.082	.437	.060	.502
	75:25	**.286**	.396	.124	.493	.070	.584
	25:75	**.195**	.364	.120	.485	.120	.561
300	150:150	.059	.200	.015	.260	.010	.281
	225:75	**.115**	.249	.021	.324	.040	.300
	75:225	.054	.196	.027	.257	.034	.294

Note. *A* = accuracy, which is the difference between the sample mean and δ.
P = precision, which is the standard deviation of the sample estimates.

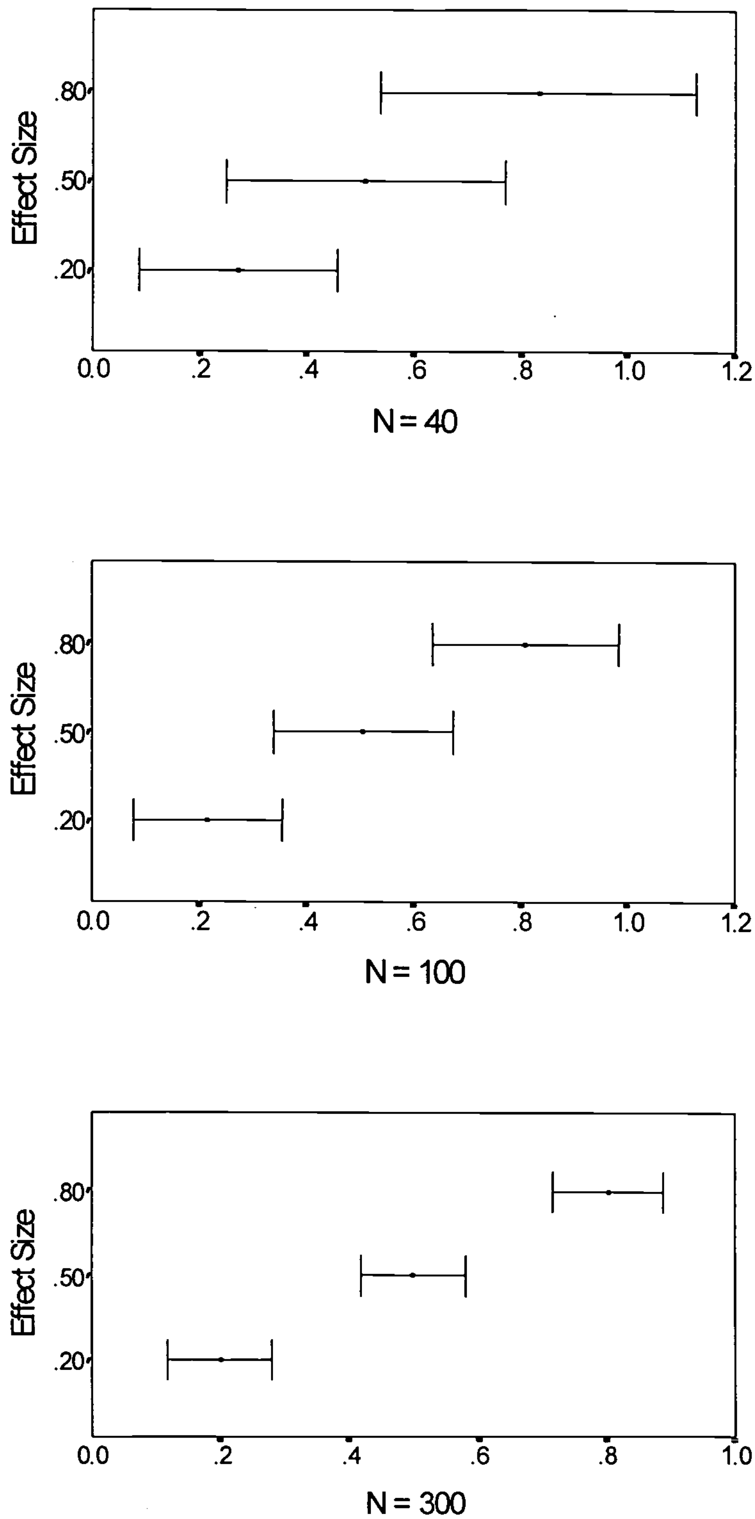

Figure 1: 3-Point Summary of the Glass Index under Optimal Conditions (Equal *n*).

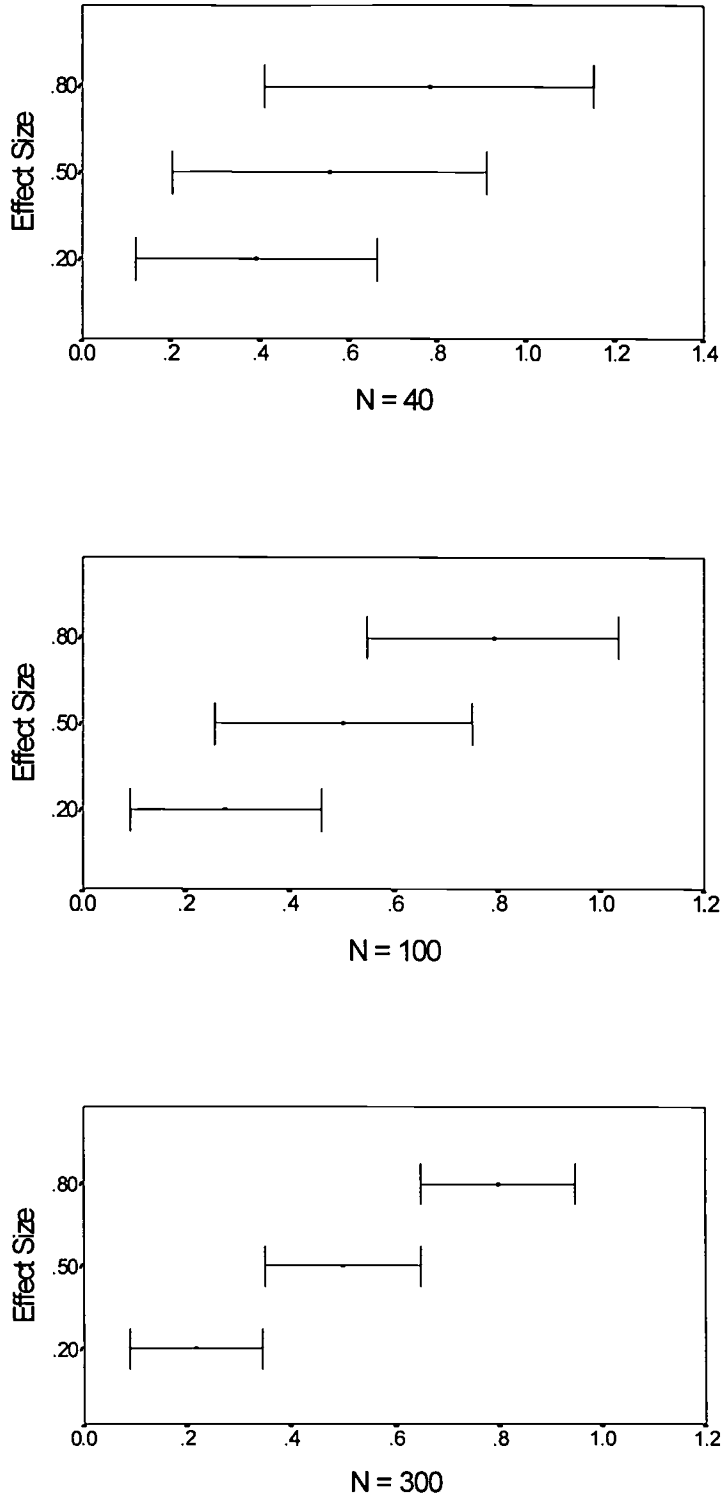

Figure 2: 3-Point Summary of the Glass Index Under Moderate Variance Heterogeneity (1:4) and Normality (0,0) and Equal *n*.

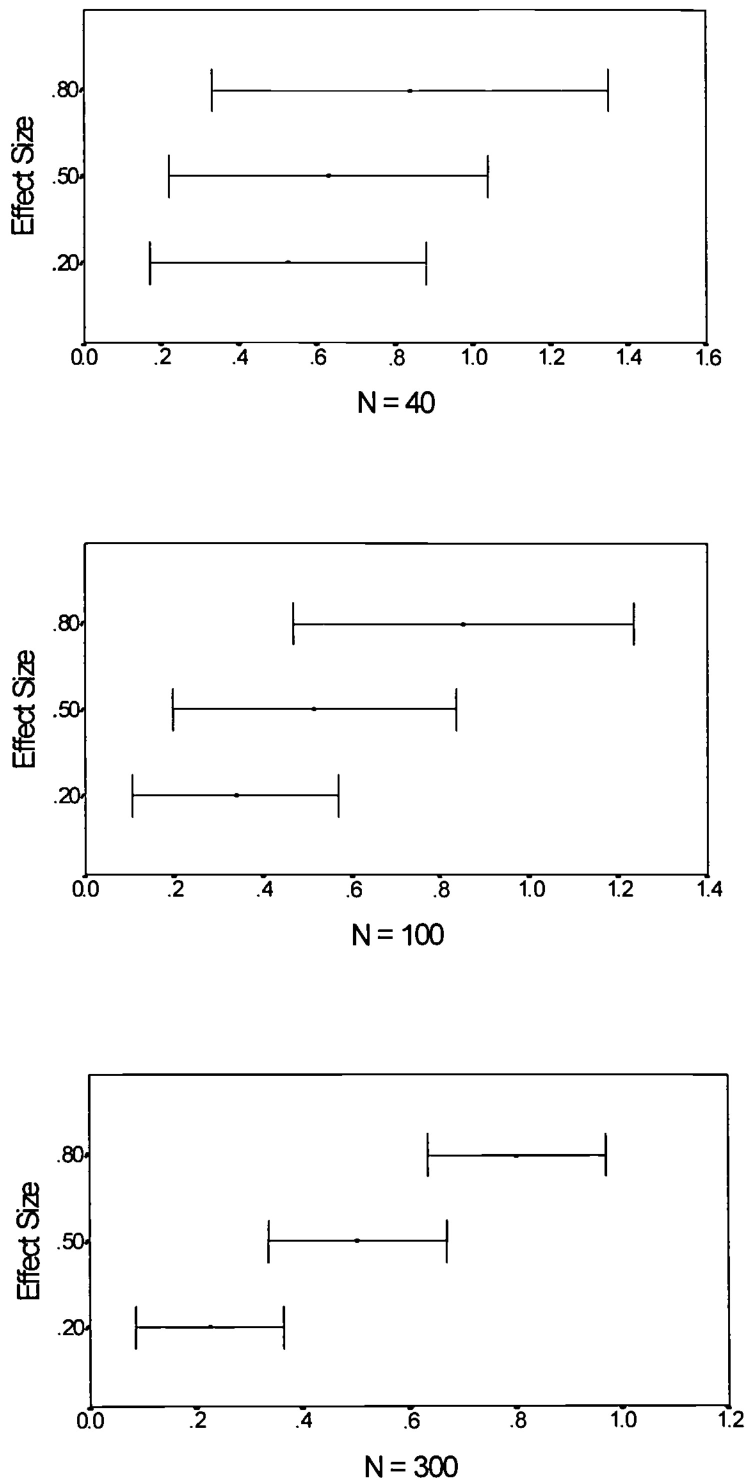

Figure 3: 3-Point Summary of the Glass Index Under Extreme Variance Heterogeneity (1:8) and Normality (0,0) and Equal *n*.

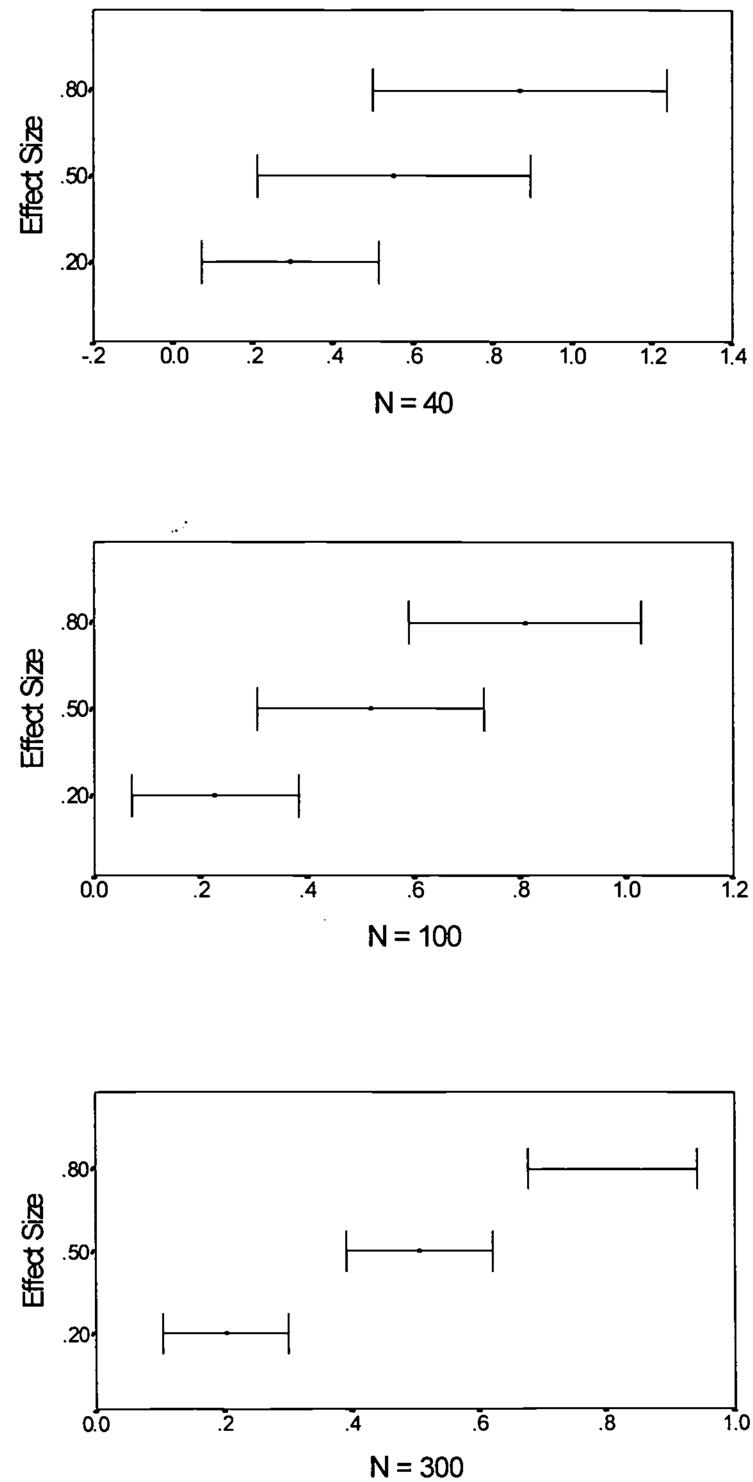

Figure 4: 3-Point Summary of the Glass Index under Equal Variances (1:1) and Nonnormality (1.75, 3.75) and Equal *n*.

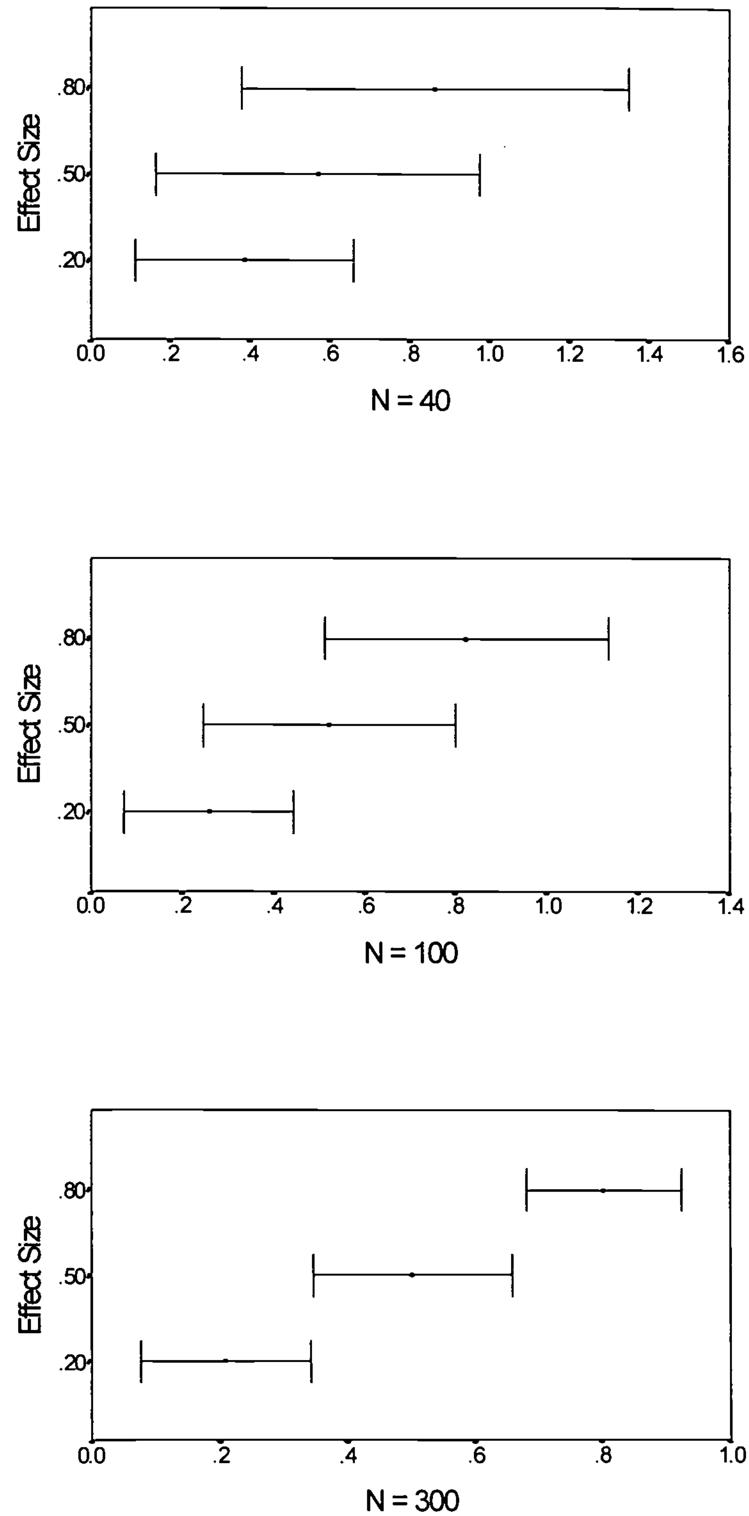

Figure 5: 3-Point Summary of the Glass Index under Moderate Variance Heterogeneity (1:4) and Nonnormality (1.75, 3.75) and Equal *n*.

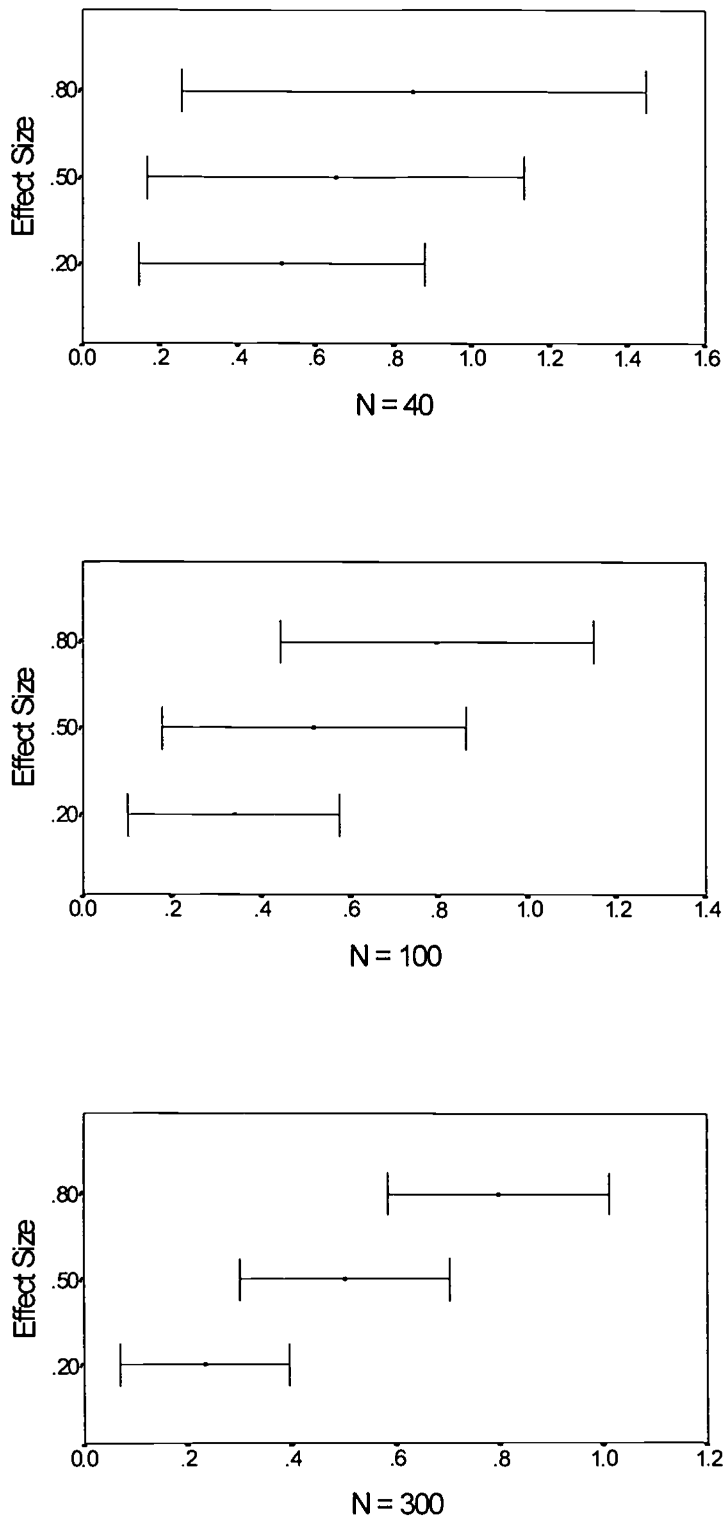

Figure 6: 3-Point Summary of the Glass Index under Extreme Variance Heterogeneity (1:8) and Nonnormality (1.75, 3.75) and Equal *n*.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM032495

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Revisiting the Efficacy of Glass's estimator of Effect Size For Program Impact Analysis

Author(s): Brian Hess and Stephen Olejnik

Corporate Source: University of Georgia

Publication Date: 4/14/01

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☒	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Brian Hess	Printed Name/Position/Title: Brian Hess, Project Director	
Organization/Address: University of Georgia, Athens GA	Telephone: 813-979-7659	FAX: 813-968-4684
	E-Mail Address: Bhess@coe.uga.edu	Date: 3/27/01

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 9/97)