Lee A. Becker

<http://web.uccs.edu/lbecker/Psy590/es.htm>

Effect Size (ES)

© 2000 Lee A. Becker

I. Overview

Effect size (ES) is a name given to a family of indices that measure the magnitude of a treatment effect. Unlike significance tests, these indices are independent of sample size. ES measures are the common currency of meta-analysis studies that summarize the findings from a specific area of research. See, for example, the influential meta-analysis of psychological, educational, and behavioral treatments by Lipsey and Wilson (1993).

There is a wide array of formulas used to measure ES. For the occasional reader of meta-analysis studies, like myself, this diversity can be confusing. One of my objectives in putting together this set of lecture notes was to organize and summarize the various measures of ES.

In general, ES can be measured in two ways:

a) as the standardized difference between two means, or

b) as the correlation between the independent variable classification and the individual scores on the dependent variable. This correlation is called the "effect size correlation" (Rosnow & Rosenthal, 1996).

These notes begin with the presentation of the basic ES measures for studies with two independent groups. The issues involved when assessing ES for two dependent groups are then described.

II. Effect Size Measures for Two Independent Groups

1. Standardized difference between two groups.

Cohen's d

$d = M_1 - M_2 / \sigma$ where $\sigma = \sqrt{[\sum(X - M)^2 / N]}$ where X is the raw score, M is the mean, and N is the number of cases.	Cohen (1988) defined d as the difference between the means, $M_1 - M_2$, divided by standard deviation, σ, of either group. Cohen argued that the standard deviation of either group could be used when the variances of the two groups are homogeneous. In meta-analysis the two groups are considered to be the experimental and control groups. By convention the subtraction, $M_1 - M_2$, is done so that the difference is positive if it is in the direction of *improvement* or in the predicted direction and negative if in the direction of *deterioration* or opposite to the predicted direction. d is a descriptive measure.
$d = M_1 - M_2 / \sigma_{pooled}$ $\sigma_{pooled} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$	In practice, the pooled standard deviation, σ_{pooled}, is commonly used (Rosnow and Rosenthal, 1996). The pooled standard deviation is found as the root mean square of the two standard deviations (Cohen, 1988, p. 44). That is, the pooled standard deviation is the square root of the average of the squared standard deviations. When the two standard deviations are similar the root mean square will be not differ much from the simple average of the two variances.
$d = 2t / \sqrt{(df)}$ or $d = t(n_1 + n_2) / [\sqrt{(df)}\sqrt{(n_1 n_2)}]$	d can also be computed from the value of the t test of the differences between the two groups (Rosenthal and Rosnow, 1991). . In the equation to the left "df" is the degrees of freedom for the t test. The "n's" are the number of cases for each group. The formula without the n's should be used when the n's are equal. The formula with

	separate n's should be used when the n's are not equal.
$d = 2r / \sqrt{(1 - r^2)}$	d can be computed from r, the ES correlation.
$d = g\sqrt{(N/df)}$	d can be computed from Hedges's g.

The interpretation of Cohen's *d*

Cohen (1988) hesitantly defined effect sizes as "small, $d = .2$," "medium, $d = .5$," and "large, $d = .8$", stating that "there is a certain risk in inherent in offering conventional operational definitions for those terms for use in power analysis in as diverse a field of inquiry as behavioral science" (p. 25).

Cohen's Standard	Effect Size	Percentile Standing	Percent of Nonoverlap
	2.0	97.7	81.1%
	1.9	97.1	79.4%
	1.8	96.4	77.4%
	1.7	95.5	75.4%
	1.6	94.5	73.1%
	1.5	93.3	70.7%
	1.4	91.9	68.1%
	1.3	90	65.3%
	1.2	88	62.2%
	1.1	86	58.9%
	1.0	84	55.4%
	0.9	82	51.6%
LARGE	0.8	79	47.4%
	0.7	76	43.0%
	0.6	73	38.2%
MEDIUM	0.5	69	33.0%
	0.4	66	27.4%
	0.3	62	21.3%
SMALL	0.2	58	14.7%
	0.1	54	7.7%
	0.0	50	0%

Effect sizes can also be thought of as the average percentile standing of the average treated (or experimental) participant relative to the average untreated (or control) participant. An ES of 0.0 indicates that the mean of the treated group is at the 50th percentile of the untreated group. An ES of 0.8 indicates that the mean of the treated group is at the 79th percentile of the untreated group. An effect size of 1.7 indicates that the mean of the treated group is at the 95.5 percentile of the untreated group.

Effect sizes can also be interpreted in terms of the percent of nonoverlap of the treated group's scores with those of the untreated group, see Cohen (1988, pp. 21-23) for descriptions of additional measures of nonoverlap.. An ES of 0.0 indicates that the distribution of scores for the treated group overlaps completely with the distribution of scores for the untreated group, there is 0% of nonoverlap. An ES of 0.8 indicates a nonoverlap of 47.4% in the two distributions. An ES of 1.7

	indicates a nonoverlap of 75.4% in the two distributions.

Hedges's *g*

$g = M_1 - M_2 / S_{pooled}$ where $S = \sqrt{[\sum(X - M)^2 / N-1]}$ and $S_{pooled} = \sqrt{MSwithin}$	Hedges's *g* is an inferential measure. It is normally computed by using the square root of the Mean Square Error from the analysis of variance testing for differences between the two groups. Hedges's *g* is named for Gene V. Glass, one of the pioneers of meta-analysis.
$g = t\sqrt{(n_1 + n_2)} / \sqrt{(n_1 n_2)}$ or $g = 2t / \sqrt{N}$	Hedges's *g* can be computed from the value of the *t* test of the differences between the two groups (Rosenthal and Rosnow, 1991). The formula with separate n's should be used when the n's are not equal. The formula with the overall number of cases, N, should be used when the n's are equal.
$\sigma_{pooled} = S_{pooled} \sqrt{(df / N)}$ were df = the degrees of freedom for the MSerror, and N = the total number of cases.	The pooled standard deviation, σ_{pooled}, can be computed from the unbiased estimator of the pooled population value of the standard deviation, S_{pooled}, and vice versa, using the formula on the left (Rosnow and Rosenthal, 1996, p. 334).
$g = d / \sqrt{(N / df)}$	Hedges's *g* can be computed from Cohen's *d*.
$g = [r / \sqrt{(1 - r^2)}] / \sqrt{[df(n_1 + n_2) / (n_1 n_2)]}$	Hedges's *g* can be computed from r, the ES correlation.

Glass's delta

$\Delta = M_1 - M_2 / \sigma_{control}$	Glass's delta is defined as the mean difference between the experimental and control group divided by the standard deviation of the control group.

2. Correlation measures of effect size

The ES correlation, $r_{Y\lambda}$

$r_{Y\lambda} = r_{dv,iv}$	The effect size correlation can be computed directly as the point-biserial correlation between the dichotomous independent variable and the continuous dependent variable.
CORR = dv with iv	The point-biserial is a special case of the Pearson product-moment correlation that is used when one of the variables is dichotomous. As Nunnally (1978) points out, the point-biserial is a shorthand method for computing a Pearson product-moment correlation. The value of the point-biserial is the same as that obtained from the product-moment correlation. You can use the CORR procedure in SPSS to compute the ES correlation.
$r_{Y\lambda} = \Phi = \sqrt{(X^2(1) / N)}$	The ES correlation can be computed from a single degree of freedom Chi Square value by taking the square root of the Chi Square value divided by the number of cases, N. This value is also known as Phi.
$r_{Y\lambda} = \sqrt{[t^2 / (t^2 + df)]}$	The ES correlation can be computed from the t-test value.
$r_{Y\lambda} = \sqrt{[F(1,\_) / (F(1,\_) + df\ error)]}$	The ES correlation can be computed from a single degree of freedom F test value (e.g., a oneway analysis of variance with two groups).
$r_{Y\lambda} = d / \sqrt{(d^2 + 4)}$	The ES correlation can be computed from Cohen's d.
$r_{Y\lambda} = \sqrt{\{(g^2 n_1 n_2) / [g^2 n_1 n_2 + (n_1 + n_2)df]\}}$	The ES correlation can be computed from Hedges's g.

The relationship between d, r, and r²

Cohen's Standard	d	r	r²	
				As noted in the definition sections above, d and be converted to r and vice versa.
	2.0	.707	.500	
	1.9	.689	.474	For example, the d value of .8
	1.8	.669	.448	corresponds to an r value of .371.

				The square of the r-value is the percentage of variance in the dependent variable that is accounted for by membership in the independent variable groups. For a *d* value of .8, the amount of variance in the dependent variable by membership in the treatment and control groups is 13.8%. In meta-analysis studies *r*s are typically presented rather than *r²*.
	1.7	.648	.419	
	1.6	.625	.390	
	1.5	.600	.360	
	1.4	.573	.329	
	1.3	.545	.297	
	1.2	.514	.265	
	1.1	.482	.232	
	1.0	.447	.200	
	0.9	.410	.168	
LARGE	0.8	.371	.138	
	0.7	.330	.109	
	0.6	.287	.083	
MEDIUM	0.5	.243	.059	
	0.4	.196	.038	
	0.3	.148	.022	
SMALL	0.2	.100	.010	
	0.1	.050	.002	
	0.0	.000	.000	

3. Computational Examples

The following data come from Wilson, Becker, and Tinker (1995). In that study participants were randomly assigned to either EMDR treatment or delayed EMDR treatment. Treatment group assignment is called TREATGRP in the analysis below. The dependent measure is the Global Severity Index (GSI) of the Symptom Check List-90R. This index is called GLOBAL4 in the analysis below. The analysis looks at the the GSI scores immediately post treatment for those assigned to the EMDR treatment group and at the second pretreatment testing for those assigned to the delayed treatment condition. The output from the SPSS MANOVA and CORR(elation) procedures are shown below.

```
 Cell Means and Standard Deviations
 Variable .. GLOBAL4          GLOBAL INDEX:SLC-90R POST-TEST
     FACTOR          CODE                  Mean  Std. Dev.        N   95 percent
Conf. Interval

  TREATGRP        TREATMEN                 .589      .645        40       .383
.795
  TREATGRP        DELAYED                 1.004      .628        40       .803
1.205
 For entire sample                         .797      .666        80       .648
.945

* * * * * * * * * * * * * A n a l y s i s   o f   V a r i a n c e -- Design   1 * * *
* * * * * * * * *

 Tests of Significance for GLOBAL4 using UNIQUE sums of squares
```

```
Source of Variation          SS      DF      MS       F   Sig of F

WITHIN CELLS               31.60      78     .41
TREATGRP                    3.44       1    3.44    8.49     .005

(Model)                     3.44       1    3.44    8.49     .005
(Total)                    35.04      79     .44


                   - -   Correlation Coefficients   - -

            GLOBAL4

TREATGRP      .3134
           (   80)
           P= .005
```

Look back over the formulas for computing the various ES estimates. This SPSS output has the following relevant information: cell means, standard deviations, and ns, the overall N, and MSwithin. Let's use that information to compute ES estimates.

$d = M_1 - M_2 / \sqrt{[(\sigma_1^2 + \sigma_2^2)/2]}$ $= 1.004 - 0.589 / \sqrt{[(0.628^2 + 0.645^2)/2]}$ $= 0.415 / \sqrt{[(0.3944 + 0.4160)/2]}$ $= 0.415 / \sqrt{(0.8144/2)}$ $= 0.415 / \sqrt{0.4052}$ $= 0.415 / 0.6366$ $= .65$	**Cohen's *d*** Cohen's *d* can be computed using the two standard deviations. What is the magnitude of *d*, according to Cohen's standards? The mean of the treatment group is at the _____ percentile of the control group.
$g = M_1 - M_2 / \sqrt{MSwithin}$ $= 1.004 - 0.589 / \sqrt{0.41}$ $= 0.415 / 0.6408$ $= .65$	**Hedges's *g*** Hedges's *g* can be computed using the MSwithin. Hedges's *g* and Cohen's *d* are similar because the sample size is so large in this study.
$\Delta = M_1 - M_2 / \sigma_{control}$ $= 1.004 - 0.589 / 0.628$ $= 0.415 / 0.628$ $= .66$	**Glass's delta** Glass's delta can be computed using the standard deviation of the control group.
$r_{Y\lambda} = \sqrt{[F(1,\_) / (F(1,\_) + df\ error)]}$ $= \sqrt{[8.49 / (8.49 + 78)]}$ $= \sqrt{[8.49 / 86.490]}$ $= \sqrt{0.0982}$ $= .31$	**Effect size correlation** The effect size correlation was computed by SPSS as the correlation between the iv (TREATGRP) and the dv (GLOBAL4), $r_{Y\lambda} = .31$ The effect size correlation can also be

	computed from the F value.

The next computational is from the same study. This example uses Wolpe's Subjective Units of Disturbance Scale (SUDS) as the dependent measure. It is a single item, 11-point scale (0 = neutral; 10 = the highest level of disturbance imaginable) that measures the level of distress produced by thinking about a trauma. SUDS scores are measured immediately post treatment for those assigned to the EMDR treatment group and at the second pretreatment testing for those assigned to the delayed treatment condition. The SPSS output from the T-TEST and CORR(elation) procedures is shown below.

```
t-tests for Independent Samples of TREATGRP    TREATMENT GROUP


                             Number
Variable                     of Cases    Mean        SD    SE of Mean
---------------------------------------------------------------------
SUDS4   POST-TEST SUDS

TREATMENT GROUP                40        2.7250     2.592      .410
DELAYED TRMT GROUP             40        7.5000     2.038      .322
---------------------------------------------------------------------

        Mean Difference = -4.7750

        Levene's Test for Equality of Variances: F= 1.216  P= .274


     t-test for Equality of Means                            95%
Variances   t-value     df    2-Tail Sig    SE of Diff      CI for Diff
----------------------------------------------------------------------
Unequal       -9.16   73.89        .000         .521     (-5.814, -3.736)
----------------------------------------------------------------------
                      - -  Correlation Coefficients  - -

           SUDS4

TREATGRP      .7199
           (    80)
           P= .000

(Coefficient / (Cases) / 2-tailed Significance)
```

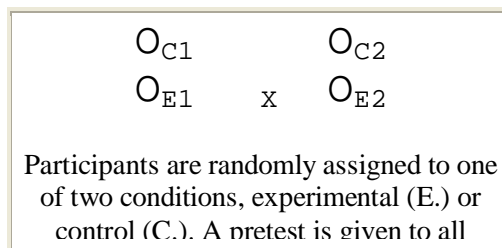Use the data in the above table to compute each of the following ES statistics:

	Cohen's *d* Compute Cohen's *d* using the two standard deviations. How large is the *d* using Cohen's interpretation of effect sizes?
	Cohen's *d* Compute Cohen's *d* using the value of the *t*-test statistic.

	Are the two values of *d* similar?
	Hedges's *g* Compute Hedges's *g* using the *t*-test statistic.
	Glass's delta Calculate Glass's delta using the standard deviation of the control group.
	Effect size correlation The effect size correlation was computed by SPSS as the correlation between the iv (TREATGRP) and the dv (SUDS4), $r_{Y\lambda}$ = . Calculate the effect size correlation using the *t* value.
	Effect size correlation Use Cohen's *d* to calculate the effect size correlation.

III. Effect Size Measures for Two Dependent Groups.

There is some controversy about how to compute effect sizes when the two groups are dependent, e.g., when you have matched groups or repeated measures. These designs are also called *correlated designs.* Let's look at a typical repeated measures design.

A Correlated (or Repeated Measures) Design

$$O_{C1} \qquad O_{C2}$$
$$O_{E1} \quad X \quad O_{E2}$$

Participants are randomly assigned to one of two conditions, experimental (E.) or control (C.). A pretest is given to all

participants at time 1 ($O_{.1.}$). The treatment is administered at "X". Measurement at time 2 (O_{E2}) is posttreatment for the experimental group. The control group is measured a second time at (O_{C2}) without an intervening treatment.. The time period between $O_{.1}$ and $O_{.2}$ is the same for both groups.

This research design can be analyzed in a number of ways including by gain scores, a 2 x 2 ANOVA with measurement time as a repeated measure, or by an ANCOVA using the pretest scores as the covariate. All three of these analyses make use of the fact that the pretest scores are correlated with the posttest scores, thus making the significance tests more sensitive to any differences that might occur (relative to an analysis that did not make use of the correlation between the pretest and posttest scores).

An effect size analysis compares the mean of the experimental group with the mean of the control group. The experimental group mean will be the posttreatment scores, O_{E2}. But any of the other three means might be used as the control group mean. You could look at the ES by comparing O_{E2} with its own pretreatment score, O_{E1}, with the pretreatment score of the control group, O_{C1}, or with the second testing of the untreated control group, O_{C2}. Wilson, Becker, and Tinker (1995) computed effect size estimates, Cohen's d, by comparing the experimental group's posttest scores (O_{E2}) with the second testing of the untreated control group (O_{C2}). We choose O_{C2} because measures taken at the same time would be less likely to be subject to history artifacts, and because any regression to the mean from time 1 to time 2 would tend to make that test more conservative.

Suppose that you decide to compute Cohen's d by comparing the experimental group's pretest scores (O_{E2}) with their own pretest scores (O_{E1}), how should the pooled standard deviation be computed? There are two possibilities, you might use the original standard deviations for the two means, or you might use the paired t-test value to compute Cohen's d. Because the paired t-test value takes into account the correlation between the two scores the paired t-test will be larger than a between groups t-test. Thus, the ES computed using the paired t-test value will always be larger than the ES computed using a between groups t-test value, or the original standard deviations of the scores. Rosenthal (1991) recommended using the paired t-test value in computing the ES. A set of meta-analysis computer programs by Mullen and Rosenthal (1985) use the paired t-test value in its computations. However, Dunlop, Cortina, Vaslow, & Burke (1996) convincingly argue that the original standard deviations (or the between group t-test value) should be used to compute ES for correlated designs. They argue that if the pooled standard deviation is corrected for the amount of correlation between the measures, then the ES estimate will be an overestimate of the actual ES. As shown in Table 2 of Dunlop et al., the overestimate is dependent upon the magnitude of the correlation between between the two scores. For example, when the correlation between the scores is at least .8, then the ES estimate is more than twice the magnitude of the ES computed using the original standard deviations of the measures.

The same problem occurs if you use a one-degree of freedom F value that is based on a repeated measures to compute an ES value.

In summary, when you have correlated designs you should use the original standard deviations to compute the ES rather than the paired t-test value or the within subject's F value.

IV. Meta Analysis

Overview

A meta-analysis is a summary of previous research that uses quantitative methods to compare outcomes across a wide range of studies. Traditional statistics such as t tests or F tests are inappropriate for such comparisons because the values of those statistics are partially a function of the sample size. Studies with equivalent differences between treatment and control conditions can have widely varying t and F statistics if the studies have different sample sizes. Meta analyses use some estimate of effect size because effect size estimates are not influenced by sample sizes. Of the effect size estimates that were discussed earlier in this page, the most common estimate found in current meta analyses is Cohen's d.

In this section we look at a meta analysis of treatment efficacy for posttraumatic stress disorder (Van Etten & Taylor, 1998). For those of you interested in the efficacy of other psychological and behavioral treatments I recommend the influential paper by Lipsey and Wilson (1993).

The Research Base

The meta analysis is based on 61 trials from 39 studies of chronic PTSD. Comparisons are made for Drug treatments, Psychological Treatments and Controls.

Drug treatments include: selective serotonin reuptake inhibitors (SSRI; new antidepressants such as Prozac and Paxil), monoamine oxidase inhibitors (MAOI, antidepressants such as Parnate and Marplan), tricyclic antidepressants (TCA; antidepressants such as Toffranil), benzodiazepines (BDZ; minor tranquilizers such as Valium), and carbamazepine (Carbmz; anticonvulsants such as Tegretol).

The psychotherapies include: behavioral treatments (primarily different forms of exposure therapies), eye movement desensitization and reprocessing (EMDR), relaxation therapy, hypnosis, and psychodynamic therapy.

The control conditions include: pill placebo (used in the drug treatment studies), wait list controls, supportive psychotherapy, and no saccades (a control for eye movements in EMDR studies).

Data Reduction Procedures

Effect sizes were computed as Cohen's *d* where a positive effect size represents improvement and a negative effect size represents a "worsening of symptoms."

Ninety-percent confidence intervals were computed. Comparisons were made base on those confidence intervals rather than on statistical tests (e.g., *t* test) of the mean effect size. If the mean of one group was not included within the 90% confidence interval of the other group then the two groups differed significantly at $p < .10$.

Comparisons across conditions (e.g., drug treatments vs. psychotherapies) were made by computing a weighted mean for each group were the individual trial means were weighted by the number of cases for the trial. This procedure gives more weight to trials with larger *n*s, presumably the means for those studies are more robust.

Treatment Comparisons

For illustrative purposes lets look at the Self Report measures of the Total Severity of PTSD Symptoms from Table 2 (Van Etten & Taylor, 1998).

Overall treatments. The overall effect size for psychotherapy treatments (M = 1.17; 90% CI = 0.99 - 1.35) is significantly greater than both the overall drug effect size (M = 0.69; 90% CI = 0.55 - 0.83) and the overall control effect size (M = 0.43; 90% CI = 0.33 - 0.53). The drug treatments are more effective than the controls conditions.

Within drug treatments. Within the drug treatments SSRI is more effective than any of the other drug treatments.

Within psychotherapies. Within the psychotherapies behavior modification and EMDR are equally effective. EMDR is more effective than any of the other psychotherapies. Behavior modification is more effective than relaxation therapy.

Within Controls. Within the control conditions the alternatives the pill placebo and wait list controls produce larger effects than the no saccade condition.

Across treatment modalities. EMDR is more effective than each of the drug conditions except the SSRI drugs. SSRI and EMDR are equally effective. Behavior modification is more effective than TCAs, MAOIs and BDZs, it is equally effective as the SSRIs and Carbmxs.

Behavior Modification and EMDR are more effective than any of the control conditions.

Self Report for Total Severity of PTSD Symptoms		
Condition	M	90% CI
TCA	.054	.034 - 0.74
Carbmz	0.93	
MAOI	0.61	0.38 - 0.84
SSRI	1.38	1.02 - 1.74
BDZ	0.49	
Drug Tx	0.69	0.55 - 0.83
Behav Tx	1.27	0.80 - 1.74
EMDR	1.24	0.99 - 1.49
Relaxat'n	0.45	
Hypnosis	0.94	
Dynamic	0.90	
Psych Tx	1.17	0.99 - 1.35
Pill Placebo	0.51	0.29 - 0.73
WLC	0.44	0.28 - 0.60
Supp Psyc	0.34	0.01 - 0.67

It is also interesting to note that the drop out rates for drug therapies (M = 31.9; 90% CI = 25.4 - 38.4) are more than twice the rate for psychotherapies (M = 14.0; 90% CI = 10.8 - 17.2).

No sacc	0.22	
Controls	0.43	0.33 - 0.53

Fail Safe N

One of the problems with meta analysis is that you can only analyze the studies that have been published. There is the **file drawer problem**, that is, how many studies that did not find significant effects have not been published? If those studies in the file drawer had been published then the effect sizes for those treatments would be smaller.

The fail safe N is the number of nonsignificant studies that would be necessary to reduce the effect size to an nonsignificant value, defined in this study as an effect size of 0.05. The fail safe Ns are shown in the table at the right.

The fail safe Ns for Behavior therapies, EMDR, and SSRI are very large. It is unlikely that there are that many well constructed studies sitting in file drawers. On the other hand, the fail safe N's for BDZ, Carbmz, relaxation therapy, hypnosis, and psychodynamic therapies are so small that one should be cautious about accepting the validity of the effect sizes for those treatments.

Condition	Fail Safe N
TCA	41
Carbmz	23
MAOI	66
SSRI	95
BDZ	10
Behav Tx	220
EMDR	139
Relaxat'n	8
Hypnosis	18
Dynamic	17

V. Effect Size Measures in Analysis of Variance

Measures of effect size in ANOVA are measures of the degree of association between and effect (e.g., a main effect, an interaction, a linear contrast) and the dependent variable. They can be thought of as the correlation between an effect and the dependent variable. If the value of the measure of association is squared it can be interpreted as the proportion of variance in the dependent variable that is attributable to each effect. Four of the commonly used measures of effect size in AVOVA are:

- Eta squared, η^2
- partial Eta squared, η_p^2
- omega squared, ω^2
- the Intraclass correlation, ρ_I

Eta squared and partial Eta squared are estimates of the degree of association for the sample. Omega squared and the intraclass correlation are estimates of the degree of

association in the population. SPSS for Windows displays the partial Eta squared when you check the "display effect size" option in GLM.

See the following SPSS lecture notes for additional information on these ANOVA-based measures of effect size: Effect size measures in Analysis of Variance

---

VI. References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Dunlop, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170-177.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48,* 1181-1209.

Mullen, H., & Rosenthal, R. (1985). *BASIC meta-analysis: Procedures and programs.* Hillsdale, NJ: Earlbaum.

Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

Rosenthal, R. & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw Hill.

Rosenthal, R. & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99,* 400-406.

Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods, 1,* 331-340.

Van Etten, Michelle E., & Taylor, S. (1998). Comparative efficacy of treatments for posttraumatic stress disorder: A meta-analysis. *Clinical Psychology & Psychotherapy*, *5,* 126-144.

Wilson, S. A., Becker, L. A., & Tinker, R. H. (1995). Eye movement desensitization and reprocessing (EMDR) treatment for psychologically traumatized individuals. *Journal of Consulting and Clinical Psychology, 63,* 928-937.