UNDERSTANDING STATISTICS, *3*(1), 1–25
Copyright © 2004, Lawrence Erlbaum Associates, Inc.

TEACHING ARTICLES

Interpreting Effect Sizes in Contrast Analysis

R. Michael Furr
Department of Psychology
Appalachian State University

Although researchers are becoming more aware of the benefits of reporting effect sizes, the usefulness of effect sizes can be enhanced if researchers have a firm understanding of how to interpret the various effect sizes that are available. This article articulates and illustrates the interpretation of 3 pairs of effect sizes developed for use in contrast analysis (Rosenthal, Rosnow, & Rubin, 2000). Three ways of conceptualizing the effect sizes are discussed: (a) as correlations between predicted and observed data, (b) as proportion of variance accounted for, and (c) as parallel to a multiple regression approach. It is hoped that this interpretive aid helps increase the frequency with which effect sizes are reported and the effectiveness with which they are used.

contrast analysis, effect size, significance testing, correlation

Social scientists are increasingly promoting the importance and utility of effect sizes (e.g., Cohen, 1994). One outcome of the vociferous debate over the role of significance testing has been a deeper appreciation of the role of effect sizes. In fact, the *Publication Manual of the American Psychological Association* (American Psychological Association [APA], 2001) tells authors that "For readers to fully understand your findings, it is almost always necessary to include some index of effect size or strength of relation in your Results section" (p. 25). Along with this endorsement by APA, a growing number of journals have instituted explicit requirements or recommendations encouraging the reporting of effect sizes (Keselman et al., 1998; Kirk, 1996; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000; Wilkinson & The APA Task Force on Statistical Inference, 1999).

Requests for reprints should be sent to R. Michael Furr, Department of Psychology, Appalachian State University, Boone, NC 28607. E-mail: furrrm@appstate.edu

Despite the growing appreciation of effect sizes among methodologists, the wider research community appears to be adopting effect sizes rather slowly (Thompson, 1999b). The strong stances recently taken by journal editors and in the *Publication Manual of the American Psychological Association* (APA, 2001) will likely accelerate the reporting of effect sizes, but this is only half the battle. Social research will be advanced not only through the reporting effect sizes, but also through the accurate interpretation of effect sizes. As Vacha-Haase et al. (2000) pointed out, “many researchers do not fully understand the logic of their statistical tests … and therefore may remain oblivious to the need for effect-size reporting *and interpretation*” (p. 414, italics added). Effect sizes can be most fully appreciated and, more important, most usefully interpreted if researchers have a solid grasp of their logic and meanings. This article describes and illustrates the interpretations of several effect sizes in the context of contrast analysis.

Along with an increased appreciation for effect sizes, many researchers have recently argued for the utility of contrast analyses (e.g., Furr & Rosenthal, 2003a; Loftus, 1996). Contrast analysis is designed to address focused analytic questions; in other words, to evaluate theories efficiently. For example, if a clinical researcher predicts that a treatment group will show more rapid symptom decrease than a control group, he or she could conduct a contrast analysis to obtain a significance test and effect size directly reflecting the degree to which the obtained pattern of results matches the predicted pattern of results. The popular traditional alternative would be to compute an analysis of variance (ANOVA) and hope for a significant main effect and interaction, which would be probed with a series of conservative post hoc analyses. Such an approach would likely lead to no single significance test or effect size directly related to the researcher’s specific theoretical prediction. In their comprehensive discussion of contrast analysis, Rosenthal, Rosnow, and Rubin (2000) outlined three basic correlational effect sizes. Each of these correlational effect sizes can be interpreted in two ways.

First, as with any correlation, a *correlational effect size* reflects the association between two things. Specifically, it can be interpreted as a correlation between the observed data and a theoretically predicted pattern of data. However, to fully understand the meaning of such an effect size, researchers must understand *which parts* of the observed and predicted data are being correlated. For example, does the data refer to the individual scores, group means, or an adjusted set of scores? Second, as with any correlation, a correlational effect size from contrast analysis can be squared and interpreted as “the proportion of variation accounted for” by the contrast. However, to fully understand the meaning of such an effect size, researchers must understand *which part* of the variation is being considered. Is it the total variation, the between-groups variation, or the error variation?

This article illustrates and articulates the meaning of three fundamental effect sizes associated with contrast analysis. In doing so, it discusses each from both the

TABLE 1
Interpreting Effect Size Correlations in Contrast Analysis

Type of Effect Size	*Unsquared*	*Squared*	*Regression Equivalent*
Effect size	Correlation between the contrast weights and the individuals' observed scores	Proportion of *total variation* that is explained by the contrast	Zero-order correlation
Alerting	Correlation between the contrast weights and the observed group means	Proportion of *between-group (explained) variation* that is explained by the contrast	Correlation (at group level) between contrast weights and group means
Contrast	Correlation between the contrast weights and the individuals' scores adjusted for (removing) between-group variation related to other contrasts (i.e., between-group variation unrelated to the given contrast)	Proportion of *variation unrelated to other contrasts* that is explained by the contrast	Partial correlation (correlation between contrast weights and the individuals' observed scores, partialling out the other contrast weights)

"unsquared" and "squared" perspectives: $r_{effect\ size}$ (and $r^2_{effect size}$), $r_{alerting}$ (and $r^2_{alerting}$), and $r_{contrast}$ (and $r^2_{contrast}$). In addition, this article discusses convergence between the contrast analysis effect sizes and regression-based statistics. Table 1 presents a summary of the interpretations to be presented. Rosenthal et al. (2000) outlined some of the topics that are articulated in this article, but their book deals with a host of rather technical facets of contrast analysis. The purpose of this article is to focus solely on articulating the interpretation of the correlational effect sizes at a relatively nontechnical, intuitive level—to distill the issues involved in interpretation into one accessible source.

ILLUSTRATIVE DATA AND CONTRASTS

Imagine that a researcher has four groups of majors, each with five students measured on empathy. Table 2 presents the participants' data and group means.[1] With four majors included in the study, the main effect of major has three degrees of free-

[1]These data are fabricated simply to illustrate the relevant issues and to allow readers to conduct their own computations with relative ease. The sample sizes and effect sizes are not representative of those typically found in the behavioral sciences.

TABLE 2
Example Data

Major							
Psychology		*Education*		*Business*		*Chemistry*	
Participant	*Empathy*	*Participant*	*Empathy*	*Participant*	*Empathy*	*Participant*	*Empathy*
1	51	6	62	11	50	16	50
2	56	7	67	12	49	17	45
3	61	8	57	13	47	18	40
4	58	9	65	14	45	19	49
5	54	10	59	15	44	20	41
M	56		62		47		45
Variance	14.5		17		6.5		20.5

Note. Grand $M = 52.5$.

dom, is significant at $p < .05$, and accounts for 80% of the total variation in empathy scores, as shown by the ANOVA results presented in Table 3. Furthermore, imagine that, on the basis of theoretical predictions, the researcher is interested in three specific hypotheses:

Contrast A: Psychology majors have higher empathy scores than Education majors.
Contrast B: Business majors have higher empathy scores than Chemistry majors.
Contrast C: On average, Psychology and Education majors have higher empathy scores than Business and Chemistry majors.

In this case, the three contrasts are orthogonal (i.e., independent) and completely account for the between-groups effect. The independence and exhaustiveness is reflected in the fact that the sums of squares for the three contrasts sum to the sum of squares for the main effect of Major ($SS_A + SS_B + SS_C = SS_{Between}$; see Table 3 for the ANOVA results associated with each contrast). When researchers examine a set of independent and exhaustive contrasts, they will fully account for all between-groups variation. That is, the between-groups variation will be completely partitioned among the contrasts. However, it is not necessary that researchers use multiple contrasts in the analysis of a given data set, nor is it necessary that multiple contrasts (if they are used) be orthogonal. The logic and interpretation remains the same, even though the effect sizes might be calculated in slightly different ways. The contrasts used in a given set of analysis should primarily depend on the researcher's goals and theoretical questions. To more fully understand the effect sizes associated with contrast analysis and the results presented in Table 3, some background on contrast analysis is necessary.

BASICS OF CONTRAST ANALYSIS

Researchers using contrast analysis to examine a hypothesis first quantify the hypothesis and then evaluate the degree to which the quantified hypothesis matches the quantitative data that are actually obtained during data collection. A basic between-groups contrast analysis of a specific hypothesis can be seen as a four-step process (Furr & Rosenthal, 2003a) in which the first two steps lead to quantification of the hypothesis (predictions about the pattern of data to be collected) and the third and fourth steps perform significance tests and compute effect sizes reflecting the degree to which the predicted data match the actual data. Researchers who are well-acquainted with contrast analysis can use familiar standard contrast weights presented in many textbooks, if indeed those weights are relevant for the hypothesis in question, and can proceed to the third step described later. For researchers less familiar with contrast analysis or for researchers dealing with hypotheses that do not fit standard sets of contrast weights, the first two steps might help identify appropriate ways of quantifying hypotheses and beginning contrast analysis.

The first step is to translate each hypothesis into numbers. A straightforward way to do this when working with group means, as in this example, is to consider each hypothesis separately and predict the approximate relative mean scores that each group should have, given the contrast in question. In this example, the researcher has three specific hypothesis, and thus creates three sets of "quantified" hypotheses. Assuming that scores on the empathy measure typically range from 40 to 70, Table 4 presents sets of predicted values that reflect the three hypotheses. For Contrast A, the two values in the set (70 and 60) reflect the prediction that the

TABLE 3
Analysis of Variance of Example Data

Source	*SS*	*df*	*MS*	*F/t*[a]	*p*[b]	η^2
Total sample						
Between-groups (major)	945	3	315	21.54	< .0001	.80
Contrast A	90	1	90	–2.48	.99	
Contrast B	10	1	10	0.83	.21	
Contrast C	845	1	845	7.60	< .0001	
Within-groups (error)	234	16	14.625			
Total	1,179	19				
Psychology and education majors only						
Between-groups (major)	90	1	90	5.71	.05	.42
Contrast A	90	1	90	–2.39	.98	
Within-groups (error)	126	8	15.750			
Total	216	9				

[a]The test statistics are an *F* for the main effect of major and a *t* for the contrasts. [b]The *p* values for the contrasts are one-tailed.

TABLE 4
Creating Weights for Contrasts A, B, and C

	Predicted Means			*Contrast Weights (λs)*		
Major	*Contrast A*	*Contrast B*	*Contrast C*	*Contrast A*	*Contrast B*	*Contrast C*
Psychology	70		65	+1		+1
Education	60		65	–1		+1
Business		50	45		+1	–1
Chemistry		40	45		–1	–1

Psychology group will have a higher mean empathy score than the Education group. Note that the other two groups are irrelevant to this contrast. For Contrast B, the two values (50 and 40) reflect the prediction that the Business group will have a higher mean empathy score than the Chemistry group. For Contrast C, the four values (65, 65, 45, and 45) reflect the prediction that the Psychology and Education students will have a higher mean than the Business and Chemistry students. Because Contrast C does not differentiate between the Psychology and Education students, both groups are given the same predicted mean (65), as are the Business and Chemistry students (45). The real goal in this step is, for each hypothesis separately, to identify which group or groups are predicted to have relatively higher means than other groups.

The second step is to translate each quantified hypothesis into a set of contrast weights (λs). This can be accomplished by computing the mean of the predicted values for a given contrast and subtracting this mean from each of the predicted values associated with the contrast. For Contrast A, the mean is 65, so the contrast weights for the Psychology and Education groups could be 5 and –5 respectively. Analyses are simplified when contrast weights are the smallest whole numbers that could represent the theory. For Contrast A, the researcher could divide the contrast weights by 5 to obtain weights of +1 and –1. This transformation simplifies computations but has no impact on the resulting *p* values or effect sizes. This procedure retains the predicted pattern and ensures that the weights for a given contrast sum to zero. Table 4 presents weights associated with each of the three contrasts.[2]

[2]Although the contrast analyses in this illustration are relatively simple, these two steps are entirely applicable to even more complex contrast analyses. For example, a "trend analysis" could evaluate the specific hypothesis that the psychology students have a larger mean than the education students, who in turn have a larger mean than the business students, who have a larger mean than the chemistry students. A researcher testing this hypothesis could use the predicted values of 70, 60, 50, and 40, respectively, all in a single set, which would be transformed to contrast weights of 3, 1, –1, and –3 respectively. Similarly, the procedures could be used in cases in which there are an uneven number of groups. For example, a researcher could test the hypothesis that psychology and education students will have a larger mean than business, chemistry, and dance students. One way of setting up this prediction would be to

The third step is to compute a significance test (t test) and probability level for each contrast.

$$t = \frac{\sum_{j=1}^{k} M_j \lambda_j}{\sqrt{s^2_{pooled}[\sum_{j=1}^{k} \frac{\lambda_j^2}{n_j}]}}, \qquad (1)$$

where M_j is the mean for group j (of k groups), λ_j is the contrast weight for group j, n_j is the number of participants in group j, and s^2_{pooled} is the pooled within-group variance (across *all* groups). When all groups have the same sample size, s^2_{pooled} is simply the mean of the k variances, or one could use the MS_{Within} from the ANOVA. For contrast A:

$$t = \frac{(56)(+1) + (62)(-1)}{\sqrt{14.625[(\frac{+1^2}{5}) + (\frac{-1^2}{5})]}} = \frac{-6}{2.42} = -2.48.$$

The Psychology group was predicted to have a higher mean than the Education group, but the observed means actually went in the opposite direction. This results in a negative t value and a one-tailed probability level of $p = .99$ for this contrast.[3] The sum of squares associated with a given contrast can be computed from the t value and the pooled variance (or MS_{Within}, equivalently):

$$SS = t^2(s^2_{pooled}). \qquad (2)$$

For Contrast A:

$$SS = (-2.48)^2(14.625) = 90.$$

The fourth step is to compute an effect size and at least three pairs of useful correlational effect sizes are available: $r_{effect\,size}$ (and $r^2_{effect\,size}$), $r_{contrast}$ (and $r^2_{contrast}$),

---

use the predicted values of 65, 65, 45, 45, and 45, respectively, which would be transformed into contrast weights of 3, 3, –2, –2, and –2.

[3]The use of a one-tailed probability level follows from the logic of contrast analysis. Contrast analysis is based on the evaluation of hypotheses about specific patterns of data, such as the hypothesis that Group A will have a larger mean than Group B. The logic of probability values suggests that, when one has a clearly directional hypothesis, one-tailed tests are appropriate. Nevertheless, researchers often choose to present two-tailed probability levels, even when evaluating clear directional hypotheses. Despite the potential suspicion with which one-tailed probability values may be viewed by reviewers and readers, they are entirely consistent with the logic of contrast analysis.

and $r_{alerting}$ (and $r^2_{alerting}$). All three pairs reflect different kinds of effects and are potentially quite useful. The following sections of this article articulate and illustrate their logic and meanings. The three "unsquared" effect sizes are described and illustrated, then the three "squared" effect sizes are discussed, and finally the convergence with regression is presented.

"UNSQUARED" EFFECT SIZES: CORRELATIONS BETWEEN OBSERVED AND PREDICTED PATTERNS OF DATA

As just outlined, the weights associated with a given contrast are essentially a quantification of the predicted pattern of data, based on theory or prior research. Accordingly, "unsquared" effect sizes reflect the degree to which observed data match (i.e., are correlated with) the predicted pattern of data. Each of the three contrast analysis effect sizes is associated with a different form of the observed data. See Table 5 for a summary of the unsquared effect sizes for each of the three contrasts.

$r_{effect\ size}$

Perhaps the most straightforward effect size for a contrast is $r_{effect\ size}$, which is the correlation between individuals' observed scores and the contrast weights that reflect the predicted pattern of data. Table 6 presents the data in a way that might illustrate this most clearly for Contrast A. Each participant has two scores: (a) the person's observed empathy score and (b) the contrast weight associated with his or her Major, as defined for the Contrast A (see Table 4). The $r_{effect\ size}$ is the correlation between the two sets of scores.

For Contrast A, focusing only on the Psychology and Education students, we temporarily ignore, or set aside, the Business and Chemistry majors (see the top half of Table 6). The correlation between the empathy scores and the contrast

TABLE 5
Summary of Effect Sizes for Contrasts A, B, and C

Contrast	$r_{effect\ size}$	$r_{alerting}$	$r_{contrast}$	$r^2_{effect\ size}$	$r^2_{alerting}$	$r^2_{contrast}$
Setting aside irrelevant groups						
Contrast A	–.65	–1.00	–.65	.42	1.00	.42
Contrast B	.29	1.00	.29	.08	1.00	.08
Including all groups						
Contrast A	–.28	–.31	–.53	.08	.10	.28
Contrast B	.09	.10	.20	.01	.01	.04
Contrast C	.85	.95	.88	.72	.89	.78

TABLE 6
Illustration of Unsquared Effect Sizes for Contrast A

$r_{effect\ size}$			$r_{alerting}$			$r_{contrast}$		
Participant	*Empathy*	*Contrast Weight*	*Major*	*M Empathy*	*Contrast Weight*	*Participant*	*Adjusted Empathy*	*Contrast Weight*
Setting aside business and chemistry majors								
1	51	+1	Psy	56	+1	1	51	+1
2	56	+1	Edu	62	–1	2	56	+1
3	61	+1				3	61	+1
4	58	+1				4	58	+1
5	54	+1				5	54	+1
6	62	–1				6	62	–1
7	67	–1				7	67	–1
8	57	–1				8	57	–1
9	65	–1				9	65	–1
10	59	–1				10	59	–1
Including business and chemistry majors								
1	51	+1	Psy	56	+1	1	44.5	+1
2	56	+1	Edu	62	–1	2	49.5	+1
3	61	+1	Bus	47	0	3	54.5	+1
4	58	+1	Che	45	0	4	51.5	+1
5	54	+1				5	47.5	+1
6	62	–1				6	55.5	–1
7	67	–1				7	60.5	–1
8	57	–1				8	50.5	–1
9	65	–1				9	58.5	–1
10	59	–1				10	42.5	–1
11	50	0				11	55.5	0
12	49	0				12	54.5	0
13	47	0				13	52.5	0
14	45	0				14	50.5	0
15	44	0				15	49.5	0
16	50	0				16	57.5	0
17	45	0				17	52.5	0
18	40	0				18	47.5	0
19	49	0				19	56.5	0
20	41	0				20	48.5	0

Note. Psy = psychology; Edu = education; Bus = business; Che = chemistry.

weights for the 10 Psychology and Education students, $r_{effect\ size} = -.65$. The correlation is negative because the Psychology mean score was lower than the Education mean score, which contradicts predictions. Instead of computing the correlation directly, researchers could obtain this $r_{effect\ size}$ from the ANOVA of the data from only the groups involved in the contrast (the bottom half of Table 3 presents the ANOVA results for only the Psychology and Education students):

$$r_{effect\ size} = \sqrt{\frac{t^2}{F_{Between}(df_{Between}) + df_{Within}}}, \tag{3}$$

where t is the t value associated with the contrast (see Equation 1), $F_{Between}$ and $df_{Between}$ are from the main effect of major in the ANOVA, and df_{Within} is from the error term in the ANOVA.[4] For Contrast A:

$$r_{effect\ size} = \sqrt{\frac{(-2.39)^2}{5.714(1) + 8}} = .65 \rightarrow -.65.$$

Researchers using this formula must be aware of the appropriate sign of the effect size.

In some cases, the researcher might choose to include the Business and Chemistry students in the analysis of Contrast A, rather than setting them aside. For example, if the researcher actually hypothesized that the Business and Chemistry students' empathy scores would fall in between the Psychology and Education students' scores, or if the researcher wished to put the contrast in the context of the total variation in the whole sample, then he or she would compute $r_{effect\ size}$ using the empathy scores and contrast weights for all 20 participants ($r_{effect\ size} = -.28$). Or, the researcher could use Equation 3, but use the t value and $F_{Between}$ from the ANOVA conducted on the whole sample:

$$r_{effect\ size} = \sqrt{\frac{(-2.48)^2}{21.54(3) + 16}} = .28 \rightarrow -.28.$$

The procedure that includes all participants has a nice convergence with the results of the full ANOVA and this convergence will be discussed in more detail in the section on squared effect sizes.

[4]Although the logic and interpretation is exactly the same, a technical consideration is the possibility of using the error term from the ANOVA conducted on the whole sample because it is based on more observations and thus would be considered a better estimate of the population error term. In this case, the t value for Contrast A = –2.48, $F_{\text{Between}} = 6.15$, and $r_{effect\ size} = -.66$.

$r_{alerting}$

A second effect size for a contrast is $r_{alerting}$, which is the correlation between the observed group means and the contrast weights reflecting the predicted pattern of group means. Again, Table 6 presents the data in a way that illustrates this for Contrast A. Each Major has two scores: (a) the observed mean empathy score and (b) the contrast weight associated with the Major, as defined by the given contrast. The $r_{alerting}$ is the correlation between the two sets of scores.

Although $r_{alerting}$ can be quite informative for many purposes, for contrasts that focus on only two groups, as does Contrast A, $r_{alerting}$ is not very useful. In such cases, the correlation between group means and predicted means has zero degrees of freedom and thus will either be +1 (the mean difference is in the predicted direction), –1 (the mean difference is in the opposite of the predicted direction), or 0 (the means are not different).

However, for cases in which there are more than two groups, as in Contrast C or in Contrast A when the Business and Chemistry groups are not set aside, $r_{alerting}$ can be very useful. For Contrast A, including the Business and Chemistry groups, $r_{alerting} = -.31$. For Contrast C, $r_{alerting} = .95$. Researchers can obtain $r_{alerting}$ from:

$$r_{alerting} = \sqrt{\frac{t^2}{F_{Between}(df_{Between})}}, \quad (4)$$

where t is the t value associated with the contrast (see Equation 1), and $F_{Between}$ and $df_{Between}$ are from the main effect of major in the original ANOVA (see Table 3). For Contrast A including all groups:

$$r_{alerting} = \sqrt{\frac{(-2.48)^2}{21.54(3)}} = .31 \rightarrow -.31.$$

Again, researchers using this formula must be aware of the appropriate sign of the effect size. For Contrast C:

$$r_{alerting} = \sqrt{\frac{(7.60)^2}{21.54(3)}} = .95.$$

$r_{contrast}$

A third effect useful size in contrast analysis, the $r_{contrast}$, is a partial correlation. The $r_{contrast}$ effect size highlights the unique association between the contrast and that part of the outcome variable that is unrelated to other known sources of variation. The goal in computing $r_{contrast}$ for a given contrast is to remove variability in the outcome that is associated with any possible contrasts other than the given contrast.

One way to obtain this effect size is to create adjusted scores for each individual; another method that could be used when multiple contrasts are included is to conduct a regression analysis. The regression procedure will be discussed later. The two effect sizes previously illustrated ($r_{effect\ size}$ and $r_{alerting}$) are correlations between observed data and the predicted pattern of data (either at the individual level or the group level), but the $r_{contrast}$ can be interpreted as a correlation between *adjusted* observed data and the predicted pattern of data. As detailed elsewhere (Rosenthal et al., 2000), the individuals' observed scores are adjusted by eliminating "all between-group sources of variation other than the contrast in question" (Rosenthal et al., 2000, p. 42). For the adjusted score method, a predicted score is computed for each person, based on the contrast weights (the predicted group mean subtracted from the person's deviation from the actual group mean). Then $r_{contrast}$ is the correlation between the adjusted scores and the contrast weights:

Adjusted score = (Individual's observed score – Group mean) + (Predicted score);

or,

Adjusted score = Residual + (Grand mean + Prediction weight);

or,

$$\text{Adjusted score} = (X_{ij} - M_j) + (M_{\bullet\bullet} + \lambda_j \frac{\sum_{j=1}^{k} M_j \lambda_j}{\sum_{j=1}^{k} \lambda_j^2}), \quad (5)$$

where X_{ij} is the observed score for individual *i* in group *j*, M_j is the mean for group *j*, $M_{\bullet\bullet}$ is the grand mean, λ_j and is the contrast weight associated with group *j* for the contrast in question. Because Contrast A concerns only the Psychology and Education majors, we temporarily set aside the Business and Chemistry majors—the only effect of which is that the Grand Mean (of the 10 Psychology and Education majors) is 59 instead of 52.5 (as it is for the entire 20-person sample). Thus, for Participant 1 in the Psychology group, the adjusted score for Contrast A is:

$$(51 - 56) + (59 + (+1)\frac{56(+1) + 62(-1)}{(+1)^2 + (-1)^2}) = -5 + (59 - 3) = 51.$$

Table 6 presents all participants' adjusted scores, based on Contrast A. For this contrast, the effect size $r_{contrast}$ = –.65. Although the logic of the Adjusted scores helps reveal the meaning of the $r_{contrast}$ effect size, the procedure is obviously unwieldy and researchers can compute the value more easily from:

$$r_{contrast} = \sqrt{\frac{t^2}{t^2 + df_{Within}}}. \tag{6}$$

In the current data:

$$r_{contrast} = \sqrt{\frac{-2.39^2}{-2.39^2 + 8}} = .65 \rightarrow -.65.$$

Note that for Contrast A, the $r_{contrast}$ computed when the Business and Chemistry students are set aside is equivalent to the $r_{effect\ size}$. This is because, by focusing only on two groups, there is no between-groups variation other than the variation between the two groups, thus nothing is partialled when $r_{contrast}$ is computed in this case. In cases in which other between-groups variation (apart from the contrast in question) does exist, $r_{contrast}$ and $r_{effect\ size}$ lead to importantly different results. For example, if we compute $r_{contrast}$ for Contrast A, without setting aside the Business and Chemistry majors, we obtain $r_{contrast} = -.53$. Similarly, if we compute $r_{contrast}$ for Contrast C, we obtain $r_{contrast} = .88$.

"SQUARED" EFFECT SIZES: "PROPORTION OF VARIATION ACCOUNTED FOR"

Many researchers may be more familiar with interpreting effect sizes as "proportion of variance accounted for" than as the degree of relation between predicted and observed data. For example, R^2 (from regression) and η^2 (from ANOVA) are popular effect sizes reflecting the proportion of total variation in scores on a dependent variable that is accounted for by variation on some predictor variable or set of predictor variables. An understanding of the effect sizes associated with contrast analysis might be facilitated by thinking of the ANOVA between-groups sums of squares ($SS_{Between}$ in Table 3) as "explained" variation and thinking of the error or within-groups sums of squares (SS_{Within} in Table 3) as "unexplained" variation.

From the ANOVA of the data in Table 2, SS_{Total} represents the degree to which all 20 students' empathy scores differ from each other and $SS_{Between}$ represents the differences among the 20 empathy scores that are associated with the fact that some majors tend to have higher empathy scores than other majors. So the mean empathy score differences between majors accounts for, or *explains,* some amount of the total differences among the individual students. As a matter of fact, the fact that some majors seem to have generally higher empathy scores than others explains 80% of the total variation in empathy:

$$\frac{SS_{Between}}{SS_{Total}} = \frac{SS_{Between}}{SS_{Between} + SS_{Within}} = \frac{945}{1179} = .80$$

(this is η^2 from the overall ANOVA).

Note that, at this point, this is a rather vague form of "explained"—yes, some groups are different from others, but the nature of the between-groups differences remains to be explored. For example, which groups are different from which? Are the differences between some groups bigger than the differences between others? Contrast analyses reveal these more specific facets of the general between-groups "explanation."

SS_{Within} represents the variation among empathy scores that is *not explainable* by differences between majors. For example, Allen and Beth are both Psychology majors, but Allen has a lower empathy score than Beth. Because they are both in the same major, the difference between the two participants' scores is "within-major" variation, which is unrelated to (and thus not explainable by) the fact that different majors tend to have different scores. There could be meaningful reasons why Allen and Beth have different empathy levels—maybe there are gender differences or perhaps intelligence is related to empathy. However, those potential explanations are *not* included in the current analyses, so the within-major variation is left unexplained.

$SS_{Contrast}$ for a given contrast represents the variation among empathy scores that is explained by that particular contrast. For example, as Table 3 shows, the sums of squares associated with Contrast A is 90—this is the "piece" of the explained empathy variation that is specifically explained by the average difference between Psychology and Education majors. Each of the squared effect sizes associated with contrasts identifies the "proportion of variation" in relation to different pieces of the variation.

$r^2_{effect\,size}$

Again, perhaps the most straightforward and familiar effect size for a given contrast is $r^2_{effect\,size}$, which can be interpreted as *the proportion of total variation that is explained by the contrast*:

$$r^2_{effect\,size} = \frac{\text{The piece of the variation explained by the given contrast}}{\text{Total variation}},$$

or,

$$r^2_{effect\,size} = \frac{SS_{Piece\ of\ the\ explained}}{SS_{Explained} + SS_{Unexplained}},$$

or,

$$r^2_{effect\,size} = \frac{SS_{Contrast}}{SS_{Total}}. \tag{7}$$

For Contrast A in this example,

$$r^2_{effect\,size} = \frac{SS_A}{SS_{Total}} = \frac{90}{1179} = .076.$$

So, about 8% of the empathy differences among all the students is explained by the fact that Psychology and Education majors have different mean empathy scores. However, consider again the issue, raised earlier, of how to treat the "irrelevant" Business and Chemistry groups for Contrast A. The SS_{Total} (1,179) entered into Equation 7 does not treat the Business and Chemistry majors as irrelevant—thus $r^2_{effect\,size}$ reflects the ratio of "the average empathy difference between Psychology and Education students" to "empathy differences among *all* students." Thus, this value of $r^2_{effect\,size}$ tells us the proportion of *all the differences among all the participants*, including those participants in the Business and Chemistry groups, that is explained by the average difference between Psychology and Education students. Note that $r^2_{effect\,size} = .076 = -.28^2$, where $-.28 = r_{effect\,size}$ from the "unsquared" analysis mentioned earlier.

Alternatively, we might choose to focus only on the differences among the 10 participants in the Psychology and Education groups, temporarily treating Business and Chemistry as "set-aside" groups. As shown in the bottom half of Table 3, the ANOVA of only the 10 Psychology and Education majors reveals that contrast variation is $SS_A = 90$, within-group variation is $SS_{Within} = 126$, and total variation is $SS_{Total} = 216$. Applying these values to Equation 7,[5]

$$r^2_{effect\,size} = \frac{SS_A}{SS_{Total}} = \frac{90}{216} = .42.$$

So, about 42% of the empathy differences *among the 10 Psychology and Education students* is explained by the fact that Psychology and Education majors have different mean empathy scores. Compare this to the previous finding that 8%

[5]As described in footnote 4, researchers treating business and chemistry as set-aside groups might opt to use the error term from the overall ANOVA because it is based on 20 observations instead of only 10. The MS_{Within} from the overall ANOVA is 14.625, which can be interpreted (roughly) as the average error per participant. Because a sums of squares value equals a mean squares value times degrees of freedom, we can obtain an estimate of SS_{Within} for the set-aside analysis by multiplying the MS_{Within} value from the overall ANOVA by the df_{Within} from the set-aside analysis (SS_{Within} for set aside = 14.625 × 8 = 117). To compute the SS_{Total} value (for Equation 6) for the set-aside analysis, we add the new SS_{Within} value to the SS_A value (117 + 90 = 207). Thus, $r^2_{effect\,size} = \frac{90}{207} = .43$.

of the empathy differences *among all 20 students* are explained by the fact that Psychology and Education majors have different mean empathy scores. Deciding which one of these effect sizes (treating Business and Chemistry as set-aside groups or not) is appropriate depends on the question that a researcher wishes to answer. Researchers might even choose to report both versions of $r^2_{effect size}$.

$r^2_{alerting}$

The second squared effect size, $r^2_{alerting}$, is also relatively straightforward, and can be interpreted as *the proportion of the explained variation that is explained by the contrast:*

$$r^2_{alerting} = \frac{\text{The piece of the variation explained by the given contrast}}{\text{All explained variation}},$$

or,

$$r^2_{alerting} = \frac{SS_{Piece\,of\,the\,explained}}{SS_{Explained}},$$

or,

$$r^2_{alerting} = \frac{SS_{Contrast}}{SS_{Between}}. \quad (8)$$

For Contrast A in this example,

$$r^2_{alerting} = \frac{SS_A}{SS_{Between}} = \frac{90}{945} = .095.$$

So, about 10% of the variation in mean empathy differences among all four groups is explained by the fact that Psychology and Education majors have different mean empathy scores. This value, although informative, may not exactly reflect the spirit of Contrast A, for which the Business and Chemistry majors could be considered irrelevant. Alternatively, the researcher could again focus only on the ANOVA for the 10 Psychology and Education majors. However such an analysis would be pointless in this context—it would tell the researcher that 100% of the variation in mean empathy differences among the Psychology group and the Education group is explained by the fact that Psychology and Education majors have different mean empathy scores. That is, the Contrast A is *all* of the between-groups variation, when the Business and Chemistry groups are set aside.

Contrast C provides a good illustration of the usefulness of $r^2_{alerting}$. In this case, there are no irrelevant groups, so for Contrast C:

$$r^2_{alerting} = \frac{SS_C}{SS_{Between}} = \frac{845}{945} = .89.$$

This value shows "where the action is" in the differences between groups. Specifically, it tells us that 89% of the variation among the four groups arises from the fact that Psychology and Education majors have a larger mean empathy score than do Business and Chemistry majors.

$r^2_{contrast}$

One way to approach the $r^2_{contrast}$ effect size is to consider it in light of the other two squared effect sizes. The $r^2_{effect size}$ reflects the proportion of the *total variation* that is explained by a particular contrast and the $r^2_{alerting}$ reflects the proportion of the *explained variation* that is explained by a particular contrast. The $r^2_{contrast}$ effect size can be seen as a way of expressing the particular contrast in relation to the *unexplained variation*, or the variation that is unexplained by any other contrasts:

$$r^2_{contrast} = \frac{\text{The piece of the variation explained by the given contrast}}{\text{Variation unexplained by any other contrast}},$$

or,

$$r^2_{contrast} = \frac{\text{The piece of the variation explained by the given contrast}}{\text{The piece of the variation explained by the given contrast + Unexplained variation}},$$

or,

$$r^2_{contrast} = \frac{SS_{Piece\,of\,the\,explained}}{SS_{Piece\,of\,the\,explained} + SS_{Unexplained}},$$

or,

$$r^2_{contrast} = \frac{SS_{Contrast}}{SS_{Contrast} + SS_{Within}}. \quad (9)$$

In this example, For Contrast A:

$$r^2_{contrast} = \frac{SS_A}{SS_A + SS_{Within}} = \frac{90}{90 + 234} = .28.$$

So, according to the $r^2_{contrast}$ for Contrast A, the mean empathy difference between Psychology and Education students accounts for 28% of the variation in empathy among all 20 students that is unrelated to all other between-group contrasts. That is, $r^2_{contrast}$ Contrast A accounts for 28% of the variation in empathy that is left over after all other between-group contrasts have been taken into consideration ("removed"). Or, put yet another way, it accounts for 28% of the variation in empathy that is related to the combination of mean differences between Psychology and Education students and of Unexplained variation. As seen with the unsquared effect sizes illustrated earlier, if Business and Chemistry majors are set aside, then $r^2_{contrast}$ is equivalent to $r^2_{effect\,size}$.

To summarize the squared effect sizes, the differences lie in the denominators of the formulae (illustrated with regard to Contrast A):

$$r^2_{effect\,size} = \frac{SS_A}{SS_A + SS_B + SS_C + SS_{Within}} = \frac{SS_A}{SS_{Between} + SS_{Within}} = \frac{90}{90 + 10 + 845 + 234} = \frac{90}{1{,}179} = .076,$$

$$r^2_{alerting} = \frac{SS_A}{SS_A + SS_B + SS_C} = \frac{SS_A}{SS_{Between}} = \frac{90}{90 + 10 + 845} = \frac{90}{945} = .095,$$

$$r^2_{contrast} = \frac{SS_A}{SS_A + SS_{Within}} = \frac{90}{90 + 234} = \frac{90}{324} = .278.$$

CONVERGENCE BETWEEN CONTRAST EFFECT SIZES AND MULTIPLE REGRESSION

Some researchers might find the connection between the contrast effect sizes and multiple regression analysis to be particularly revealing. Cohen (1968) pointed out that many ANOVA procedures can be performed fruitfully as multiple regression analysis, and contrast analysis is no exception. The "effect size" and "contrast" effect sizes have direct parallels to the output of typical regression analysis that focuses on individual-level data. Because the "alerting" effect size is at the group level, instead of the person level, it could be integrated with the other two effect sizes within a multilevel modeling (or hierarchical linear modeling) approach.

Imagine that the researcher conducts a regression analysis predicting individuals' scores on empathy from three variables reflecting Contrasts A, B, and C (see Table 7 for the likely data structure for such an analysis). The squared zero-order correlation for a predictor contrast (or its squared semi-partial correlation in this example because all the contrasts are orthogonal) for a predictor is equivalent to

TABLE 7
Data for Regression Analysis of Contrasts A, B, and C

Participant	*Empathy*	*Contrast A*	*Contrast B*	*Contrast C*
1	51	+1	0	+1
2	56	+1	0	+1
3	61	+1	0	+1
4	58	+1	0	+1
5	54	+1	0	+1
6	62	–1	0	+1
7	67	–1	0	+1
8	57	–1	0	+1
9	65	–1	0	+1
10	59	–1	0	+1
11	50	0	+1	–1
12	49	0	+1	–1
13	47	0	+1	–1
14	45	0	+1	–1
15	44	0	+1	–1
16	50	0	–1	–1
17	45	0	–1	–1
18	40	0	–1	–1
19	49	0	–1	–1
20	41	0	–1	–1

$r^2_{effect size}$, and the squared partial correlation for a predictor contrast is equivalent to $r^2_{contrast}$. A *partial correlation* for a particular predictor is the correlation between that portion of the outcome that is unrelated to all other predictors and that portion of the predictor that is unrelated to all other predictors. That is, the other predictors are partialled from the focal predictor and from the outcome variable. For Contrast A in this example, the partial correlation is the correlation between that part of empathy that is unrelated to Contrast B and C, and that part of Contrast A that is unrelated to Contrast B and C. In cases where all contrasts are orthogonal, as in this example, each contrast is inherently unrelated to all other contrasts.

As shown in Figure 1, contrast analysis and the $r^2_{effect size}$, $r^2_{alerting}$, and $r^2_{contrast}$ effect sizes can be expressed through a Venn diagram (e.g., Cohen & Cohen, 1983, p. 89), which might reveal the convergence with regression the "proportion of variation accounted for" interpretation even more clearly. Each contrast variable is represented by a circle, as is empathy. Each of the three contrasts explains, or is associated with, a *separate* part of empathy, which arises from the orthogonality of the three contrasts. In other words, each piece of variation explained by the three contrasts is "unique" variation. For example, the "a" portion of the diagram is that part of empathy uniquely explained by Contrast A, and the total variation in empathy is (a + b + c + d). Thus, for Contrast A:

$$r^2_{effect\,size} = \frac{a}{a+b+c+d} = \text{ squared correlation}$$

$$\mathrm{r}^2_{contrast} = \frac{a}{a+d} = \text{squared partial correlation}$$

$$r^2_{alerting} = \frac{a}{a+b+c} = \text{ no direct parallel in multiple regression at the individual level .}$$

CONCLUSIONS

Effect sizes are becoming increasingly integrated into and expected as part of statistical analysis in the social sciences. Many effect sizes are available to researchers, but many discussions of effect sizes are long on computation and rather short on interpretation. Consequently, researchers and reviewers could be forgiven for being somewhat unclear about the exact meaning of a given effect size or of the difference between effect sizes. This article is intended to help articulate and clarify the meaning of three correlational effect sizes associated with contrast analysis.

Because contrast analysis and the effect sizes discussed in this article are but a piece of the large puzzle of understanding one's data and evaluating one's theories, a few broader issues merit comment. First, recent recommendations (e.g., Cohen, 1990; Loftus, 1996; Wilkinson et al., 1999) highlight the usefulness of exploratory data analysis, and particularly the graphical presentation of data. Whether used to aid the detection of problems in the data, such as outliers; to potentially uncover unforeseen differences or trends, such as quadratic associations; or simply to present data in a clear and efficient modality, graphical examination of data might pre-

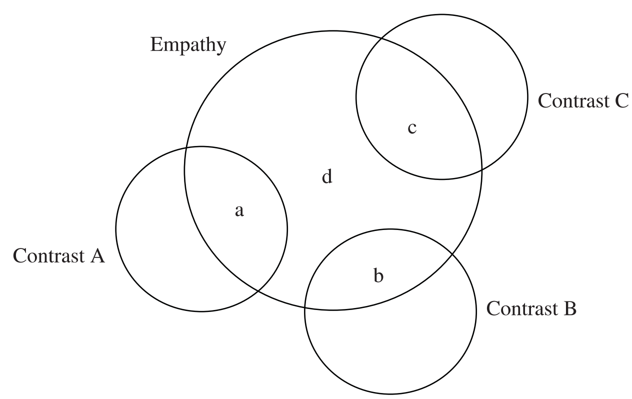

FIGURE 1 Venn diagram for effect sizes in contrast analysis.

cede any quantitative examinations of data. Behrens (1997) presented an excellent discussion of many useful graphical strategies and exploratory data analysis in general.

Recent recommendations (e.g., Cohen, 1990; Loftus, 1996; Wilkinson et al. 1999) also highlight a second broad statistical consideration—the importance of confidence intervals. Even among those who question the logic or utility of significance testing as it has typically been practiced, confidence intervals are recognized as a useful complement to effect sizes. The effect sizes discussed in this article are no exception. To compute the confidence intervals around $r_{effect\ size}$ and $r_{contrast}$, one uses the typical procedures for computing confidence intervals around correlations, as discussed in many basic statistics textbooks (e.g., Howell, 2002, p. 279). Conversely, because $r_{alerting}$ is computed with *groups* as the unit of observation, its confidence intervals are somewhat problematic. The degrees of freedom used to calculate a correlation's confidence interval are based on the number of units of observation and most $r_{alerting}$ effect sizes will have few degrees of freedom and thus have large confidence intervals. For example, the confidence interval for the $r_{alerting}$ associated with Contrast C ($r_{alerting}$ = .95) is $-.13 = \leq \rho_{alerting} \geq +1$.

A third broad statistical issue concerns alternative effect sizes. This article focuses on three correlational effect sizes of particular interest for contrast analysis, but many other effect sizes are available and potentially informative. For example, Thompson (1999a) distinguished between "uncorrected" and "corrected" effect sizes. All the effect sizes outlined in this article are "uncorrected," but others (e.g., adjusted R^2, ω^2) include adjustments that account for methodological issues such as sample size and number of variables. Similarly, Rosenthal (1994) distinguished between effect sizes derived from correlations (as are those included in this article) and those derived from Cohen's effect size *d* (the standardized mean difference between groups). Both sets of effect sizes can be transformed to and from each other, so the preference for one or the other may depend most on the analytic context and the way in which the researcher wishes to frame his or her findings.

A fourth broad issue concerns the use of contrast effect sizes in studies with different or more complex designs. The primary goal of this article is to provide a relatively intuitive presentation of the logic and meaning of three basic correlational effect sizes in contrast analysis, and the example is based on a fairly simple design. For different or more complex designs, the interpretation and logic of effect sizes remains the same. For example, all the effect sizes outlined in this article can be used for contrast analysis in repeated measures designs (e.g., Furr & Rosenthal, 2003b), although the calculations depend on the exact nature of the variables. In addition, the logic of the effect sizes outlined in this article are applicable to designs that include more factors or that do not have equal numbers of participants in each condition, although, again, there are several ways in which such values could be calculated (see Rosenthal et al., 2000).

Differentiating and Using the Effect Sizes

The effect sizes described in this article differ from each other in a number important ways, and researchers might wonder if some are more preferable than others. The short answer to this question is that the different effect sizes provide answers to different questions, so the appropriate effect sizes are the ones that answer the relevant analytic questions. In a given analytic situation, a researcher might choose to use several of the effect sizes, recognizing that each provides distinct and potentially important information. Nevertheless, the differences among the effect sizes are worth further consideration.

As outlined in this article, a researcher might consider reporting either unsquared or squared effect sizes (or both). If an analytic question seems to be intuitively framed in terms of "to what degree to the observed data match the predictions?," then unsquared effect sizes will be appropriate. For analytic questions more intuitively framed as "to what degree does the contrast account for variation in the data?," then the squared effect sizes will be more useful. In general though, the unsquared approaches might be preferable for at least two reasons. First, they allow researchers to retain information about the direction of the relation. In this article, analysis of Contrast A revealed that the data contradicted the hypothesis and this contradiction is reflected in the sign of the unsquared effect sizes, but is lost in the squared effect sizes. Thus, unsquared effect sizes provide direct information about both the magnitude and direction of association between predictions and observations. Second, the unsquared effect sizes avoid the interpretational ambiguity and criticisms leveled at squared effect sizes. For example, some researchers have argued that squaring puts the effect sizes on a nonintuitive and inappropriate metric of variation, that squaring is based on a statistical model that is often inappropriate for making conclusions about amount of determination, and that squaring often leads to an underestimation of the importance of effects (Abelson, 1985; D'Andrade & Dart, 1990; Ozer, 1985; Rosenthal & Rubin, 1979).

A second distinction arises when some data could be treated as irrelevant to a given contrast and thus be set aside. This distinction arises for contrasts that focus on only some participants or groups, such as Contrasts A and B in these examples; it does not arise for contrasts that include all groups, such as Contrast C. So, for contrasts in which some data could be set aside, which one is correct—the set-aside procedure or the full sample procedure? They are both reasonable and potentially useful, but they differ primarily in terms of which sample is of interest. If researchers wish to focus only on a subsample (e.g., only the differences between the Psychology and the Education majors), then the set-aside procedure might be more appropriate. This might particularly hold true when using a nonsquared effect size (e.g., $r_{effect\ size}$)—when the effect size is to be interpreted as the correlation between observed data and predicted data. In such cases, the predicted data really only involve a subset of participants and to include predicted data for the full sample might misrepresent the meaning of the contrast. Conversely, when using squared effect sizes (e.g., $r^2_{effect\,size}$), both the set-aside and the full-sample analysis could be quite meaningful. For many research situations, it may indeed be reasonable to report the degree to which a given contrast explains variation within a particular subsample and within the full sample. Note that the set-aside effect sizes will always be equal to, or more likely, greater than, full-sample effect sizes (see Table 5). Finally, if researchers are attempting to represent results with regard to the total sample, then the full-sample effect sizes are of course more appropriate than set-aside effect sizes. See Rosenthal et al. (2000, pp. 56–61, 102–112) for a variety of additional considerations about the set-aside strategies.

Aside from the issues of squaring and setting aside, how can the differences among the three basic effect sizes—$r_{effect\,size}$, $r_{alerting}$, and $r_{contrast}$—be summarized? Both $r_{effect\,size}$ and $r_{alerting}$ have relatively straightforward interpretations—they are zero-order correlations between observed data and predicted data. In addition, they both reflect the degree to which participants in one group (or groups) have a larger mean score than participants in another group (or groups). They differ only with regard to the level of aggregation. The $r_{effect\,size}$ analysis is at the level of the individual and thus includes *all participants*—so individual differences within a group (i.e., within-group variation or error variation) are included in the analysis. The $r_{alerting}$ analysis is at the level of the group and thus includes only group means—individual differences among the participants within groups are not included in the analysis. A consequence of this difference is that $r_{effect\,size}$ will always be smaller than $r_{alerting}$ (assuming that there is some degree of within-group variation).

Finally, $r_{contrast}$ is conducted at the individual level, but is a way of partialling out or controlling for variation associated with other contrasts. The distinction between $r_{contrast}$ and $r_{effect\,size}$ becomes more apparent when considering the fact that different research designs will likely include different factors, different numbers of factors, and different kinds of participants. The $r_{contrast}$ effect size describes a particular contrast in relation to error variation only, whereas $r_{effect\,size}$ describes a particular contrast in relation to all variation within a given study. Consider the possibility that two researchers each conduct a study of some phenomenon and obtain the same exact $r_{contrast}$. However one researcher included a larger number of factors in her study, which increases the total variation in the dependent variable. This would result in the two researchers obtaining different (perhaps drastically different) $r_{effect\,size}$ values. Thus, $r_{contrast}$ (and $r^2_{contrast}$) might be particularly useful for comparing contrast analyses across studies, and thus would be quite useful as a meta-analytic tool.

The "Size" of Effect Sizes

An additional issue involves concluding whether a given effect size value is "large" or of practical importance. Many researchers have acknowledged that statistical significance, size of effect, and importance of effect are separate issues (e.g., Rosnow & Rosenthal, 1989). Although Cohen (1988) provided guidelines for gauging the size of effects, a strict adherence to these guidelines is probably counterproductive (Thompson, 1999a). McCartney and Rosenthal (2000) outlined three considerations in interpreting effect size and practical importance. First, an effect should be considered in the context of methodological issues such as measurement and research design. For example, the measurement technologies in the social sciences are generally less precise than those in other sciences and this limits the effect sizes that we are likely to observe. Second, an effect should be considered in the context of relevant research literature. Is a given effect relatively larger or smaller than effects observed in similar research areas? Third, an effect can be presented as a binomial effect size display (BESD; Rosenthal & Rubin, 1982), which translates a given effect size into a 2 × 2 table of outcomes that is relatively intuitive, particularly for nonscientists (see also Rosenthal et al., 2000). The BESD, along with a cost-benefit analysis, can begin to reveal the potential practical importance of a given effect.

In sum, more and more researchers in the social sciences are recognizing the importance of reporting effect sizes. It is hoped that this article enhances the frequency and effectiveness with which effect sizes are used in social science research.

ACKNOWLEDGMENT

I thank Robert Rosenthal and Timothy Huelsman for their comments on an earlier version of this article.

REFERENCES

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin, 97*, 129–133.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2*, 131–160.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70*, 426–443.

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

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304–1312.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997–1003.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

D'Andrade, R., & Dart, J. (1990). The interpretation of r versus r^2 or why percentage of variance accounted for is a poor measure of size of effect. *Journal of Quantitative Anthropology, 2*, 47–59.

Furr, R. M., & Rosenthal, R. (2003a). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics, 2,* 45–67.

Furr, R. M., & Rosenthal, R. (2003b). Repeated-measure contrasts for "multiple-pattern" hypotheses. *Psychological Methods, 8,* 275–293.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Belmont, CA: Duxbury.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R., Donahue, B., et al. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68,* 350–386.

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

Loftus, G. R. (1996). Psychology will be a much better science when we change the way that we analyze data. *Current Directions in Psychological Science, 5*, 161–171.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development, 71*, 173–180.

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97*, 307–315.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Sage.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, England: Cambridge University Press.

Rosenthal, R., & Rubin, D. B. (1979). A note on percent of variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology, 9*, 395–396.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*, 166–169.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44*, 1276–1284.

Thompson, B. (1999a). If statistical significance tests are broke/misused, what practices should supplement or replace them? *Theory & Psychology, 9,* 165–181.

Thompson, B. (1999b). Why "encouraging" effect size reporting isn't working: The etiology of researcher resistance to changing practices. *Journal of Psychology, 133*, 133–140.

Vacha-Haase, T., Nilsson, J. E., Reetz, D. R., Lance, T. S., & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory & Psychology, 10*, 413–425.

Wilkinson, L., & the APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604.