Practical Guide for Reporting Effect Size in Quantitative Research in the *Journal of Counseling & Development*

Jerry Trusty, Bruce Thompson, and John V. Petrocelli

The purpose of this article is to assist researchers in meeting the requirement of reporting effect sizes in quantitative research studies submitted to the Journal of Counseling & Development. *This requirement is detailed in the "Guidelines for Authors" included in this issue. The authors provide practical information on generating, reporting, and interpreting effect size estimates for various types of statistical analyses. Information is provided on the meaning of effect sizes within the larger knowledge base.*

Whereas statistical significance tests assess the reliability of the relationship between independent and dependent variables, effect sizes assess the strength of the relationship. When reported and interpreted appropriately, effect size estimates are practical, straightforward, and relevant to research questions and hypotheses (Wilkinson & APA Task Force on Statistical Inference, 1999). Effect sizes can serve as an important mechanism for communicating our professional knowledge base to others within and outside our profession, including our clients (Thompson, 2002a). Many researchers who publish articles in the *Journal of Counseling & Development (JCD)* are doctoral-level professionals. Many *JCD* readers, however, are master's-degree level professional counselors. Therefore, researchers should (a) describe the significance of their findings in relation to counseling practice and (b) present findings in ways that are readily understood by counselors. Effectively providing effect size information is an important means for accomplishing both these goals (see American Psychological Association, 2001, pp. 5, 25–26).

For these various reasons, 23 journals currently require effect size reporting (cf. Harris, 2003; Snyder, 2000). Indeed, as Fidler (2002) recently noted, "Of the major American associations, only all the journals of the American Educational Research Association have remained silent on all these issues" (p. 754).

The purpose of the present article is to review briefly some of the frequently reported effect size choices, how these effect indices can be obtained from computer software, and how they are reported and interpreted. Only a few of the several dozen effect size choices (Huberty, 2002; Kirk, 1996) are summarized here. Readers seeking more detail may consult Snyder and Lawson (1993) or Rosenthal (1994). More contemporary treatments are provided by Kirk (in press), Hill and Thompson (in press), and Thompson (in press). The book by Kline (in press) is a particularly excellent resource.

HOW TO GENERATE, REPORT, AND INTERPRET EFFECT SIZE

Kieffer, Reese, and Thompson (2001) and Thompson and Snyder (1998) investigated the research and statistical techniques used in counseling journals during the past 15 years. These authors reported that of the various statistical procedures (beyond basic descriptive statistics), forms of analysis of variance (ANOVA; e.g., multivariate analysis of variance [MANOVA] and multivariate analysis of covariance [MANCOVA]) were the most frequently used statistical procedures. Also, only a small percentage of *JCD* authors reported effect sizes (Thompson & Snyder, 1998). Therefore, we cover ANOVA and related techniques most thoroughly in this article. Bangert and Baumberger (in press) and Petrocelli (2003) reported that regression analyses and chi-square analyses were also relatively common in *JCD* throughout the last decade. Therefore, we provide guidelines regarding effect sizes for these two types of statistical analysis. Bangert and Baumberger did find a higher incidence of effect size reporting in regression analyses, in the form of reporting multiple correlation squared (R^2).

Because SPSS seems to be the statistical package most commonly used by counseling researchers, all practices in

Jerry Trusty, *Department of Counselor Education, Counseling Psychology and Rehabilitation Services, Pennsylvania State University;* ***Bruce Thompson,*** *Department of Educational Psychology, Texas A&M University;* ***John V. Petrocelli,*** *Department of Psychology, Indiana University. Correspondence concerning this article should be addressed to Jerry Trusty, 327 CEDAR Building, Penn State University, University Park, PA 16802-3110 (e-mail: jgt3@psu.edu).*

this section are described using SPSS Version 11.0. Procedures in other versions of SPSS and other computer software programs (e.g., SAS, Minitab) are similar.

In all instances of reporting strength of association or mean differences, researchers should use the terminology *effect size* in their writing. This language alerts the reader to the fact that an estimate of practical significance is being reported. Otherwise, some readers may not recognize particular statistics as being estimates of effect size.

Effect Size in Analysis of Variance and Related Techniques

For ANOVA techniques, SPSS provides estimates of effect size in the form of eta^2 (η^2). Eta^2 is an estimate of the proportion of variability in the dependent variable(s) explained, or accounted for, by membership in the groups defining the independent variable(s). Eta^2 is equivalent to the R^2 of a dummy-coded variable in regression (Pedhazur, 1982). Eta^2 estimates of effect size are not automatically provided in SPSS; that is, the researcher must specify that this estimate of effect size is desired.

In SPSS, eta^2 is available through *Compare Means* procedures, but only in the *Means* choice (click on *Analyze, Compare Means, Means*), and not in the *T-Test* choices or the *One-Way ANOVA* choice However, eta^2 is consistently available through the *General Linear Model* procedures. Therefore, if you desire to perform a one-way ANOVA, for example, use *General Linear Model* instead of the *One-Way ANOVA* procedure accessed through *Compare Means* in the *Analyze* pull-down menu. Otherwise, these and other effect sizes may also be computed by hand (Snyder & Lawson, 1993).

From the *Analyze* pull-down menu in SPSS, select *General Linear Model*. Then select either *Univariate*, *Multivariate*, or *Repeated Measures*, depending on your design. After you specify your dependent variable(s) and independent variables(s)—termed *Fixed Factor(s)* in the box—click on the *Options* button. One choice in the *Display* box is *Estimates of Effect Size*. By clicking (checking) on that option, you will get *partial eta^2* estimates of effect size in your output file. The partial eta^2 is an estimate of the amount of variance in the dependent variable(s) attributable to the particular effect of interest, and these statistics should be reported as partial eta^2 statistics (see Tabachnick & Fidell, 2001, pp. 52–53). Partial eta^2 estimates are displayed for each effect in the analysis (i.e., multivariate effects, main effects, interaction effects, effects of covariates). In the case of multiple dependent variables (e.g., MANOVA), partial eta^2 estimates are displayed for each dependent variable. For researchers who use SPSS syntax, the subcommand line is /PRINT = ETASQ, or alternatively, /PRINT = EFSIZE. Whether through the pull-down menu or through syntax, this procedure for obtaining eta^2 estimates is available for all types of analysis of variance designs, including one-way univariate and multivariate designs, factorial designs, and repeated measures designs.

Eta^2 estimates are referred to as *variance-accounted-for* statistics. The interpretation of a partial eta^2 value of .08 for a particular independent variable, for example, would be "Regarding effect size, 8% of the variability (or differences) in the dependent variable scores was explained or predicted with knowledge of group membership on the independent variable." The complement of the Wilks's lambda statistic (1 – Wilks's lambda) has been used to indicate variance-accounted-for in multivariate tests. However, this practice may produce inflated estimates of explained variance, whereas the partial eta^2 produces more reasonable estimates (see Tabachnick & Fidell, 2001, pp. 338–339).

Researchers may choose to use the SPSS *MANOVA command* for performing multivariate analyses. This command is available only through SPSS syntax and not through the SPSS pull-down menus. If the MANOVA command is used, researchers can use the squared canonical correlation as an overall estimate of variance-accounted-for effect size. The subcommand for obtaining the canonical correlation is / PRINT = SIGNIF(EIGEN). Researchers can use structure coefficients (SPSS subcommand: / DISCRIM = CORR) to gauge the relative contributions of particular dependent-variable differences to overall effects.

There are situations in which *standardized-difference* effect sizes are needed in the reporting of results. This is particularly the case in experimental or quasi-experimental designs when the mean difference between experimental and control groups is of interest. Standardized-difference effect size statistics such as Cohen's d can easily be calculated by hand, and researchers, by requesting descriptive statistics in analyses, can get all the statistics they need from the SPSS output (e.g., group means, standard deviations) to calculate Cohen's d and similar statistics. Thompson (2002a) provided formulas for calculating Cohen's d and for converting standardized-difference effect sizes into variance-accounted-for effect sizes, and Olejnik and Algina (2000) provided information on effect size calculations in several analysis of variance designs.

Effect Size in Regression Procedures

As stated earlier, effect sizes are commonly reported in regression analyses in the form of R^2, which is another variance-accounted-for effect size. Similar to the interpretation of eta^2 and the squared canonical correlation, an R^2 of .32, for example, would be interpreted, "Regarding effect size, 32% of the variability in the dependent variable was explained, or accounted for, by the independent variables." R^2 is automatically included in linear regression output in SPSS, and variations and extensions of regression (e.g., ordinal regression, logistic regression, structural equation modeling) include specific types of R^2 statistics.

One shortcoming of an overall R^2 is that effects are not revealed for particular independent or predictor variables. There are various approaches for gauging the effects of particular predictor variables. One method is to report standardized regression coefficients (standardized betas, path coefficients). Another is to report the change in the R^2 when particular variables are entered hierarchically into regression equations (incremental increase in R^2). Yet another is to remove a predictor from an equation and note the incremental decrease in the R^2. Structure coefficients have also been highly recommended (Courville & Thompson, 2001). Researchers should be aware of caveats regarding the use of standardized betas, increments, and structure coefficients (see Courville & Thompson, 2001; Pedhazur, 1982).

Multicollinearity and suppressor effects in regression equations can have considerable influences on results (see Cohen & Cohen, 1975; Tabachnick & Fidell, 2001), and therefore researchers should be alert for these situations and interpret results accordingly. In addition, it is advisable to report the adjusted R^2 in regression analyses (see Thompson, 2002a).

One purpose of reporting effect sizes is to provide future researchers with a consistent gauge of the strengths of associations. A future researcher who reads your article may be interested only in the relationship between one pair of variables in your analysis; therefore, inclusion of a bivariate (zero-order) correlation matrix along with your regression analysis will likely contribute to the knowledge base (see Pedhazur, 1982). Doing so will be especially useful to researchers conducting meta-analyses; otherwise, future researchers may choose to disregard your study simply because there is no reliable way to estimate the desired effect size (see Rosenthal & DiMatteo, 2001).

Effect Size in Chi-Square Procedures

As in ANOVA or regression procedures, there are numerous indicators of strength of association in chi-square analyses, and there is no one indicator that accounts for every type of association (Norusis, 1993). The use of one indicator or the other depends on the scaling properties of the variables involved. Also, some indices are symmetric (no variable is designated as the dependent variable) or asymmetric (one or the other variable is designated as the dependent variable), and some indices are symmetric only.

Chi-square tests are accessed through the *Crosstabs* choice in the *Descriptive Statistics* procedure in the *Analyze* pull-down menu (click on *Analyze, Descriptive Statistics, Crosstabs*). SPSS provides numerous chi-square measures of association including Phi (ϕ), Cramer's V, Kendall's tau-b (τ_b) and tau-c (τ_c), Goodman and Kruskal's gamma (Γ), Cohen's kappa (κ), Goodman and Kruskal's lambda (λ), Goodman and Kruskal's tau (τ), and Somers's *d*. Click on the *Statistics* button to access these and other statistics. In SPSS output, these various statistics are presented in terms of the scaling of the variables (e.g., nominal by nominal) and other criteria (e.g., symmetric or asymmetric). Researchers decide, based on variable scaling and the design and purposes of the research, which statistic is most appropriate for reporting effect size.

Interpretations of chi-square effect sizes are dependent on the specific statistic used. For example, Goodman and Kruskal's lambda is appropriate when both variables have nominal (categorical) scaling. Lambda represents the degree of reduction in the error of predicting the values (categories) of one variable based on the values (categories) of another variable. SPSS output provides lambda statistics for the symmetric case (no dependent variable) and both asymmetric cases (each variable as the dependent variable). Researchers may benefit from using an SPSS *Base-System User's Guide* (e.g., Norusis, 1993), a *Base Applications Guide* (e.g., SPSS, 1999), and the *Help* menus in SPSS as resources for selecting and interpreting estimates of effect sizes in chi-square analyses.

ATTRIBUTING FURTHER MEANING TO EFFECT SIZE

Basic reports and interpretations of effect sizes are only one part of the research task. The manuscript is incomplete unless these effects are evaluated in the context of the study and in the larger context of knowledge. It is common for researchers to use Jacob Cohen's benchmarks for *small*, *medium*, and *large* effect to evaluate their effect size results. However, Cohen did not intend for these benchmarks to be used with the rigidity that some researchers apply (Thompson, 2002a, 2002b).

Small effect sizes for very important outcomes can be extremely important, as long as they are replicable. For example, the variance-accounted-for effect size for "not smoking" on cancer incidence is only around 2%, yet, as Gage (1978) pointed out,

> Sometimes even very weak relationships can be important. . . . [O]n the basis of such correlations, important public health policy has been made and millions of people have changed strong habits. (p. 21)

Conversely, large effect sizes may be trivial if they involve trivial outcomes. Spuriously large effect sizes can result from method variance and response biases (Thorndike, 1997) and from specification errors and outliers (Pedhazur, 1982). Misleadingly small effect sizes can result from measurement error and skewed distributions of variables (Tabachnick & Fidell, 2001). Skewed distributions of categorical variables can especially limit the magnitude of variance-accounted-for effect sizes (for an example, see Magidson, 1993, p. 139).

Effect size interpretation should be informed by an explicit factual comparison of detected effects with the effects reported in the related prior literature (Thompson, 2002b). However, interpretation also necessarily involves some subjective value judgment on the part of the researcher. The key is to be explicit in providing the reasoning underlying these judgments. As Huberty and Morris (1988, p. 573) noted, "As in all of statistical inference, subjective judgment cannot be avoided. Neither can reasonableness!"

REFERENCES

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

Bangert, A. W., & Baumberger, J. (in press). Research and statistical techniques used in the *Journal of Counseling & Development*, 1990–2001. *Journal of Counseling & Development*.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement*, *61*, 229–248.

Fidler, F. (2002). The fifth edition of the APA *Publication Manual:* Why its statistics recommendations are so controversial. *Educational and Psychological Measurement, 62,* 749–770.

Gage, N. L. (1978). *The scientific basis of the art of teaching.* New York: Teachers College Press.

Harris, K. (2003). Instructions for authors, *Journal of Educational Psychology. Journal of Educational Psychology, 95,* 201.

Hill, C. R., & Thompson, B. (in press). Computing and interpreting effect sizes. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 19). New York: Kluwer.

Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement, 62,* 227–240.

Huberty, C. J., & Morris, J. D. (1988). A single contrast test procedure. *Educational and Psychological Measurement, 48,* 567–578.

Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in *AERJ* and *JCP* articles from 1988 to 1997: A methodological review. *Journal of Experimental Education, 69,* 280–309.

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

Kirk, R. E. (in press). The importance of effect magnitude. In S. F. Davis (Ed.), *Handbook of research methods in experimental psychology.* Oxford, United Kingdom: Blackwell.

Kline, R. (in press). *Beyond significance testing.* Washington, DC: American Psychological Association.

Magidson, J. (1993). *SPSS for Windows CHAID release 6.0.* Chicago: SPSS Inc.

Norusis, M. J. (1993). *SPSS for Windows base system user's guide release 6.0.* Chicago: SPSS Inc.

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

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). Fort Worth, TX: Holt, Rinehart and Winston.

Petrocelli, J. V. (2003). Hierarchical multiple regression in counseling research: Common problems and possible remedies. *Measurement and Evaluation in Counseling and Development, 36,* 9–22.

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: Russell Sage Foundation.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52,* 59–82.

Snyder, P. (2000). Guidelines for reporting results of group quantitative investigations. *Journal of Early Intervention, 23,* 145–150.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education, 61,* 334–349.

SPSS. (1999). *SPSS base 10.0 applications guide.* Chicago: Author.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Thompson, B. (2002a). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling & Development, 80,* 64–71.

Thompson, B. (2002b). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(3), 24–31.

Thompson, B. (in press). Research synthesis: Effect size. In G. Camilli, P. B. Elmore, & J. Green (Eds.), *Complementary methods for research in education.* Washington, DC: American Educational Research Association.

Thompson, B., & Snyder, P. A. (1998). Statistical significance and reliability analyses in recent *JCD* research articles. *Journal of Counseling & Development, 76,* 436–441.

Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604. (Reprint available through the APA Home Page: http://www.apa.org/journals/amp/amp548594.html)