DOCUMENT RESUME

ED 471 293 TM 034 670

AUTHOR	Tapia, Martha; Marsh, George E., II
TITLE	Interpreting and Reporting Effect Sizes in Research Investigations.
PUB DATE	2002-11-00
NOTE	14p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Chattanooga, TN, November 6-8, 2002).
PUB TYPE	Information Analyses (070) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Effect Size; *Research Reports; *Statistical Significance
IDENTIFIERS	*American Psychological Association

ABSTRACT

Since 1994, the American Psychological Association (APA) has advocated the inclusion of effect size indices in reporting research to elucidate the statistical significance of studies based on sample size. In 2001, the fifth edition of the APA "Publication Manual" stressed the importance of including an index of effect size to clarify research reports. Many journal now require authors to include effect-size statistics, but there is little guidance for researchers to indicate the effect size index they should use for univariate and multivariate research designs. This paper reviews the methods suggested for reporting effect size, noting similarities and differences among them. A table summarizes the formulas for 16 common effect size statistics. (Contains 20 references.) (Author/SLD)

Running head: Effect Sizes

INTERPRETING AND REPORTING EFFECT SIZES IN RESEARCH INVESTIGATIONS

Martha Tapia

Berry College

and

George E. Marsh II

The University of Alabama

Paper presented at the Annual Meeting of the
Mid-South Educational Research Association,
Chattanooga, Tennessee
November 6-8, 2002

ABSTRACT

Since 1994, the American Psychological Association has advocated the inclusion of effect-size reporting in research to elucidate the statistical significance of studies based on sample size. In 1999, the APA Task Force on Statistical Inference emphasized that effect sizes should always be reported along with p values. In 2001, the 5th edition of the APA Publication Manual stressed the importance of including an index of effect size to clarify how much difference exists. As a result, many research journals require authors to include effect-size statistics. While researchers will comply and follow editorial leadership in this regard, there is little guidance for investigators in which statistics they should use to report effect size for univariate and multivariate research designs. This paper is an attempt to review the variety of methods suggested for reporting effect size.

Interpreting and Reporting Effect Sizes in Research Investigations

Introduction

Statistical significance tests have been essential for social science and educational research for the past 70 years, but they have been criticized for dependency on sample size. In an effort to correct for this limitation, the 1994 edition of the American Psychological Association (APA) publication manual encouraged the use of effect-size reporting and many journals now require it. In 1999, the APA Task Force on Statistical Inference emphasized that effect sizes should always be reported along with p values. Subsequently, in 2001, the 5th edition of the APA Publication Manual stressed the importance of including an index of effect size. Although the concept of effect size has existed for many years, it remains perplexing to investigators and reports of effect sizes remains infrequent. Effect size is the difference between the null and alternative hypotheses, and can be measured either using raw or standardized values. At issue is the probability of getting a statistically significant result if there is a real effect in the population under examination. If a test is not significant, it is important to know if this is because there is no effect or because the research design did not detect it.

Recent research of the literature has revealed at least 61 different effect-size statistics (Elmore, 2001). Due to the large number, selection and appropriate interpretation of effect-size statistics is problematic. Few software programs contain automatic methods for their determination The purpose of this paper is to provide an overview of formulas for computing corrected and uncorrected effect-size statistics, and review suggested guidelines for their uses in data analysis and reporting for univariate and multivariate studies.

Effect Sizes

There are two major classes of effect sizes and a third, "miscellaneous" category described by Kirk (1996)): (a) variance-accounted-for effect sizes, and b) standardized mean differences effect sizes. Variance-accounted-for effect sizes (VAFE) can always be computed since all parametric analyses are correlational (Knapp, 1978; Thompson, 1984, 1991). Effect sizes in this class include indices such as r^2, R^2, and η^2. A VAFE size is the ratio of explained variance to total variance. For example, it can be obtained by dividing the sum-of-squares for an effect by the sum-of-squares total.

In ANOVA, the resulting effect size is called eta squared (η^2). In multiple regression, the resulting effect size is called the squared multiple correlation (R^2). The formula in either case is

$$R^2 = \eta^2 = \frac{SOS_{Explained}}{SOS_{Total}}$$

Variance-accounted for effect sizes can range from 0 to 1. Hence the amount of variance accounted for by the independent variable, namely $SOS_{Explained}$, can explain a range of variance from none to the total exhibited. Hence, effect size represents the percentage of the total variance explained by the independent variable.

The other class is the standardized difference effect size, representing the mean differences in units of common population standard deviations. Standardized difference effect sizes vary in how they can be used to estimate the standard deviation for the population. Effect sizes in this class include indices such as Cohen's *d*, Glass'Δ, and Hedges' *g*.

Cohen's *d* is the most common example of a standardized effect-size statistic. It uses all the variance across the groups (SD_{pooled}) because it is based on a larger N. The formula is

$$d = \frac{(M_{experimental} - M_{control})}{SD_{pooled}}$$

Another example of standardized difference effect size is Glass's Δ, which uses the SD of only the control group as an estimate of the SD of the population. This statistic expresses effect in standard deviation units and can be positive or negative and not bounded by 1 or 0 as the variance-accounted for effect size. It is exactly equivalent to a Z-score of the standard Normal distribution. Hence, it can be converted into statements about overlap between the two samples in terms of a comparison of percentiles.

The existence of two different metrics with different ranges of values complicates interpretation of effect sizes. However, effect sizes in these two classes can be transformed into metrics of the other. For example, Cohen's *d* can be converted to an *r* (Cohen, 1988):

$$r = \quad d / [(d^2 + 4)^{.5}]$$

When total size is small or group sizes are disparate, the following formula can be used (Aaron, Kromrey & Ferron, 1998):

$$r = \quad d / [(d^2 + [(N^2 - 2N)/(n_1\ n_2\ ^4)^{.5}]$$

Also, an *r* can be converted to a *d* (Friedman, 1968):

$$d = \quad [2r] / [(1 - r^2)^{.5}].$$

Interpretation of the magnitude was recommended by Cohen (1988), who cautiously characterized effects as "small," "medium," and "large" for *d* and r^2. A small effect size (d = .2) is less than a medium effect (d = .5) and this is less than a large effect (d = .8). Cohen interprets a medium effect as one that is visible to the naked eye of a careful observer. A small effect size, although noticeable, is not so small as to be trivial. Table 1 shows a summary of Cohen's interpretations.

Table 1. Interpretation of Effect Size

Characterization	*d*	r^2
"low"	0.2	1.0%
"medium"	0.5	5.9%
"large"	0.8	13.8%

Although these standards are commonly used when reporting effect sizes, Huck (2000) suggests establishing standards based on the raw units of the instrument used as a dependent variable. For example, if the task is the completion of a math drill, the researcher might determine that a nontrivial effect should consist of a ten-second drop in time between the experimental and control group. This "standard" should be based on the standard deviation of the population from which the inference is drawn. If one standard deviation for this math test is 20 seconds, it might be argued that half a standard deviation difference (10 seconds) is visible to the naked eye and could be regarded as a medium effect. The appraisal of effect sizes inherently requires the researcher to introduce personal value judgments about the practical or clinical importance of effects. As Baugh and Thompson (2001) stress, even small effect estimates may be important when the outcomes are critical, such as in life-or-death matters.

In addition to standardized difference and VAFE sizes, there are "uncorrected" and "corrected" effect sizes. The theory of "ordinary least squares" used in "classical statistical methods, such as ANOVA and regression, tend to capitalize on all the variance present in the observed sample scores. This variance includes the "sampling error variance" that is unique to the sample under study. Hence, the VAFE sizes, such as η^2 and R^2, which use the variance, tend to overestimate the effects that would be replicated in the population or in future samples.

The extent of overestimation or positive bias in the sample VAFE size estimate can be corrected. The corrected effect size is obtained by removing the estimated sampling error variance. Corrected estimates are always less than or equal to uncorrected estimates. The corrected VAFE sizes include indices such as adjusted R^2, Hays's ω^2, and Herzberg's R^2. For standardized mean difference effect size, a corrected effect size is Thompson's "corrected"*d*. There is more sampling error variance when (a) sample sizes are smaller, (b) the number of observed variables is larger, or (c) the population effect is smaller. Hence, it is better to use corrected effect-size statistics if any one of the following is true:

- F, t, or R^2 values are just above the critical level for statistical significance
- N is small
- An initial calculation of an uncorrected effect-size statistics suggests that the effect size is small

Snyder and Lawson (1993) also suggest use of corrected effect sizes when the ratio of participants to dependent variables is less than 5:1.

Examples of uncorrected effect-size measures are η^2, R^2, Cohen's d, and Glass's Δ. Some corrected effect measures are adjusted R^2, Hays's ω^2, ε^2, and the Ezekiel formula (Thompson, 2002). Selecting the appropriate effect-size measure among so many options is complex, not only because of the range of available choices, but also because there is a lack of common agreement in the field (Thompson, 1999; Snyder & Thompson, 1998; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000). Journal editors apparently welcome any choice of statistics that can be substantiated with reason. Although present circumstances are inconclusive, selection of an appropriate statistic can be made by determining that it is in concordance with the statistical analyses of the data.

The choice of effect size measure should depend primarily upon the researcher's intention to generalize results to other samples or to the population. If a researcher wants to use results from a previous sample to generalize to future samples, then examples of effect-size measures to use are η^2, partial η^2, Herzberg and Lord formulas. Examples of effect-size measures designed for developing population expectations are adjusted R^2, Hays's ω^2, and the Interclass correlation ρ_{I}.

Although ANOVA can be considered a special case of regression analysis (Cohen, 1968), different statistics are used with each analysis. Effect-size measures used in ANOVA are η^2, partial η^2, ε^2, Hays's ω^2, and Cohen's d. For regression analysis some effect-size measures are R^2, adjusted R^2, ε^2, and the Ezekiel formula.

Fixed designs and random-effect design have also different effect-size measures associated with them. Fixed models assume that levels in factors are fixed in an ANOVA design or the values of the predictor factors are fixed in a regression mode. That is, either all the levels of independent variables are used or the researcher wants to generalize to the levels actually used in the study. A replication would need to use the same levels. In a random effect design, the researcher randomly selects the levels of the independent values to be used. Generalizations can be made to other levels, and replication studies could use other randomly selected levels. While the Ezekiel formula and ε^2.are exclusively used for fixed designs, the Herzberg formula and Hays's ω^2 have alternative formulas for fixed or random effects.

A univariate design examines the relationship between one or more independent variables and a single dependent variable. A multivariate design examines multiple dependent variables. Canonical correlations and multivariate analysis of variance (MANOVA) are examples of multivariate techniques. Different effect-size statistics are used for univariate and multivariate analyses. Effect-size statistics used for univariate analyses are η^2, partial η^2, ε^2, Hays's ω^2, R^2, Ezekiel formula, and Cohen's d. For multivariate analyses the effect-size statistics to use are D^2 and $1-\lambda$ (Stevens, 1992). Table 2 contains formulas for common effect-size statistics.

Conclusions

The incorrect interpretation of statistical significance has stimulated a movement to report results that include effect size for significant and nonsignificant results. It is assumed that use of effect size can avoid interpretations that may be erroneously applied to the general population. In other words, reports of a significant difference should be clarified with the size of the difference. This review has provided a survey of various methods that have been recommended.

There is no common agreement about the statistical methods and disagreement about the interpretations of various effect-size statistics that may be used. Whenever possible, there should be an objective method of determination. While overstating the significance of research results can be ameliorated with effect-size reports, there needs to be further research and clear strategies for effect-size reporting, perhaps by disciplines, and especially in fields and research topics where interpretation of effect size relies upon subjective interpretation. In such cases, the researcher should provide a clear rationale for the approach.

Table 2. Formulas of Common Effect Size Statistics.

Cohen's *d*	$(M_{experimental} - M_{control}) / SD_{pooled}$
Glass's Δ	$(M_{experimental} - M_{control}) / SD_{control}$
Hedges's g	$(M_{experimental} - M_{control}) / SD_{pooled}$
Eta squared, η^2	$SS_{exffect} / SS_{total}$
Partial eta squared, η_p^2	$SS_{exffect} / (SS_{exffect} + SS_{exrror})$
Epsilon squared, ε^2	$(SS_{exffect} - (df_{effect})(MS_{exrror})) / SS_{total}$
Omega squared fixed, ω^2	$(SS_{exffect} - (df_{effect})(MS_{exrror})) / (MS_{error} + SS_{total})$
Omega squared random,	$(MS_{exffect} - MS_{exrror}) / (MS_{effect} + (df_{total})(MS_{exrror}))$
Interclass correlation, ρ_I	$(MS_{exffect} - MS_{error}) / (MS_{exffect} + (df_{effect})(MS_{exrror})$
R^2	$SS_{exffect} / SS_{total}$
adjusted R^2	$R^2 - ((1 - R^2) * (k / (n - k - 1)))$
Herzberg fixed	$1 - ((n - 1) / (n - k - 1)) * (1 - R^2)$
Herzberg random	$1 - ((n - 1)/ (n - k - 1)) * ((n - 2)/ (n - k - 2)) * ((n + 1)/n) * (1 - R^2)$
Ezekiel	$1 - ((n - 1) / (n - k - 1)) * (1 - R^2)$
Lord	$1 - (1 - R^2) * ((n + k + 1) / (n - k - 1))$
Mahalanobis D^2	$4F((N - 2) / N) * (df_1 / df_2)$

References

Aaron, B., Kromrey, J.D. & Ferron, J.M. (1998). *Equating r-based and d-based effect size indices: Problems with a commonly recommended formula.* Paper presented at the annual meeting of the Florida Educational Research Association, Orlando, FL. (ERIC Document Reproduction Service No. ED 433 353)

American Psychological Association (1994). *Publication manual of the American Psychological* (4th ed.). Washington, D.C.: Author.

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

Baugh, F., & Thompson, B. (2001). Using effect sizes in social science research: New APA and journals mandates for improved methodology practices. *Journal of Research in Education, 11*(1), 120-129.

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: Erlbaum.

Elmore, P. (2001). *A primer on basic effect size concepts.* Paper presented at the American Educational Research Association, Seattle. (ERIC Document Reproduction Service No. ED 453260)

Friedman, H. (1968). Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin, 70,* 245-251.

Huck, S. W. (2000). *Reading statistics and research.* New York: Longman.

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

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin*, *85*, 410-416.

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

Snyder, P. A., & Thompson, B. (1998). Use of tests of statistical significance and other analytic choices in a school psychology journal: Review of practices and suggested alternatives. *School Psychology Quarterly*, *13*, 335-348.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation*. Newbury Park, CA: SAGE.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. *Measurement and Evaluation in Counseling and Development*, *24*(2), 80-95.

Thompson, B. (1999). *Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap*. Paper presented at the Invited address presented at the Annual meeting of the American Educational Research Association, Montreal. (ERIC Document Reproduction Service No. ED 429110).

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

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 and Psychology*, *10*, 413-425.

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

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

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM034670

I. DOCUMENT IDENTIFICATION:

Title: INTERPRETING and Reporting Effect Sizes in Research Investigations	
Author(s): MARTHA TAPIA and George E. Marsh II	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

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

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

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

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

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

Sign here, → please

Signature: Martha Tapia	Printed Name/Position/Title: MARTHA TAPIA / PROF / Dr.	
Organization/Address: Berry College / P.O. BOX 495014 MT. Berry, GA, 30149-5014	Telephone: 706-290-2662	FAX: 706-238-7849
	E-Mail Address:	Date:

(Over)

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

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

Publisher/Distributor:
Address:
Price:

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

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

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

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

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org