

Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research

TIMOTHY R. LEVINE

Michigan State University

CRAIG R. HULLETT

University of Wisconsin-Madison

Communication researchers, along with social scientists from a variety of disciplines, are increasingly recognizing the importance of reporting effect sizes to augment significance tests. Serious errors in the reporting of effect sizes, however, have appeared in recently published articles. This article calls for accurate reporting of estimates of effect size. Eta squared (η^2) is the most commonly reported estimate of effect sized for the ANOVA. The classical formulation of eta squared (Pearson, 1911; Fisher, 1928) is distinguished from the lesser known partial eta squared (Cohen, 1973), and a mislabeling problem in the statistical software SPSS (1998) is identified. What SPSS reports as eta squared is really partial eta squared. Hence, researchers obtaining estimates of eta squared from SPSS are at risk of reporting incorrect values. Several simulations are reported to demonstrate critical issues. The strengths and limitations of several estimates of effect size used in ANOVA are discussed, as are the implications of the reporting errors. A list of suggestions for researchers is then offered.

Null hypothesis testing with standard tests of statistical significance have long been the decision rules of choice in most quantitative communication research. There is, however, a growing recognition of the limitations associated with significance testing and p -values as the sole criterion for interpreting the meaning of results (e.g., see Boster, this issue). As a consequence, many communication journals have adopted editorial policies that require estimates of effect sizes and statistical power be reported in addition to significance tests. For example, the current editorial policies of both *Communication Monographs* (CM) and *Human Communication Research* (HCR) require authors to report the effect sizes for statistical tests. As we argue here, there are good reasons to report estimates of effect size in addition to p -values.¹

Timothy R. Levine (Ph.D., Michigan State University, 1992) is a Professor in the Department of Communication at Michigan State University. *Craig R. Hullett* (Ph.D., Michigan State University, 2000) is an Assistant Professor in the Department of Communication Arts at the University of Wisconsin-Madison. Electronic correspondence can be directed to the first author at levinet@msu.edu.

The analysis of variance (ANOVA) is one of the most frequently used statistical analyses in quantitative communication research. Although a number of different estimates of effect size are available when using ANOVA (e.g., omega squared, epsilon squared, eta squared; Keppel, 1982), eta squared seems to be, by far, the most frequently reported. For example, a brief perusal of recent issues of *CM* (Volume 66, No. 3) and *HCR* (Volume 25, No. 3) showed that ANOVA was used in 6 out of the 10 articles reporting statistical analyses of data. Eta squared was reported as the estimate of effect size in five out of the six of these cases (the sixth case did not report effect sizes at all). Thus, eta squared is currently a frequently reported and important descriptive statistic in communication research.

The current authors have become aware of published errors connected with the reporting of eta squared. Specifically, impossibly large estimates of eta squared have been appearing in submitted and published reports of communication research. Although we do not know how widespread these errors are, we suspect they are quite common and most often are likely to go unnoticed. We further believe that these errors may be attributable to a reporting error that we have discovered on the statistical software SPSS for Windows printouts. Values labeled as eta squared on (at least some) SPSS printouts are really partial eta squared (Cohen, 1973), and consequently, researchers may often be unknowingly reporting partial eta squared values as if they were eta squared. Because partial eta squared values may, in some cases, be widely discrepant from the values of omega squared, epsilon squared, and eta squared, these reporting errors may lead to serious substantive errors in the interpretation of results. For these reasons, a closer look at eta squared and partial eta squared is warranted.

To further compound the problem, most popular statistics texts for the social sciences are, at best, of little help on this issue. For example, although the issue has been recognized for at least 30 years (see Kennedy, 1970), an examination of the more than 20 statistical texts showed that partial eta squared is mentioned by name in only one (Pedhazur, 1997). Further, not all texts provide equivalent formulas for the more commonly mentioned eta squared. For example, Rosenthal and Rosnow (1985, 1991) provide a formula for eta squared that is equivalent to Cohen's (1973) formula for partial eta squared and discrepant from the formulas for eta squared given in Keppel (1982), Kerlinger (1986), and Kirk (1995). In short, there seems to be much confusion in the literature regarding eta squared and partial eta squared.

The primary aim of this paper is to alert communication researchers to potential errors stemming from the use of SPSS to obtain estimates of eta squared in ANOVA. As a secondary goal, this paper strives to clarify issues concerning the development and appropriate use of eta squared and

partial eta squared in ANOVA. To this end, we begin with a brief discussion of significance testing and effect size. The two alternative formulas for eta squared are provided, and a distinction is drawn between partial eta squared and eta squared based on Cohen (1973) and Pedhazur (1997). The reporting of effect size in SPSS is discussed and the results of tests of simulated data are reported. Examples from published communication articles are cited. Finally, suggestions are offered for future research.

SIGNIFICANCE TESTING AND EFFECT SIZE

Significance tests are employed as the primary decision strategy in most communication research using statistical analyses to test hypotheses or to answer research questions. This significance testing, in its traditional and most common form, might be described as "null hypothesis testing" (Abelson, 1995). In null hypothesis testing, researchers use statistical tests (e.g., z , t , F , Π^2) to ask the question, "if the null hypothesis is true (i.e., the variables are totally unrelated in the population), what is the probability of obtaining the relationship (or a stronger relationship) that was found in my sample?" The resulting p -value is used to form a decision. If the p -value is less than the traditional .05, the null hypothesis is considered to be rejected, the results are said to be statistically significant, and support is inferred for the relationship under study (Gigerenzer et al., 1995). However, if a p -value is .05 or greater, then the relationship between variables is considered to be inconclusive.

There are several limitations with null hypothesis testing, and these are noted as common practice in many, if not most, methods texts (e.g., Abelson, 1995; Keppel, 1982; Kerlinger, 1986; Kirk, 1995). First, the null hypothesis is almost never literally true, so rejecting it is relatively uninformative. Second, significance tests are highly dependent on sample size. When the sample size is small, strong and important effects can be non-significant (i.e., a Type II error is made). Alternatively, when sample sizes are large, even trivial effects can have impressive looking p -values. In short, p -values from null hypothesis significance tests reflect both the sample size and the magnitude of the effects studied.

What is needed is an estimate of the magnitude of effect that is relatively independent of sample size.² The estimates of magnitude of effect or effect size tell us how strongly two or more variables are related, or how large the difference between groups is. Abelson (1995) predicts that "as social scientists move gradually from reliance on single studies and obsession with null hypothesis testing, effect size measures will become more and more popular" (p. 47).

Methods texts often list omega squared, epsilon squared, and eta

squared as estimates of effect size in ANOVA (e.g., Keppel, 1982; Kirk, 1995; Maxwell & Delaney, 1990). Estimates of each of these tend to differ only slightly, especially with moderate or large samples.³ As noted above, of these three, eta squared seems to be the preferred choice among communication researchers.

TWO FORMULAS FOR ETA SQUARED?

Eta squared (η^2 , also known as the correlation ratio or R^2) dates back at least to Karl Pearson (1911) and is most often defined as the sums of squares for the effect of interest divided by the total sums of squares (e.g., Cohen, 1973, 1988; Fisher, 1928, 1973; Hays, 1994; Keppel, 1982; Kerlinger, 1986; Kirk, 1995; Maxwell & Delaney, 1990; Pearson, 1911; Pedhazur, 1997; Sechrest & Yeaton, 1982). That is, eta squared is most often calculated as:

$$\eta^2 = SS_{\text{between}} / SS_{\text{total}} \quad (1)$$

However, other formulas for η^2 can be found (Kennedy, 1970). For example, Rosenthal and Rosnow (1985, 1991) provide the following formula for eta squared:

$$\eta^2 = SS_{\text{between}} / SS_{\text{between}} + SS_{\text{error}} \quad (2)$$

Although these two formulas yield the same result in one-way ANOVAs (where $SS_{\text{total}} = SS_{\text{between}} + SS_{\text{error}}$), they produce different results in more complex ANOVAs (Cohen, 1973; Kennedy, 1970; Sechrest & Yeaton, 1982), and can in fact be widely discrepant.

ETA SQUARED AND PARTIAL ETA SQUARED

Kennedy (1970) noted that "at least two formulas have been recently suggested for the purpose of calculating eta squared within an ANOVA context" (p. 886). He distinguished the "classical interpretation," given in Formula 1 above, from the "Cohen-Friedman approach." The latter used the formula:

$$\eta^2 = n_1F / n_1F + n_2 \quad (3)$$

which Kennedy showed could simplify to Formula 2 above. Cohen (1973)

replied that Formula 1 "is quite properly the formula for η^2 " (p. 108) while Formulas 2 and 3 above are for partial η^2 . Following Cohen (also Pedhazur, 1997; Sechrest & Yeaton, 1982), we will refer to Formula 1 as eta squared and Formulas 2 and 3 as partial eta squared.

A REPORTING ERROR IN SPSS

According to the help menu on SPSS for Windows, 9.0 (1998):

Eta squared is interpreted as the proportion of the *total* variability in the dependent variable that is accounted for by variation in the independent variable. It is the ratio of the between groups sum of squares to the *total* sum of squares [italics added].

This is an accurate description of eta squared, and it is consistent with the classical formulation described above as well as with Cohen's (1973) distinction between partial eta squared and eta squared. However, under the display heading in GLM univariate options of the help menu on SPSS for Windows, 9.0 (1998), the following explanation is provided:

Estimates of effect size give a *partial* eta squared value for each effect and each parameter estimate. The eta squared statistic describes the proportion of the total variability attributable to a factor [italics added].

Thus, while eta squared is defined correctly, that is, consistent with the classical formulation and Cohen's (1973), the online documentation notes that having SPSS estimate effect size produces a partial eta squared rather than eta squared. No definition of partial eta squared is provided.

We ran several complex ANOVAs on SPSS for Windows, 9.0 and requested estimates of effect size. We then calculated both eta squared and partial eta squared by hand based on the sums of squares provided on the printout. In each case, the estimate of effect size on the printout was clearly labeled as eta squared, but the by-hand calculation indicated that partial eta squared rather than eta squared was reported. Hence, we conclude that SPSS is mislabeling estimates of effect size in GLM. While the estimate of effect size is correctly noted to be partial eta squared in the documentation, it is not labeled as such on printouts. Instead, values that are really partial eta squares are incorrectly labeled. Therefore, researchers who do not carefully read the online documentation are likely to misreport effect sizes when SPSS is used for 2+ way ANOVAs.

EXAMPLES OF THE MISREPORTING ERROR

In rare cases with complex designs coupled with multiple substantial effects, the misreporting of partial eta squares as eta squares is sometimes easily detected. In such cases, the errors are easy to spot if the estimates sum to greater than 1.00. This is because eta squares are additive and the sum can never exceed 1.00 (i.e., one cannot account for more than 100% of the variance in a dependent measure) while partial eta squares are not additive, and if added, may sum to over 1.00 (Sechrest & Yeaton, 1982). Hence, if an article appears to account for more than 100% of the variance with the eta squares reported, it is likely that the estimates are partial eta squares rather than eta squares.

As an extreme example, consider the results of LePoire and Yoshimura (1999). They report the results of several 7-way mixed ANOVAs. In reporting effects of a variety of independent variables on their first dependent measure (kinesic involvement), they report that their analysis:

showed significant main effects for communication, $F(1,74) = 212.13, p < .001, \eta^2 = .74$, and manipulation, $F(1,74) = 29.06, p < .001, \eta^2 = .28$; two-way interactions for communication by role, $F(1,74) = 184.94, p < .001, \eta^2 = .71$, role by manipulation, $F(1,74) = 90.20, p < .001, \eta^2 = .55$, communication by manipulation, $F(1,74) = 354.94, p < .001, \eta^2 = .83$, communication by time, $F(1,74) = 4.05, p < .05, \eta^2 = .05$; and a three-way interaction between communication by role by manipulation, $F(1,74) = 339.38, p < .001, \eta^2 = .82$. (p. 12)

Here the errors are easily detected. If we sum the reported effects (i.e., $.74 + .28 + .71 + .55 + .83 + .05 + .82$) it appears that these authors claim to account for 398% of the variance in kinesic involvement. Because this is obviously mathematically impossible, reporting errors clearly exist.

In many, and probably most cases, however, the misreporting of partial eta squared as eta squared should be difficult to detect. If the designs are less complex, or if multiple substantial effects are absent, then the reported effects may not sum to a value greater than 1.00 even if reporting errors do exist.

To assess the prevalence of less obvious errors, we applied Formula 3 (which can be used to estimate partial eta squared from F) to results reported in recent issues of *CM* (Volume 66, No. 3) and *HCR* (Volume 25, No. 3). Recall that five articles in these issues reported eta squared as the estimate of effect size. In three of these five, the designs were such (i.e., more than one independent groups factor) that reporting errors were possible. In each of those three, values reported as eta squared were consistent with our calculations of partial eta squared.⁴ These findings are at least suggestive of reporting errors.

EFFECTS OF THE REPORTING ERROR

The primary effect of reporting partial eta squared as eta squared is that scholars may be systematically overestimating the size of their effects. The degree of overestimation should be a function of the complexity of the design and the size of other effects in the analysis. All things being equal, the greater the number of factors in an analysis, the larger the discrepancy between eta squared and partial eta squared. That is, we would expect errors to be larger in 5-way ANOVAs than in 2-way ANOVAs. The size of errors should be further exacerbated by the inclusion of repeated factors that create additional error terms. Errors in mixed-model ANOVAs are potentially greater than errors in independent-groups ANOVAs, with greater numbers of repeated factors producing increasingly more extreme errors.

Within similar designs, as the number of substantial independent effects increase, the difference between eta squared and partial eta squared will increase. For example, compare examples 1 and 2 to example 3 in Table 1. In example 1, there are two large main effects in a 2-way ANOVA. In example 2, both the main effects and the interaction are substantial. In example 3, however, there are two small main effects and no interaction. The discrepancy between eta squared and partial eta squared is large in the first two examples, but trivial in the third. Hence, even in a 2-way design, partial eta squared can be substantially greater than eta squared.

Perhaps the most important implication of these reporting errors is for meta-analyses, where researchers accumulate effect sizes across studies. If some researchers correctly report eta squared while others misreport partial eta squared values as eta squared, meta-analyses are likely to overestimate population effects. Even more disturbing, however, is that heterogeneity in effect sizes should be evident in future meta-analyses. Such heterogeneity may lead meta-analyses to identify false moderators or to conclude that author effects exist.

APPROPRIATE USES OF ETA SQUARED AND PARTIAL ETA SQUARED

To avoid making reporting errors, authors using SPSS may simply decide to report partial eta squared as their estimate of effect size, but correctly label it in their results section. Such a decision may be made for ease (because it does not require hand calculation), because authors prefer larger looking effect sizes, or because authors assume that if their statistical software reports it, it must be the most useful estimate. We be-

TABLE 1
The Results of Some Simulated 2 x 2 ANOVAs

<i>Sources of variation</i>	SS	df	MS	F	p	η^2	<i>Partial</i> η^2	ω^2	ϵ^2
<i>Example 1: Two large and equal main effects with a large N^a</i>									
Factor A	2500	1	2500	1237	.01	.43	.76	.43	.43
Factor B	2500	1	2500	1237	.01	.43	.76	.43	.43
A x B	00	1	0	00	ns				
Error	800	396	2	.02					
Total	5800								
<i>Example 2: Two large main effects and a large interaction with a large N^b</i>									
Factor A	625	1	625	309	.01	.23	.44	.23	.23
Factor B	625	1	625	309	.01	.23	.44	.23	.23
A x B	625	1	625	309	.01	.23	.44	.23	.23
Error	800	396	2.02						
Total	2675								
<i>Example 3: Two weaker main effects with a smaller N</i>									
Factor A	10	1	10	4.5	.05	.10	.11	.08	.08
Factor B	10	1	10	4.5	.05	.10	.11	.08	.08
A x B	0	1	0	0					
Error	80	36	2.22						
Total	100								

NOTE: ^aSquared zero-order correlations = .43; squared partial correlations =.76. ^bMain effects: Squared zero-order correlations = .23; squared partial correlations =.30.

lieve such decisions would be extremely unfortunate. To explain why, we need to discuss the advantages and disadvantages of eta squared and partial eta squared.

Eta squared has a number of properties that make it a useful estimate of effect size for communication research. First, eta squared is often interpreted in terms of the percentage of variance accounted for by a variable or model. Many people find this to be a useful way of understanding effect size. Second, when there is one degree of freedom in the numerator, the square root of eta squared equals r . This makes eta squared useful because most researchers have a good understanding of the meaning of a correlation. This also makes the reported effects useful for subsequent meta-analyses because the results can be easily converted to d or r . When there is more than one degree of freedom in the numerator, eta squared equals R^2 which is also well understood. Third, eta squared has the property that the effects for all components of variation (including error) will sum to 1.00. This quality is intuitively appealing (Sechrest & Yeaton, 1982).

Fourth, a fundamental characteristic of effect size is that as the number of independent causes of an effect increase, the effect size for individual causes decreases (Ahadi & Diener, 1989). Eta squared possesses this property while partial eta squared does not. Fifth, eta squared is addressed in many methods texts while partial eta squared is rarely discussed. Consequently, its properties are better understood by social scientists, especially those with limited mathematical training. Sixth, because eta squared is always either equal to partial eta squared or smaller, it may be seen as a more conservative estimate than partial eta squared and this may be appealing to many readers, reviewers, and editors. Finally, eta squared is very easy to calculate.

Despite the advantages, eta squared has some limitations. Perhaps the harshest criticism came from Fisher (1928) who noted "as a descriptive statistic the utility of the correlation ratio is extremely limited" (p. 224). One reason for Fisher's concern may have been that eta squared tends to be upwardly biased, especially when the sample size is small.⁵ Under such circumstances, omega squared or epsilon squared are better choices because these offer corrections (Hays, 1994; Keppel, 1982; Kirk, 1995; Maxwell & Delaney, 1990). Epsilon squared may be especially familiar to communication researchers as it is equivalent to shrunken R^2 which is labeled as adjusted R^2 on SPSS regression printouts. Because communication research tends toward larger sample sizes (and few levels of the independent variable), the most commonly noted limitation of eta squared is of little practical importance. Nevertheless, several methods texts suggest that omega squared or epsilon squared is preferred to eta squared (e.g., Hays, 1994; Keppel, 1982; Kirk, 1995; Maxwell & Delaney, 1990).

In contrast, partial eta squared has few of the advantages of eta squared. It is not a percentage of the total sums of squares, it is not additive (Cohen, 1973; Sechrest & Yeaton, 1982), and is not equivalent to the familiar r^2 or R^2 . It should be of little use in meta-analyses, and is less well known and understood. Further, if eta squared is criticized for being an overestimate, then partial eta squared is much more susceptible to this criticism because it is often larger still. In sum, our examination of the literature revealed little reason for the reporting of partial eta squared.

One notable exception, however, is offered by Cohen (1973; cf. Pedhazur, 1997). Cohen argued that under certain circumstances, if one wishes to compare the size of an effect of an identical manipulation across studies with different designs, partial eta squared may prove more comparable than eta squared. To illustrate Cohen's point, consider a hypothetical researcher, Dr. Beta. Dr. Beta is involved in an ongoing program of research investigating the communicative effects of factor A. In the first study, Dr. Beta manipulates factor A and tests the effects of factor A with a one-way, independent-groups, fixed-effects ANOVA. The hypothetical results of

TABLE 2
Simulated Effects of an Additional Factor on the Comparability of Effect Size Estimates

<i>Sources of Variation</i>	SS	df	MS	F	p	η^2	<i>Partial</i> η^2	ω^2	ϵ^2
<i>One-way ANOVA table</i>									
Factor A	25	1	25	33	.05	.25	.25	.24	.24
Error	75	98	.76						
Total	100								
<i>Two-way example 1: Inclusion of factor B that reduces error</i>									
Factor A	25	1	25	48	.05	.25	.50	.24	.24
Factor B	25	1	25	48	.05	.25	.50	.24	.24
A x B interaction	0	1	0	0	ns	.00	.00		
Error	50	96	.52						
Total	100								
<i>Two-way example 2: Inclusion of factor B that instills variation</i>									
Factor A	25	1	25	32	.05	.20	.25	.19	.19
Factor B	25	1	25	32	.05	.20	.25	.19	.19
A x B interaction	0	1	0	0	ns	.0	.00		
Error	75	96	.77						
Total	125								

NOTE: $N = 100$.

this analysis are presented in the one-way table in Table 2. Dr. Beta finds that factor A had a substantial effect on the dependent variable, $F(1, 25) = 32.89$, $p < .05$, $\eta^2 = .25$, *partial* $\eta^2 = .25$, $\omega^2 = .24$, $\epsilon^2 = .24$. Based on this initial success, Dr. Beta seeks to expand the research to include factor B. Factor A is crossed with factor B in a carefully controlled experiment, creating a 2×2 independent-groups, fixed-effects ANOVA. Dr. Beta is interested to see how the effects of factor A compare across studies. Two possible outcomes are shown in Table 2. In each case, $SS_A = 25.0$. In two-way example 1, Factor B functions to reduce the error term. That is, the inclusion of the additional independent variable explained previously unexplained variance in the dependent measure. Under these circumstances, η^2 , ω^2 , and ϵ^2 all produce estimates comparable to the initial study with a one-way design, while *partial* η^2 differs widely between studies. This is just the sort of circumstance that lead Kennedy (1970) to critique *partial* η^2 . However, consider 2-way example number 2 in Table 2. In this case, the inclusion of factor B does not explain previously unexplained variance, but instead instills additional variation in the dependent variable. In this case, *partial* η^2 offers the more comparable estimate of effect

size. Situations such as this led Cohen (1973) to argue that partial eta squared can be more comparable when additional manipulated or control variables are added to a design. The reader should note, however, that Cohen did not advocate the wholesale use of partial eta squared, but instead advocated its use for very specific purposes under very specific conditions.

The major limitation with the use of partial eta squared is that it requires researchers to understand the effects that the inclusion of additional variables have on both error terms and the total variability in the dependent variable. That is, while clean examples were provided as illustrations in Table 2, in actual practice, the inclusion of additional manipulated or control variables may both reduce error and instill additional variation. It is the authors' view that few, if any, programs of research in communication are advanced enough to fully understand the bases of error and total variability in a study. For this reason, caution is surely warranted.

If any parallel is to be drawn, partial eta squared appears to be equivalent to the squared partial correlation. Agresti and Finlay (1997) define the squared partial correlation as the partial proportion of the variance explained in the dependent variable explained uniquely by the variable of interest divided by the proportion of variance in the dependent variable unexplained by the other variable(s). This definition reduces to SS factor A in the numerator and SS total minus SS factor B in the denominator. This formula is the same as that used for partial eta squared. To check this claim, one can see that in Table 1, example 1 this formula yields $2500 / (5800 - 2500) = .76$, which equals partial eta squared. To check this result, we hand calculated the partial correlation for factor A, using the correlation coefficients calculated from each main effect (in each case, $r = .66$). We obtained partial correlations of .87 that, when squared, equal partial eta squared.

In example 2, however, the squared partial correlation (.30) is considerably lower than the estimates of partial eta squared. This deviation, however, is due to the interaction between the factors. Essentially, the interaction term becomes another variable, and thus a higher-order partial correlation is warranted rather than the first order partial correlation for example 1. Higher order partial correlations call for calculations of the multiple correlation for all of the independent variables with the dependent variable, as well as the multiple correlation of the independent variables whose effects are being controlled. The equation is as follows:

$$r^2_{YX_1 \cdot X_2, X_3} = [R^2_{Y(X_1, X_2, X_3)} - R^2_{Y(X_2, X_3)}] / [1 - R^2_{Y(X_2, X_3)}] \quad (4)$$

For the present purposes, let Y represent the dependent variable, X_1 equal factor A, X_2 equal factor B, and X_3 equal the interaction term. We created a dummy variable to represent the interaction term where codes of "0" were assigned to (a) the condition where factor A = 1 and factor B = 1, (b) the condition where A = 2 and B = 2, and (c) where "1" was assigned for the remaining two conditions (the crossover interaction). We regressed our dependent variable onto the original two factors as well as this new dummy variable to find R^2 total (.701), then removed factor A from the regression equation to obtain the R^2 for factor B and the dummy variable (.467). When these values were entered into Formula 4, we obtained the value we initially obtained as partial eta squared (.44). Thus, the data were again consistent with our contention that partial eta squared is equivalent with the squared partial correlation.

SUGGESTIONS FOR RESEARCHERS

Based on our exploration of eta squared and partial eta squared, we offer three suggestions for communication researchers.

1. Researchers should most often report eta squared, omega squared, or epsilon squared rather than partial eta squared.
2. Researchers should take care to accurately and correctly report effect sizes in their papers and articles. Effect sizes that are really partial eta squares should not be reported as eta squares.
3. Researchers who report partial eta squared should read Cohen (1973) and Kennedy (1970) so they are informed of its proper use, and should report eta squared in addition to partial eta squared.

SUMMARY AND CONCLUSIONS

The importance of reporting effect sizes to augment significance tests is commonly recognized in communication research. Eta squared (the ratio of sum of squares for an effect to the total sum of squares) appears to be the effect size estimate of choice for communication researchers. However, serious reporting errors connected with eta squared have appeared in recently published articles. Communication researchers appear to be reporting partial eta squared (sum of squares effect over sum of squares effect plus the sum of squares error) but labeling it as eta squared. We argue that this practice will have negative consequences on the field, and that the extent of this error is more extreme with complex ANOVAs (especially mixed-model ANOVAs) as well as when multiple effects are substantial. We suspect that these errors stem from a reporting error on SPSS printouts. The estimate of effect size that SPSS reports as eta squared is really a partial eta squared.

After reviewing the strengths and limitations of several estimates of effect size used in ANOVA, we conclude that communication researchers would be well served by reporting eta squared, omega squared, or epsilon squared rather than partial eta squared. If partial eta squared is reported at all, it should be correctly labeled as such and should be accompanied by estimates of eta squared. We hope that this paper will alert communication researchers to some often unrecognized issues in the reporting and interpretation of effect sizes.

NOTES

1. We are using the term "effect size" in a general sense including what also might be labeled variance accounted for, magnitude of effect, and strength of association. Readers should note, however, that our use of the term is inconsistent with distinctions made by some authors. For example, Kirk (1995) as well as Maxwell and Delaney (1990) distinguish between estimates of strength of association (which includes eta squared) and estimates of effect size.

2. Murray and Dosser (1987) argue that all measures of effect size depend on sample size, either directly or indirectly.

3. Eta squared tends to be a biased estimate of the population effect size when N is small or when there are several levels of the independent variable. Omega squared and epsilon squared offer corrections for this bias (Hays, 1994; Keppel, 1982; Maxwell & Delaney, 1990).

4. In some cases the reported values differed from the calculated values by as much as .02. In each case, the reported eta squared was larger than the partial eta squared calculated from F . Because partial eta squared can never be greater than eta squared, either there are minor errors in our estimation technique or some authors take liberties in rounding their estimates. If the reported values were lower than our calculation, we would have inferred that they were accurately reporting eta squared.

5. Another concern of Fisher's (1928) was that the distributions of eta squared was not known to him, and hence calculating confidence interval was problematic. The sampling distribution of η^2 is now known to be a beta distribution (Murray & Dosser, 1987).

REFERENCES

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Erlbaum.
- Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences*. Upper Saddle River, NJ: Prentice Hall.
- Ahadi, S., & Diener, E. (1989). Multiple determinants and effect size. *Journal of Personality and Social Psychology*, *56*, 398-406.
- Boster, F. J. (2002). On making progress in communication science. *Human Communication Research*, *28*, 473-490.
- Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, *33*, 107-112.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Fisher, R. A. (1928). *Statistical methods for research workers* (2nd ed.). New York: Hafner.
- Fisher, R. A. (1973). *Statistical methods for research workers* (14th ed.). New York: Hafner.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Kruger, L. (1995). *The empire of chance: How probability changed science and everyday life*. Cambridge, UK: Cambridge University Press.

- Hays, W. L. (1994). *Statistics*. Fort Worth, TX: Harcourt Brace.
- Kennedy, J. J. (1970). The eta coefficient in complex ANOVA designs. *Educational and Psychological Measurement, 30*, 855–889.
- Keppel, G. (1982). *Design and analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Kerlinger, F. N. (1986). *Foundations of behavioral research*. New York: Holt, Rinehart & Winston.
- Kirk, R. E. (1995). *Experimental design*. Pacific Grove, CA: Brooks Cole.
- LePoire, B. A., & Yoshimura, S. M. (1999). The effects of expectancies and actual communication on nonverbal adaptation and communication outcomes: A test of interaction adaptation theory. *Communication Monographs, 66*, 1–30.
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data*. Pacific Grove, CA: Brooks Cole.
- Murray, L. W., & Dosser, D. A. (1987). How significant is a significant difference? Problems with the measurement of magnitude of effect. *Journal of Counseling Psychology, 34*, 68–72.
- Pearson, K. (1911). On a correction needful in the case of the correlation ratio. *Biometrika, 8*, 254–256.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research*. Fort Worth, TX: Harcourt Brace.
- Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. Cambridge, UK: Cambridge University Press.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research*. New York: McGraw Hill.
- Sechrest, L., & Yeaton, W. H. (1982). Magnitude of experimental effects in social science research. *Evaluation Review, 6*, 579–600.
- SPSS for Windows (Version 9.0) [Computer software]. (1998). Chicago: SPSS Inc.