CAUTIONARY NOTE ON REPORTING ETA-SQUARED VALUES FROM MULTIFACTOR ANOVA DESIGNS

CHARLES A. PIERCE
University of Memphis

RICHARD A. BLOCK
Montana State University

HERMAN AGUINIS
University of Colorado at Denver

The authors provide a cautionary note on reporting accurate eta-squared values from multifactor analysis of variance (ANOVA) designs. They reinforce the distinction between classical and partial eta-squared as measures of strength of association. They provide examples from articles published in premier psychology journals in which the authors erroneously reported partial eta-squared values as representing classical eta-squared values. Finally, they discuss broader impacts of inaccurately reported eta-squared values for theory development, meta-analytic reviews, and intervention programs.

Keywords: *analysis of variance; eta-squared; partial eta-squared; effect size*

Psychologists have recently devoted considerable attention to null hypothesis significance testing (NHST) (e.g., Krueger, 2001; Nickerson, 2000). Some favor NHST (e.g., Hagen, 1997; Wainer, 1999), whereas others oppose its use (e.g., Falk, 1998). Regardless of one's position on NHST, most

---

A previous version of this article was presented as part of a symposium (M. L. Williams, Chair) at the meeting of the Academy of Management, New Orleans, August 2004. The research reported herein was conducted while the first author was at Montana State University. Correspondence concerning this article should be addressed to Charles A. Pierce, Department of Management, Fogelman College of Business and Economics, University of Memphis, Memphis, TN 38152-3120; e-mail: capierce@memphis.edu.

Educational and Psychological Measurement, Vol. 64 No. 6, December 2004 916-924
DOI: 10.1177/0013164404264848

researchers agree there are alternative methods available for interpreting and reporting empirical findings. Among these alternatives is the use of measures of strength of association (Fidler, 2002; Loftus, 1996; Maxwell, Camp, & Arvey, 1981).

The goal of the present article is to provide a cautionary note on reporting accurate eta-squared values from multifactor analysis of variance (ANOVA) designs. We reinforce the distinction between classical and partial eta-squared as measures of strength of association. We acknowledge that statistics textbooks describe classical and partial eta-squared (e.g., Cohen, Cohen, West, & Aiken, 2003; Keppel, 1991; Maxwell & Delaney, 2000). We also acknowledge that others have distinguished between these two measures (e.g., Cohen, 1973; Haase, 1983; Kennedy, 1970). Nevertheless, we show that researchers in psychology apparently do not understand the distinction between classical and partial eta-squared. To demonstrate the need to make this distinction more apparent, we provide several examples from articles published in premier psychology journals in which the authors erroneously reported partial eta-squared values as representing classical eta-squared values. We therefore alert researchers in education and psychology to this reporting inaccuracy. Finally, we discuss broader impacts of inaccurately reported eta-squared values for theory development, meta-analytic reviews, and intervention programs.

Eta-Squared and Other Measures of Strength of Association

When using ANOVA to analyze data, several measures of strength of association are available. These include eta-squared (η^2), omega-squared (ω^2), and epsilon-squared (ε^2) (Cohen et al., 2003; Hays, 1994; Keppel, 1991; Maxwell & Delaney, 2000). Many former and current journal editors espouse a position consistent with the American Psychological Association (APA) in terms of requiring authors to report these measures (e.g., Kendall, 1997; Thompson, 1994; Zedeck, 2002). This position is reinforced by a statement in APA's *Publication Manual:* "For the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your Results section" (APA, 2001, p. 25). Perhaps in response to this and previous calls for a more frequent use of measures of strength of association, eta-squared is commonly reported in the education and psychology literatures.

Although eta-squared is frequently reported, it is an upwardly biased estimate of the population strength of association between an independent variable and a dependent variable, particularly when total sample size is small. Omega-squared and epsilon-squared, on the other hand, are unbiased estimates and thus should be reported when researchers want to estimate the population strength of association rather than merely provide a descriptive summary of their sample data (Cohen et al., 2003; Hays, 1994; Maxwell & Delaney, 2000). However, because common statistical software packages such as SPSS only report eta-squared values and not omega-squared or epsilon-squared values in their ANOVA output files, many researchers in education and psychology report eta-squared values. Considering the ease with which eta-squared values can be obtained, along with the fact that these values have been inaccurately reported in the psychology literature, the present article focuses on eta-squared.

Distinction Between Classical and Partial Eta-Squared

As a descriptive index of strength of association between an experimental factor (main effect or interaction effect) and a dependent variable, classical eta-squared is defined as the proportion of total variation attributable to the factor, and it ranges in value from 0 to 1 (Cohen et al., 2003; Hays, 1994; Maxwell & Delaney, 2000). From information reported in an ANOVA summary table, classical eta-squared is computed as follows:

$$\text{classical } \eta^2 = SS_{\text{factor}}/SS_{\text{total}}, \tag{1}$$

where SS_{factor} is the variation attributable to the factor and SS_{total} is the total variation.

In contrast to classical eta-squared, partial eta-squared for an experimental factor is defined as the proportion of total variation attributable to the factor, partialling out (excluding) other factors from the total nonerror variation (Cohen, 1973; Haase, 1983; Kennedy, 1970). Partial eta-squared values also range from 0 to 1. From information reported in an ANOVA summary table, partial eta-squared is computed as follows:

$$\text{partial } \eta^2 = SS_{\text{factor}}/(SS_{\text{factor}} + SS_{\text{error}}), \tag{2}$$

where SS_{factor} is the variation attributable to the factor and SS_{error} is the error variation.

In a multifactor ANOVA, $SS_{\text{factor}} + SS_{\text{error}}$ in the denominator of Equation 2 is always less than or equal to the corresponding SS_{total} in the denominator of Equation 1. Consequently, partial eta-squared is typically greater than classical eta-squared for a source of variance. There is, however, an exception: Classical and partial eta-squared are identical in a design that has only one factor. In a one-way ANOVA, $SS_{\text{total}} = SS_{\text{factor}} + SS_{\text{error}}$ and thus classical and partial eta-squared are equivalent. However, in a multifactor ANOVA, classical and partial eta-squared are equivalent for a source of variance only if it is the sole source that contributes to the total nonerror variation (Cohen, 1973; Haase, 1983; Kennedy, 1970). Hence, classical and partial eta-squared are almost always unequal.

With respect to any multifactor ANOVA, partial eta-squared values can sum to greater than 1, but classical eta-squared values cannot (Cohen, 1973; Haase, 1983). The reason why partial eta-squared values can sum to greater than 1 is that they are computed using different values for the total variation ($SS_{\text{factor}} + SS_{\text{error}}$) in the denominator of Equation 2. Thus, partial eta-squared is not a measure of unique variation in the dependent variable in that some of the nonerror variation can be accounted for by other factors in the analysis. In contrast, classical eta-squared values cannot sum to greater than 1 because each is computed using the same value for SS_{total} in the denominator of Equation 1. Thus, classical eta-squared is an additive measure of unique variation in the dependent variable in that the nonerror variation cannot be accounted for by other factors in the analysis.

Erroneously Reported Classical Eta-Squared Values

Levine and Hullett (2002) discovered that researchers in communication have mistakenly reported partial eta-squared values as representing classical eta-squared values. However, they cited only one example of this misreporting problem. More important, the sole example they provided was from the communication literature. To date, there is no evidence that this problem exists in the psychology literature. We provide several examples from articles published in premier psychology journals in which partial eta-squared values were apparently reported as representing classical eta-squared values. As did Levine and Hullett, we established this inaccuracy because the reported classical eta-squared values from each analysis sum to greater than 1. Levine and Hullett suggested the erroneously reported values might be attributable to the fact that versions of SPSS prior to 11.0 (SPSS, Inc., 2001) mislabeled partial eta-squared values as *eta-squared* values in output files. However, the misreporting problem is not limited to researchers who use SPSS. The authors of one article we found (Huang-Pollock, Carr, & Nigg, 2002) stated that they used SAS to analyze their data. Unfortunately, none of the authors of the other articles we found specified which statistical analysis software they used.

Table 1 shows examples of erroneously reported classical eta-squared values in premier psychology journals. Whereas Levine and Hullett's (2002) sole example was from a complex seven-way mixed-factor ANOVA, our examples demonstrate that the reporting inaccuracy also occurs with simpler two-way ANOVAs. To illustrate the reporting inaccuracy depicted in Table 1, consider results that Shepperd and McNulty (2002) reported for their research on individuals' affective responses to expected versus unexpected

Table 1
Examples of Erroneously Reported Classical Eta-Squared Values in Premier Psychology Journals

Authors and Journal	Analysis	Eta-Squared Values Reported for Main Effects	Eta-Squared Values Reported for Interactions	Total Variation Explained (%)
1. Bressan & Dal Martello (2002), Exp 1; *Psychological Science*	2 × 2 × 2 × 2 × 2 mixed ANOVA	.72 for belief in relatedness; .59 for biological relatedness		131
2. Huang-Pollock, Carr, & Nigg (2002), Study 2, Table 3; *Developmental Psychology*	Six 4 × 2 repeated-measures (RM) ANOVAs	.26-.59 for distractor; .56-.87 for set size	.09-.43 for Distractor × Set Size	141[a]
3. Roefs & Jansen (2002); *Journal of Abnormal Psychology*	2 × 2 × 2 × 2 mixed ANCOVA	.76 for combination; .12 for order; .28 for age as covariate	.12 for Combination × Group; .52 for Combination × Order; .06 for Age × Combination	186
4. Shepperd & McNulty (2002), Study 1; *Psychological Science*	2 × 2 ANOVA	.46 for grade expected; .85 for grade received		131
5. Shepperd & McNulty (2002), Study 2; *Psychological Science*	2 × 2 RM ANOVA	.88 for raise expected; .99 for raise received		187
6. Shepperd & McNulty (2002), Study 3; *Psychological Science*	2 × 2 ANCOVA	12 for news expected; .68 for news received; .30 for affect as covariate		110
7. Walton, Walton, & Felton (2001), Exp 1; *Journal of Educational Psychology*	4 × 3 × 2 mixed ANOVA	.22 for treatment; .49 for word type; .30 for reading clues	.09 for Treatment × Word Type	110
8. Wichman, Friel, & Harris (2001); *Memory & Cognition*	Four 2 × 3 RM ANOVAs	.35-.96 for article; .38-.91 for sentence type	.60-.69 for Article × Sentence Type	204[a]

a. Averaged across analyses reported.

positive and negative outcomes. In their Study 1, they conducted a 2 × 2 between-subjects ANOVA and reported classical eta-squared values of .46 for the main effect of grade expected and .85 for the main effect of grade received. According to these results, the two main effects accounted for an impossible 131% of the total variation in participants' responses.

Because this analysis was unlike the others shown in Table 1 in that it was based on a purely between-subjects design and, moreover, because the authors reported means, standard deviations, cell *n*s, and *F* values for the entire design, we were able to reconstruct the ANOVA summary table using Johnson's (1993) DSTAT meta-analytic software. This reconstructed ANOVA table was used to compute accurate classical eta-squared values for the purpose of determining the degree to which the erroneously reported values were inflated. As one example, for the main effect of grade expected, the erroneous classical eta-squared value of .46 was 4.2 times larger than the accurate classical eta-squared value of .11. Shepperd and McNulty (2002) concluded that their findings "provide *strong* [italics added] support" for decision affect theory (p. 87). However, because the authors interpreted inflated classical eta-squared values, they overestimated the extent to which expectations and outcomes influenced participants' predicted and actual affective responses, which calls into question whether decision affect theory received strong support.

The reporting inaccuracy illustrated in Table 1 is detectable only when classical eta-squared values sum to greater than 1. However, considering that classical eta-squared values typically range from .01 to .09 in the social sciences (Cohen, 1988) and, moreover, that most authors do not report SS_{factor}, SS_{error}, and SS_{total} values, the reporting inaccuracy is usually impossible to detect. Thus, it is impossible to estimate the pervasiveness of this reporting inaccuracy in the psychology literature. We concur with Levine and Hullett (2002), who stated, "Although we do not know how widespread these errors are, we suspect they are quite common and most often are likely to go unnoticed" (p. 613). The misreporting problem is probably more widespread than one may think considering that the sum of reported classical eta-squared values rarely approaches 1 and that the sum of partial eta-squared values rarely exceeds 1.

Broader Impacts of Erroneously Reported Eta-Squared Values

Why should this misreporting problem concern researchers in education and psychology? For any multifactor ANOVA with more than one nonzero source of variance, partial eta-squared values are greater than the corresponding classical eta-squared values. In fact, partial eta-squared values can be greater than .50 even if the corresponding classical eta-squared values are less than .20. In the six articles we use as examples, the sum of reported classical eta-squared values ranged from 1.10 to 2.04 (M = 1.53 or 153% of the total variation) for the analyses we highlighted. Thus, on average, the reported classical eta-squared values were inflated by 53%, assuming that the nonerror sources in each analysis accounted for 100% of the total variation. If we assume more realistically that the nonerror sources accounted for between 10% and 75% of the total variation, on average the reported classical eta-squared values were inflated by at least 78% and perhaps as much as 143%.

When considering the broader impacts of misreporting partial eta-squared values as representing classical eta-squared values, we acknowledge they depend on whether there was more than one factor with a SS_{factor} value greater than 0 and whether the magnitude of the disparity between the two values for a source of variance is small or large. From a theory-testing standpoint, researchers who draw conclusions from inflated classical eta-squared values will overestimate the influence of main effects and interaction effects. These misguided conclusions will impact the further development and testing of a theory. Another impact is that meta-analysts will compute inflated study-level effect size estimates (*d*s or *r*s) from classical eta-squared values without realizing they are actually partial eta-squared values. These study-level effect size estimates will result in misguided meta-analytic conclusions that, in turn, will impact the further development and testing of a theory. Meta-analysts should thus be especially cautious and rely on other reported statistics such as means, standard deviations, and sample sizes, or *d*s or *r*s, instead of eta-squared values. If other statistics are not reported and if eta-squared values are unreasonably high (i.e., they sum to nearly 1 or greater than 1), meta-analysts should contact the authors to obtain information needed to compute eta-squared values or exclude the study from the analysis. Finally, researchers who draw conclusions from inflated classical eta-squared values will make misguided recommendations regarding the use of a costly treatment or intervention program about which the degree of success was misrepresented in empirical trials.

In closing, we caution researchers on reporting accurate eta-squared values from multifactor ANOVA designs. We reinforce the distinction between classical and partial eta-squared as measures of strength of association. This distinction needs to be reinforced because researchers have erroneously reported partial eta-squared values as representing classical eta-squared values. We illustrate this erroneous reporting using six articles recently published in a diverse set of prestigious psychology journals. These illustrations show that, on average, the erroneously reported values were inflated by at least 53% and perhaps as much as 143%. Finally, we are not suggesting classical eta-squared values should be reported instead of partial eta-squared values. To the contrary, there are situations in which a partialled measure is preferred over a classical measure (Breaugh, 2003; Cohen, 1973; Haase, 1983; Keren & Lewis, 1979; Olejnik & Algina, 2000). These situations include those in which researchers who used a multifactor design (a) want to report an index of the strength of association between an independent variable and a dependent variable that excludes variance produced by other factors or (b) want to compare the strength of association between the same independent variable and dependent variable across studies that have different factorial designs. Regardless of whether they are classical or partial, eta-squared values should be accurately reported because inaccurately reported values can have a detrimental impact on theory development, meta-analytic conclusions, and the anticipated success of intervention programs.

References

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

Breaugh, J. A. (2003). Effect size estimation: Factors to consider and mistakes to avoid. *Journal of Management, 29*, 79-97.

Bressan, P., & Dal Martello, M. F. (2002). *Talis pater, talis filius*: Perceived resemblance and the belief in genetic relatedness. *Psychological Science*, *13*, 213-218.

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* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Falk, R. (1998). In criticism of the null hypothesis statistical test. *American Psychologist, 53*, 798-799.

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.

Haase, R. F. (1983). Classical and partial eta square in multifactor ANOVA designs. *Educational and Psychological Measurement, 43*, 35-39.

Hagen, R. L. (1997). In praise of the null hypothesis statistical test. *American Psychologist, 52*, 15-24.

Hays, W. L. (1994). *Statistics* (5th ed.). Belmont, CA: Wadsworth.

Huang-Pollock, C. L., Carr, T. H., & Nigg, J. T. (2002). Development of selective attention: Perceptual load influences early versus late attentional selection in children and adults. *Developmental Psychology*, *38*, 363-375.

Johnson, B. T. (1993). *DSTAT 1.10: Software for the meta-analytic review of research literatures*. Hillsdale, NJ: Lawrence Erlbaum.

Kendall, P. C. (1997). Editorial. *Journal of Consulting and Clinical Psychology, 65*, 3-5.

Kennedy, J. J. (1970). The eta coefficient in complex ANOVA designs. *Educational and Psychological Measurement, 30*, 885-889.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Keren, G., & Lewis, C. (1979). Partial omega squared for ANOVA designs. *Educational and Psychological Measurement, 39*, 119-128.

Krueger, J. (2001). Null hypothesis significance testing: On the survival of a flawed method. *American Psychologist, 56*, 16-26.

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research, 28*, 612-625.

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

Maxwell, S. E., Camp, C. J., & Arvey, R. D. (1981). Measures of strength of association: A comparative examination. *Journal of Applied Psychology, 66*, 525-534.

Maxwell, S. E., & Delaney, H. D. (2000). *Designing experiments and analyzing data: A model comparison perspective*. Mahwah, NJ: Lawrence Erlbaum.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241-301.

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

Roefs, A., & Jansen, A. (2002). Implicit and explicit attitudes toward high-fat foods in obesity. *Journal of Abnormal Psychology*, *111*, 517-521.

Shepperd, J. A., & McNulty, J. K. (2002). The affective consequences of expected and unexpected outcomes. *Psychological Science*, *13*, 85-88.

SPSS, Inc. (2001). *SPSS for Windows* (release 11.0.1) [Computer software]. Chicago: Author.

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement, 54*, 837-847.

Wainer, H. (1999). One cheer for null hypothesis significance testing. *Psychological Methods, 4*, 212-213.

Walton, P. D., Walton, L. M., & Felton, K. (2001). Teaching rime analogy or letter recoding reading strategies to prereaders: Effects on prereading skills and word reading. *Journal of Educational Psychology, 93*, 160-180.

Wichman, A. L., Friel, B. M., & Harris, R. J. (2001). The effect of lexical, pragmatic, and morphological violations on reading time and deviance ratings of English and German sentences. *Memory & Cognition*, *29*, 493-502.

Zedeck, S. (2002). An expanded scope for the *Journal of Applied Psychology*. *The Industrial-Organizational Psychologist, 40* (2), 140-141.