Psicothema

Psicothema 2016, Vol. 28, No. 3, 340-345
doi: 10.7334/psicothema2015.124

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Generalized eta squared for multiple comparisons on between-groups designs

María Eva Trigo Sánchez and Rafael Jesús Martínez Cervantes
Universidad de Sevilla

Abstract

Background: Psychological and educational researchers are experiencing many practical difficulties in following the guidelines of the American Psychological Association (APA) for their statistical analyses: one such difficulty is the reporting of an effect-size measure along with each test of statistical significance (APA, 2010). The problem is exacerbated when researchers focus on contrast analysis instead of omnibus tests and when the Type-I error rate per comparison has to be adjusted. **Method:** Several reasons for this problem are discussed, with emphasis on the facts that researchers may be presented with too many optional effect-size measures with varying degrees of adequacy in several designs, and common statistical packages fail to provide appropriate effect-size measures for contrast analysis. **Results:** This study proposes specific procedures (also implemented in spreadsheets) to compute generalized eta squared for various kinds of hypotheses, either general or specific, for one-factor and factorial between-group designs, and with manipulated and/or measured factors. **Conclusions:** Finally, conclusions are drawn concerning the need to take into account the kind of design and the kind of hypothesis in order to calculate comparable effect-size indexes across different types of studies and to prevent an overestimation of effect size.

Keywords: APA guidelines, effect size, generalized eta squared, contrast analysis, post-hoc comparisons.

Resumen

Eta cuadrado generalizado para comparaciones múltiples en diseños entregrupos. **Antecedentes:** los investigadores en Psicología y Educación están teniendo muchas dificultades prácticas para seguir la directriz de la Asociación Americana de Psicología (APA) de aportar una medida de tamaño de efecto junto con cada prueba de significación (APA, 2010). El problema se agrava cuando se realizan contrastes a priori en lugar de pruebas ómnibus y cuando la tasa de error de Tipo I por comparación tiene que ser ajustada. **Método:** se discuten diversas razones para ello, como la existencia de muchas medidas diferentes de tamaño de efecto y el hecho de que los paquetes estadísticos comunes como SPSS no proporcionan medidas apropiadas para las comparaciones múltiples. **Resultados:** se proponen procedimientos específicos (también implementados en hojas de cálculo) para calcular el índice eta cuadrado generalizado para diversos tipos de hipótesis, generales o específicas; tipos de diseños, univariables o factoriales; y con factores manipulados y/o medidos. **Conclusiones:** finalmente se concluye sobre la necesidad de tener en cuenta el tipo de diseño y el tipo de hipótesis para obtener índices de tamaño de efecto comparables entre diferentes tipos de investigaciones y que eviten una sobreestimación del mismo.

Palabras clave: recomendaciones de la APA, tamaño de efecto, eta cuadrado generalizado, comparaciones a priori, comparaciones a posteriori.

Since 1994, the American Psychological Association (APA) has recommended reporting an effect-size measure along with each test of statistical significance (APA, 1994), which is usually evaluated according to the levels proposed by Cohen (1988) for a small, medium, or large effect size. After a report by Cohen (1994) on statistical significance, and the revision of empirical studies that showed the minor influence of the APA recommendation on scientific reports published (e.g., Keselman et al., 1998; Kirk, 1996), the Task Force on Statistical Inference stressed the importance of reporting and interpreting effect-size statistics to improve scientific quality (Wilkinson & the APA Task Force on Statistical Inference, 1999). Later editions of the APA Publication Manual (2001; 2010) included this recommendation. However, a review of articles published between 2005 and 2007 revealed that fewer than half (40%) of the reported analyses in APA journals included an effect-size measure, of which only 51% interpreted the reported index (Sun, Pan, & Wang, 2010). Similar results were obtained with articles published in 2009 and 2010, due to the lack of appropriate effect-size measures for ANOVA-related analysis, such as simple effects, post-hoc comparisons, and planned contrasts (Fritz, Morris, & Richler, 2011). The current situation can be explained by a range of causes.

First of all, a large number of researchers have failed to realize both the relevance of effect sizes, and the very different information they provide to that obtained from statistical significance. Rosnow and Rosenthal (2009, pp. 1) summarize the relation between the statistical significance, effect sizes, and sample sizes:

$$\text{Significance test} = \text{Size of effect} \times \text{Size of study}$$

Received: May 13, 2015 • Accepted: March 2, 2016
Corresponding author: María Eva Trigo Sánchez
Facultad de Psicología
Universidad de Sevilla
41018 Sevilla (Spain)
e-mail: trigo@us.es

This relation makes it clear that, with a small effect (not zero), an increase in the number of observations can produce a statistically significant result. It also clarifies the importance of performing a power analysis before collecting data to guarantee that statistical significance can be reached with a determined effect size if an adequate sample size is used. Reaching an adequate sample size is especially relevant, as the power of a large amount of research studies in Psychology is under the convenient $1 - \beta = .80$ (Cohen, 1992; Sedlmeier & Gigerenzer, 1989; Valera, Sánchez, & Marín, 2000). Another useful application for effect-size measures is that, by interpreting them jointly with statistical significance, researchers can distinguish between results of a more confident nature, when both indexes lead to the same conclusion, and other situations that suggest possible threats to the statistical conclusion validity of the study, when a discrepancy arises between both indexes.

A second possible cause is that the good practice to systematically report effect-size indexes is been obstructed by the use of some statistical packages, such as SPSS, because their output cannot provide appropriate effect-size measures. With certain versions of the package and certain designs, there is also some confusion between what is claimed to be reported, eta squared, and what is actually reported, partial eta squared (Levine & Hullett, 2002; Pierce, Block, & Aguinis, 2004).

Finally, researchers may be presented with too many optional effect-size measures with varying degrees of adequacy in several designs (Ferguson, 2009). For example, Kirk (1996) identified 40 different measures. Therefore, recommending a single index would constitute the first step towards rendering the reporting and interpreting of effect-size statistics a more frequent practice. To this end, η^2, also known as the correlation ratio or R^2, is the most common measure reported by researchers in Psychology and Education because it is the index provided by common statistical packages, such as SPSS (Pierce et al., 2004). In a between-group design, eta squared is usually computed as:

$$\eta^2 = SS\text{Effect}/SS\text{Total} \quad (1)$$

where *SS*Effect is the sum of squares for the factor, and *SS*Total is the total sum of squares. In designs of greater complexity, however, with more than one factor, a different effect-size measure can be identified for each factor. This is the objective of the partial eta squared:

$$\eta^2_P = SS\text{Effect}/(SS\text{Effect} + SS\text{Error}) \quad (2)$$

where *SS*Error is the subjects-within-cells sum of squares, also named within-group sum of squares.

Olejnik and Algina (2003) have provided researchers with a generalized form of eta squared that integrates the two aforementioned indexes and can be used in designs with one or more measured or manipulated factors (p. 437):

$$\eta^2_G = \frac{SS\text{Effect}}{\delta * SS\text{Effect} + \sum SS\text{Measured} + \sum SS\text{Subjects/Cov}'} \quad (3)$$

where *SS*Effect refers to the sum of squares for the factor for which the effect-size statistic is computed; *SS*Measured refers to the sums of squares for all the blocking factors or interactions with blocking factors; *SS*Subjects/Cov includes the sums of squares involving subjects or covariates, equal to *SS*Error in a between-group design; $\delta = 1$ if the effect of interest is a manipulated factor; and $\delta = 0$ if it is a measured factor. Olejnik and Algina (2003) generated the adequate formulas for three-factor designs with manipulated or measured factors, and Bakeman (2005) emphasized the application of these formulas in repeated measures designs, with a single factor or multiple factors. The application of this general formula for different research hypotheses (contrast analysis, omnibus test, and post-hoc comparisons) on between-group designs is exemplified below.

One-factor between-group design

The main issue concerning effect-size measures in one-factor between-group designs has to do with the type of research hypothesis and tests: omnibus tests to prove a general hypothesis, or contrast analysis to prove hypotheses about the differences between at least two levels of the independent variable. Consequently, we must report omnibus or targeted effect-size indexes (Kelley & Preacher, 2012). The APA Publication Manual (2010) explicitly recognizes the importance of these targeted effect-size statistics by saying that "Multiple degree-of-freedom effect-size indicators are often less useful than effect-size indicators that decompose multiple degree-of-freedom tests into meaningful one degree-of-freedom effects." (p. 34)

Orthogonal contrasts on a manipulated factor

Let the focus of study be on research in Sport Psychology[1] that analyses the relationship between start reaction time (DV) and the type of feedback used (IV) while training 24 athletes: without feedback (a1: 0%); with continuous feedback (a2: 100%); with a progressive decrease in feedback from 100% to 0% (a3: progressive); and with self-regulated feedback (a4: self-regulated). Table 1 shows example data for this study.

Let us also suppose the following specific hypotheses derived from the theoretical background: 1) the reaction time will be shorter with feedback (a2, a3, and a4) than without feedback (a1), because the lack of feedback impedes learning; 2) continuous feedback (a2) will lead to a longer reaction time than intermittent feedback (a3 and a4) because the latter is more similar to the testing situation; and 3) considering the two intermittent feedbacks, the reaction time will be shorter when the feedback is planned by the trainer (a3) than by the athlete (a4). Such comparisons constitute an orthogonal group of Helmert contrasts, and there are several

Table 1
Example data [a]

Gender [b]	a1: 0%	a2: 100%	a3: progressive 50%	a4: self-regulated
Men	385	374	358	370
Men	387	375	372	371
Men	393	385	345	359
Women	375	371	359	373
Women	369	360	358	386
Women	383	373	356	367

[a] Data fabricated to illustrate calculations
[b] Gender will be used later as a second factor in a 4 × 2 factorial design. SPSS data files are available at http://personal.us.es/trigo/suppmaterials.htm

advantages to using this kind of comparisons. First of all, each contrast addresses a different and non-overlapped question. Second, since the number of comparisons in an orthogonal group is limited by the degrees of freedom of the variable, several classic authors on Experimental Designs consider it unnecessary to adjust the alpha level per contrast (e.g. Keppel, 1991; Keppel & Zedeck, 1989; Kirk, 1995). The introduction of adequate coefficients for these contrasts in SPSS (e.g. C1: 3, -1, -1, -1; C2: 0, 2, -1, -1; and C3: 0, 0, 1, -1) provides us with ANOVA and the contrast tests summarised in Table 2[2].

Application of the general formula by Olejnik and Algina (2003) to the planned contrasts on our manipulated factor ($\delta = 1$; ΣSSMeasured = 0; ΣSSSubjects/Cov = SSError) yields the usual partial eta squared formula for a contrast:

$$\eta^2_{G\text{-}C} = \frac{SS\text{Contrast}}{SS\text{Contrast} + SS\text{Error}} \quad (4)$$

The main inconvenience is that SPSS neither provides this index nor the sums of squares associated to each contrast, except for polynomials. To solve the gap, the sums of squares of the contrasts can be generated from n, coefficients (a) and means:

$$SSContrast = \frac{n\left(\Sigma a\overline{y}\right)^2}{\Sigma a^2} \quad (5)$$

For the third contrast:[3]

$$SS_{C3} = \frac{6[(0)(382) + (0)(373)\,(+1)\,(358) + (-1)(371)]^2}{(0)^2 + (0)^2 + (1)^2 + (-1)^2} = 507.00$$

or from t values and the mean squared error:

$$SS\text{Contrast} = t^2 * MS\text{Error} \quad (6)$$

$$SS_{C3} = -2.64^2 * 72.80 \approx 507.00$$

Therefore, by applying Equation 4:

$$\eta^2_{G\text{-}C3} = \frac{507.00}{507.00 + 1456.00} = .26$$

Table 2
Results of the Analysis of Variance for example data in a one-factor between-group design

Source	SS [a]	df	MS	F	t [b]	p
Between-group	1764.00	3	588.00	8.08		.001
C1: a1-a2,3,4	968.00	1	968.00		3.65	.002
C2: a2-a3,4	289.00	1	289.00		1.99	.060
C3: a3-a4	507.00	1	507.00		-2.64	.016
Within-group (Error)	1456.00	20	72.80			
Total	3220.00	23				

[a] SS for contrasts not provided by SPSS; computed from $n * (\Sigma a\overline{y})^2/\Sigma a^2$ (value of contrast in SPSS = $\Sigma a\overline{y}$). [b] The statistical tests provided by SPSS are F for general effect and Student t for contrasts

Another equivalent procedure to obtain this effect-size index for a contrast on a manipulated factor consists of computing it directly from the F or t value (Cohen, 1973):

$$\eta^2_{\text{p}} = \frac{df\text{Effect} * F\text{Effect}}{df\text{Effect} * F\text{Effect} + df\text{Error}} \quad (7)$$

$$\eta^2_{G\text{-}C3} = \frac{(1)(-2.64)^2}{(1)(-2.64)^2 + 20} = .26$$

Non-orthogonal contrasts on a manipulated factor

Our scientific hypothesis cannot always be tested by orthogonal contrasts. If that occurs, especially when the number of contrasts is larger than the degrees of freedom for the between-group source, in general, the Type-I error rate for contrast is adjusted to keep the familywise Type-I error rate within standard levels of .05 or .01. In practice, this means dividing our alpha level by the number of planned contrasts. To this end, if the factor has four different levels (df = 4 - 1 = 3) and, for example, four contrasts are planned, then an adjusted alpha level equal to .05 / 4 could be used. When performing this adjustment, Rosenthal, Rosnow, and Rubin (2000) recommend also adjusting the effect-size indexes to make them consistent with the significance level adjusted by the number of contrasts. Thus, an adjusted t or F must substitute a normal t or F in Equation 6 or 7. Let us suppose our third contrast forms part of a group of four planned contrasts. Its adjusted p value would be .016 × 4 = .064, corresponding[4] to an adjusted t (df = 20, p = .064) = -1.96. It is therefore possible to compute an adjusted sum of squares for the contrast by using equations equivalent to Equations 6 and 4 respectively with adjusted values:

$$Adj\ SS\text{Contrast} = \text{a}dj\ t^2 * MS\ \text{Error} \quad (8)$$

$$Adj\eta^2_{G\text{-}C} = \frac{AdjSS\text{Contrast}}{AdjSS\text{Contrast} + SS\text{Error}} \quad (9)$$

$$Adj\ SS\ \text{C3} = -1.96^2 * 72.80 = 279.67$$

With this modified formula, a lower generalized eta squared would be obtained than would with normal t values (.17 instead of .26; see upper part of Table 3):

$$\eta^2_{G\text{-}C3} = \frac{279.67}{279.67 + 1456.00} = .16$$

The same result would be obtained from a formula equivalent to Equation 7 with adjusted values of t:

$$Adj\ \eta^2 = \frac{df\text{Effect} * Adj\ F\text{Effect}}{df\text{Effect} * Adj\ F\text{Effect} + df\text{Error}} = $$

$$= \frac{df\text{Effect} * Adj\ t^2\text{Effect}}{df\text{Effect} * Adj\ t^2\text{Effect} + df\text{Error}} \quad (10)$$

$$\eta^2_{G\text{-}C3} = \frac{(1)(-1.96)^2}{(1)(-1.96)^2 + 20} = .16$$

Table 3
Generalized eta squared for contrast 3 depending on the research situation

Factor	Background	p^{a}	t^{b}	$\eta^2_{G-C} = \frac{SS\text{Contrast}^{c}}{SS\text{Contrast} + SS\text{Error}}$	η^2_{G-C3}
Manipulated	3 orthogonal	.016	-2.64	507.00 / (507.00 + 1456.00)	.26
	4 non-orthogonal	.064	-1.96	279.67 / (279.67 + 1456.00)	.16
	Tukey HSD	.069	-1.92	268.37 / (268.37 + 1456.00)	.16
Factor	**Background**	p^{a}	t^{b}	$\eta^2_{G-c} = \frac{SS\text{contrast}}{SS\text{Total}}$	η^2_{G-c3}
Measured	3 orthogonal	.016	-2.64	507.00 / 3220.00	.16
	4 non-orthogonal	.064	-1.96	279.67 / 3220.00	.09
	Tukey HSD	.069	-1.92	268.37 / 3220.00	.08

[a] Adjusted *p* in the last two rows (4 non-orthogonal and Tukey HSD)
[b] Adjusted *t* in the last two rows. [c] Adjusted *SS* in the last two rows

Contrasts on a measured, non-manipulated factor

One advantage of generalized eta squared is that the manipulated or measured nature of the factor is taken into account. Let us suppose that athletics have decided the training procedure in accordance with their preferences. Application of the general formula by Olejnik and Algina (2003) to the planned contrasts on our non-manipulated factor ($\delta = 0$; Σ *SS* Subjects/Cov = *SS*Error) fails to yield the usual partial eta squared for any contrast (lower-case letter c indicates a measured factor):

$$\eta^2_{G-c} = \frac{SS\text{contrast}}{SS\text{Measured} + SS\text{Error}} = \frac{SS\text{contrast}}{SS\text{Total}} \quad (11)$$

Thus, the generalized eta squared for the same contrast can be greater for a manipulated factor than for a measured factor, when it is equivalent to eta squared (see 3 orthogonal rows in Table 3):

$$\eta^2_{G-C3} = \frac{507.00}{507.00 + 1456.00} = .26; \ \eta^2_{G-c3} = \frac{507.00}{3220} = .16$$

Finally, if this contrast on a measured factor is part of four planned contrasts, the effect-size index would be even shorter, since the sum of squares of the contrast would be adjusted (see Table 3):

Adj SS C3 = $-1.96^2 * 72.80 = 279.67$; *Adj* $\eta^2_{G\text{-}C3} = 279.67/3220 = .09$

Omnibus test followed by post-hoc comparisons

An omnibus test permits general research questions to be evaluated. The objective, without a specific prediction from the theoretical background, is to discover whether there is any statistical difference, at least between the lowest and highest mean. Post-hoc multiple comparisons are developed after this omnibus test to discover whether there are any more significant differences. In this case most authors recommend applying a Bonferroni adjustment. SPSS offers multiple post-hoc tests, but not all include adjustment of the alpha level. This is the problem with the LSD (Least Significant Difference) test, although no previous warning of this drawback is given. Other tests are not recommended, since they fail to provide adequate control of the familywise Type-I error rate, such as Duncan's test (Davis & Gaito, 1984) and Student-Newman-Keuls's test (Keselman, Keselman, & Games, 1991). For this reason, these tests are excluded from certain simulation studies (e.g. Ramsey, 2002), although SPSS has yet to eliminate these obsolete procedures from its catalogue. On the other hand, Scheffé's test is too conservative, because it sets the familywise Type-I error rate for all the possible comparisons, both pair and complex, but at the same time, is not well programmed in SPSS, which only provides pair contrasts.

Nevertheless, it would be easy to compute an adequate effect-size measure for the omnibus test from Equation 1: the same for manipulated or measured factors. More difficulties are found, however, in the computation of appropriate indexes for post-hoc tests. One possibility would be to compute adjusted *t* values from the adjusted probability of the post-hoc test and then compute the adjusted sums of squares for the contrasts from Equation 8. Finally, Equation 9 can be applied for a manipulated factor, or Equation 11, with adjusted sums of squares in the numerator, for a measured factor. For example, the *p* value for a Tukey HSD test on the comparison between the third and fourth groups is .069. The adjusted[5] *t* value for this probability is $t = -1.92$. By using this new *t* value, a lower effect-size index would be obtained for a manipulated factor and an even lower effect-size index would be obtained for a measured factor (see Tukey HSD rows in Table 3).

Two-way between-group factorial design

Let us suppose the same previous data example, but also considering the gender of the participants (see Table 1), with the six items of data in each experimental condition having been obtained from three men and three women. The design is now a 4 × 2 factorial. The *GLM: Univariate* command in SPSS produces a factorial ANOVA with all the main and interactive effects. If all the factors of the design are manipulated factors, suitable generalized eta squared indexes can be obtained, equivalent to partial eta squared statistics, by requesting estimates of effect size in the *Options* button. Problems arise when measured factors are included in the design or the interest lies in planned contrasts, especially if interaction contrasts are desired.

Orthogonal contrasts on main and interaction effects

The *GLM: Univariate* command does not perform interaction contrasts. For example, by requesting Helmert contrasts, SPSS will only provide three main-effect Helmert contrasts on feedback and one main-effect contrast on gender. To obtain the three derived interaction contrasts, we must add another piece of syntax using the LMATRIX command (IV1 = feedback; IV2 = sex):

```
UNIANOVA DV BY IV1 IV2
/CONTRAST(IV)=Helmert
/LMATRIX = IV1*IV2 3 -3 -1 1 -1 1 -1 1; IV1*IV2 0 0 2 -2 -1
1 -1 1; IV1*IV2 0 0 0 0 1 -1 -1 1.
```

Table 4 shows the results provided by the package for general effects (see upper part of Table 4) and the planned contrast defined in the previous piece of syntax (see lower part of Table 4).

Table 4
Results of the Analysis of Variance for example data in a 4 × 2 factorial design

Source		SS	df	MS	F	p
Between-group		2268.00	7			
Feedback		1764.00	3	588.00	9.88	.001
Gender		80.67	1	80.67	1.36	.261
Feedback*Gender		423.33	3	141.11	2.37	.109
Within-group (Error)		952.00	16	59.50		
Total		3220.00	23			
Contrast	**Contrast estimate**	***SS*** [a]	***df***	***Standard error***	***t*** [b]	***p***
Feedback						
C1: a1-a2,3,4	14.67	968.00	1	3.64	4.03	.001
C2: a2-a3,4	8.50	289.00	1	3.86	2.20	.043
C3: a3-a4	-13.00	507.00	1	4.45	-2.92	.010
Gender	3.67	80.67	1	3.15	1.16	.261
Feedback*Gender						
C1: (a1-a2,3,4)(b1-b2)	36.00	162	1	21.82	1.65	.118
C2: (a2-a3,4)(b1-b2)	28.00	196	1	15.43	1.81	.088
C3: (a3-a4)(b1-b2)	9.33	65.33	1	8.91	1.05	.310

[a] Not provided by SPSS; computed from $n * (\Sigma a\bar{y})^2 / \Sigma a^2$. [b] Not provided by SPSS; computed from Contrast estimate / Standard error

In order to obtain generalized eta squared indexes for each contrast, appropriate sums of squares must be obtained for each contrast from Equations 5 or 6 (see lower part of Table 4). The formulas by Olejnik and Algina (2003) can then be adapted to contrasts instead of general effects, and to two-factor designs instead of three-factor designs. As can be observed in Table 5, the general rule is to include in the denominator of the formula: the error sum of squares, the sum of squares for the contrast of interest, and the sums of squares for the remaining general or contrast effects on measured factors. Table 5 also exemplifies the application of these formulas for main-effect contrast 3 on A and interaction contrast 3 depending on the manipulated or measured nature of the factors. As can be observed in these formulas and examples, the maximum value of generalized eta squared for contrasts on A or interaction with A is obtained when both factors are manipulated. This value decreases even when the measured factor is only the other factor, B, it decreases even more when the measured factor is only A, and takes the minimum value when both A and B are measured factors.

Non-orthogonal or post-hoc comparisons on main and interaction effects

When comparisons are developed after an omnibus test, the first problem with SPSS is the lack of post-hoc tests for the interaction effect. That problem can be solved using the same command as for planned contrasts with a Bonferroni adjustment. This is the same strategy to be used when planning non-orthogonal contrasts, especially if they exceed the number of degrees of freedom of the factor. The second problem involves combining the use of the adjusted sum of squares with the use of different denominators depending on the manipulated or measured nature of the factors. The proposal here is to use the same equations previously compiled for contrast analysis in Table 5, but with adjusted *t* values derived from the adjusted *p* values. Once again, the adjusted sum of squares can be obtained by means of Equation 8.

Discussion

Olejnik and Algina (2003) pointed out that the characteristics of the designs must be taken into account in order to provide comparable effect-size indexes across a variety of designs. We have pointed out that also the kind of hypothesis to be tested must be taken into account. The power of a study can be increased not only by introducing blocking variables, but also via the planning of contrast analysis, especially in an experimental context. Even when planned contrasts are not possible, appropriate effect-size indexes must also be reported. In a small percentage of articles, authors calculate effect-size indexes for post-hoc contrasts, usually reporting partial eta squared for the factor and eta squared or Cohen's *d* for post-hoc comparisons (Fritz et al., 2011). This practice can overestimate the effect size because generalized eta squared in post-hoc comparisons would involve eta squared when the factor is measured and partial eta squared when the factor is manipulated. Additionally, in accordance with Rosenthal et al. (2000), we have proposed an adjustment of the effect-size index for post-hoc comparisons (also for planned contrast with a Bonferroni adjustment) that further reduces the effect size.

Rosnow and Rosenthal (1996) distinguish between ways to increase the power that can be used only by original researchers, and ways to increase power and statistical validity of a published study that can be used by research consumers when the original

Table 5
Generalized eta squared formulas for main-effect and interaction contrast 3 in a 4 × 2 factorial design (adapted from Olejnik and Algina, 2003, p. 438)

Design	Contrasts (C or c) [a]	Contrast 3
	C on A / c on a	**C3 on A / c3 on a**
AB	$\frac{SSC}{SSC + SS\text{Error}}$	$\frac{507.00}{507.00 + 952.00} = .35$
Ab	$\frac{SSC}{SSC + SS\text{b} + SS\text{Ab} + SS\text{Error}}$	$\frac{507.00}{507.00 + 80.67 + 423.33 + 952.00} = .26$
aB	$\frac{SS\text{c}}{SS\text{a} + SS\text{aB} + SS\text{Error}}$	$\frac{507.00}{1764.00 + 423.33 + 952.00} = .16$
ab	$\frac{SS\text{c}}{SS\text{a} + SS\text{b} + SS\text{ab} + SS\text{Error}}$	$\frac{507.00}{1764.00 + 80.67 + 423.33 + 952.00} = .1$
	C on AB or Ab / c on aB or ab	**C3 on AB or Ab / c3 on aB or ab**
AB	$\frac{SSC}{SSC + SS\text{Error}}$	$\frac{65.33}{63.33 + 952.00} = .06$
Ab	$\frac{SSC}{SS\text{b} + SS\text{Ab} + SS\text{Error}}$	$\frac{65.33}{80.67 + 423.33 + 952.00} = .04$
aB	$\frac{SS\text{c}}{SS\text{a} + SS\text{aB} + SS\text{Error}}$	$\frac{65.33}{1764.00 + 423.33 + 952.00} = .02$
ab	$\frac{SS\text{c}}{SS\text{a} + SS\text{b} + SS\text{ab} + SS\text{Error}}$	$\frac{65.33}{1764.00 + 80.67 + 423.33 + 952.00} = .02$

[a] Upper-case letters indicate a manipulated factor; lower-case letters indicate a measured factor

researchers fail to complete the task. The latter means of increasing power would involve performing contrast and power analysis based on the reported results. To this end, the formulas and procedures recovered in this report are based on sums of squares of effects or contrasts, easily obtained from the means and number of observations usually reported; and/or in sums of squares of error, that can be derived from means, *F* values, and degrees of freedom. However, professional psychologists and researchers would be even more interested in statistical packages that include in their computer programming these indexes for multiple contrasts. “Authors of statistical software packages have a responsibility to assist researchers in following the best practices in statistics” (Kirk, 2001, p. 216), although, as we have learned so far, “these things take time” (Cohen, 1990, p. 1311). Unfortunately, neither general statistical packages nor statistical software for Meta-Analysis have incorporated indexes that are comparable across designs and type of hypothesis in the form of generalized eta squared. While they decide to modify its outputs, we have also considered it necessary to facilitate computations with several spreadsheets[6].

Acknowledgements

We would like to thank our colleague, Carlos Camacho, who encouraged us to publish this report after reading its draft. Any errors or shortcomings are our responsibility.

Footnotes

[1] Based on the topic of study by Zubiaur, Oña, and Delgado (1998).
[2] Output files are available at http://personal.us.es/trigo/suppmaterials.htm.
[3] The third contrast will be used to exemplify all calculations.
[4] Obtained from http://statpages.org/pdfs.html for $df = 20$ and $p = .060$.
[5] Obtained from http://statpages.org/pdfs.html for $df = 20$ and adjusted $p = .069$.
[6] Spreadsheets are available at http://personal.us.es/trigo/suppmaterials.htm.

References

American Psychological Association (1994). *Publication Manual of the American Psychological Association* (4th ed.). Washington: American Psychological Association.

American Psychological Association (2001). *Publication Manual of the American Psychological Association* (5th ed.). Washington: American Psychological Association.

American Psychological Association (2010). *Publication Manual of the American Psychological Association* (6th ed.). Washington: American Psychological Association.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37*, 379-384.

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.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

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

Davis, C., & Gaito, J. (1984). Multiple comparison procedures within experimental research. *Canadian Psychology/Psychologie Canadienne, 25*, 1-13.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*, 532-538.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2011). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*, 2-18.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*, 137-152.

Keppel, G. (1991). *Design and analysis: A research handbook* (3ª ed.). Upper Saddle River, NJ: Prentice-Hall.

Keppel, G., & Zedeck, S. (1989). *Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches*. New York: W. H. Freeman & Company.

Keselman, H. J., Keselman, J. C., & Games, P. A. (1991). Maximum familywise Type I error rate: The least significant difference, Newman-Keuls, and other multiple comparison procedures. *Psychological Bulletin, 110*, 155-161.

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

Kirk, R. E. (1995). *Experimental designs: Procedures for the behavioural sciences* (3ª ed.). Pacific Grove, CA: Brooks/Cole.

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

Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational & Psychological Measurements, 61*, 213-218.

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

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8*, 434-447.

Pierce, Ch. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement, 64*, 916-924.

Ramsey, P. H. (2002). Comparison of closed testing procedures for pairwise testing of means. *Psychological Methods, 7*(4), 504-523.

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

Rosnow, R. L. & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people’s published data: General procedures for research consumers. *Psychological Methods, 1*, 331-340.

Rosnow, R. L., & Rosenthal, R. (2009). Effect sizes: Why, when, and how to use them. *Zeitschrift Für Psychologie/Journal of Psychology, 217*, 6-14.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*, 309-316.

Sun, S., Pan, W., & Wang, L. L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *Journal of Educational Psychology, 102*(4), 989-1004.

Valera, A., Sánchez Meca, J., & Marín Martínez, F. (2000). Contraste de hipótesis e investigación psicológica española: análisis y propuestas [Hypothesis testing and Spanish psychological research: Analyses and proposals]. *Psicothema, 12*, 549-552.

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