

Recommended effect size statistics for repeated measures designs

ROGER BAKEMAN
Georgia State University, Atlanta, Georgia

Investigators, who are increasingly implored to present and discuss effect size statistics, might comply more often if they understood more clearly what is required. When investigators wish to report effect sizes derived from analyses of variance that include repeated measures, past advice has been problematic. Only recently has a generally useful effect size statistic been proposed for such designs: generalized eta squared (η_G^2 ; Olejnik & Algina, 2003). Here, we present this method, explain that η_G^2 is preferred to eta squared and partial eta squared because it provides comparability across between-subjects and within-subjects designs, show that it can easily be computed from information provided by standard statistical packages, and recommend that investigators provide it routinely in their research reports when appropriate.

Increasingly, investigators are implored to compute, present, and discuss effect size statistics as a routine part of any empirical report (American Psychological Association, 2001; Wilkinson & the Task Force on Statistical Inference, 1999), yet they lag in compliance (Fidler, Thomason, Cumming, Finch, & Leeman, 2004) and discussion of effect sizes in reports of psychological research is not yet widespread. Many more investigators would probably include reports of effect sizes if they understood clearly which statistics should be provided. With respect to one common data-analytic approach—the analysis of variance (ANOVA) when designs include repeated measures—there is some reason for confusion. Most of the advice given in textbooks is problematic or limited in ways that I will describe shortly, and only recently has a more generally useful effect size statistic for such cases been proposed (Olejnik & Algina, 2003).

In this article, I describe generalized eta squared (η_G^2) as defined by Olejnik and Algina (2003). I endorse their recommendation that it be used as an effect size statistic whenever ANOVAs are used, define under what circumstances it varies from eta squared (η^2) and partial eta squared (η_p^2), and show how it can be computed easily with the information provided by standard statistical packages such as SPSS. Olejnik and Algina showed how η_G^2 applies to between-subjects designs, analyses of covariance, repeated measures designs, and mixed designs in general. In contrast, I emphasize application just to designs that include repeated measures. My purposes are to explain why η_G^2 is particularly important for such de-

signs (with such designs, differences among η_G^2 , η_p^2 , and η^2 can be quite pronounced) and to make its computation easily accessible to users of such designs.

Before I describe η_G^2 in more detail, some general comments regarding limitations are in order. Not all investigations lend themselves to proportion of variance effect size measures or other standardized effect size measures. As Bond, Wiitala, and Richard (2003) argue, when measurement units (e.g., reaction time in seconds) are clearly understood, raw mean differences may be preferred. Moreover, η_G^2 assumes a traditional univariate ANOVA approach, as opposed to multivariate or multilevel approaches to designs that include repeated measures. Finally, η^2 and similar measures are often omnibus tests, which some scholars regard as insufficiently focused to capture researchers' real concerns (Rosenthal, Rosnow, & Rubin, 2000).

At first glance, the matter of effect sizes in an ANOVA context seems straightforward. ANOVAs partition a total sum of squares into portions associated with the various effects identified by the design. These portions can be arranged hierarchically, and the various effects can be tested sequentially (Cohen & Cohen, 1983). In the ANOVA literature, the effect size statistic is usually called *eta squared* (η^2) and is the ratio of effect to total variance. Thus,

$$\eta^2 = \frac{SS_{\text{effect}}}{SS_{\text{total}}}, \quad (1)$$

which is identical to the change in R^2 of hierarchic multiple regression:

$$\Delta R^2 = \frac{R^2_{\text{change}}}{R^2_{\text{total}}}. \quad (2)$$

η^2 , or the change in R^2 , seems fine in the context of many hierarchic regressions or with a one-way ANOVA, but

My appreciation to my colleague, Christopher Henrich, and to two anonymous reviewers who provided useful and valuable feedback on earlier versions of this article. Correspondence concerning this article should be addressed to R. Bakeman, Department of Psychology, Georgia State University, P.O. Box 5010, Atlanta, GA 30302-5010 (e-mail: bakeman@gsu.edu).

may become problematic when there is more than one source of variance in an ANOVA.

For example, imagine that we computed η^2 for factor A, first in a one-way and then in a two-way factorial ANOVA. In the second case, factor B may contribute nonrandom differences to SS_{total} ; thus, the two η^2 s may not be comparable. Usually, η_p^2 is presented as a solution to the comparability problem (Keppel, 1991; Tabachnick & Fidell, 2001). In the context of factorial designs, η_p^2 removes the effect of other factors from the denominator; thus,

$$\eta_p^2 = \frac{SS_{\text{effect}}}{SS_{\text{effect}} + SS_{s/\text{cells}}}. \quad (3)$$

With a simple $A \times B$ factorial design, the total sum of squares is partitioned into four components: SS_A , SS_B , SS_{AB} , and the error term, or $SS_{s/AB}$ (i.e., subjects nested within cells defined by crossing factors A and B). Thus, the η_p^2 denominators for the A effect, the B effect, and the AB interaction would be $SS_A + SS_{s/AB}$, $SS_B + SS_{s/AB}$, and $SS_{AB} + SS_{s/AB}$, respectively. In recognition of the arguments for η_p^2 , current versions of SPSS provide η_p^2 and not η^2 when effect size statistics are requested.

With a simple repeated measures design (i.e., one with no between-subjects variables and only one within-subjects variable), the total sum of squares is partitioned into three components: SS_s , SS_p , and SS_{ps} , where SS_s is the sum of squares between subjects. (I use A, B, C, etc. for between-subjects factors; P, Q, R, etc. for within-subjects factors; and s for the subjects "factor.") Here, SS_s reflects the proportion of total score variance that can be accounted for by knowing the particular subject, which will be larger the more the repeated scores correlate within subjects. As students are typically taught, a design with one repeated measures factor can be regarded as a two-factor design, with subjects as the other "factor"; each subject represents a level of the factor, and the mean square for error for the repeated measure factor is thus the $P \times$ subjects interaction, or SS_{ps} divided by its degrees of freedom. The more scores correlate within subjects, the smaller this error term will be, which gives repeated measures designs their reputation for increased power (Bakeman, 1992; Bakeman & Robinson, 2005).

Assume the repeated measures factor is age, as it would be in a longitudinal design. If the only factor is age, its effect size per η^2 would be the ratio of SS_p to the sum of SS_s , SS_p , and SS_{ps} (i.e., SS_{total}), but its effect size per η_p^2 would be the ratio of SS_p to the sum of SS_p and SS_{ps} ; that is, for η_p^2 the effect of subject would be removed from the denominator. In this case, η_p^2 would be larger than η^2 , which, from a "bigger-is-better" standpoint, might seem desirable; nevertheless, the "correction" of η_p^2 renders the effect size for age from a longitudinal design not comparable with a similar effect size for age from a cross-sectional design.

The issue of comparability of effect sizes derived from studies with similar factors but different designs (i.e., cases in which a factor is between subjects in one study

and within subjects in another) has been addressed recently by Olejnik and Algina (2003). They propose an η_G^2 (described in detail in the next few paragraphs): an effect size statistic that provides comparable estimates for an effect even when designs vary. Such comparability is important not just for meta-analyses but also for the less formal literature reviews incorporated in the introductory sections of our research reports when we want to compare the strength of a particular factor on a specific outcome across studies, even though the designs in which the variables are embedded may vary. Comparability across studies is precisely what we want in our literature; thus, η_G^2 seems clearly the best choice for an effect size statistic when ANOVAs are at issue. Still, no statistic can solve all problems. As Olejnik and Algina wrote, no effect size measures can be comparable when different populations are sampled. Thus, estimates of age effects in studies on 3- to 7-year-old children, for example, cannot be compared to similar estimates when 8- to 11-year-old children are studied. But, other things being equal, η_G^2 assures comparability of effect sizes when factors are included in different designs, regardless of whether a given factor is between subjects or within subjects.

η_G^2 differs from η^2 and from η_p^2 in its denominator. Whereas η^2 includes all component sums of squares in its denominator, η_p^2 and η_G^2 include only some of them, although typically η_G^2 includes more than η_p^2 . Computation of η_G^2 is based on the point of view that variability in the data arises from two sources: The first includes factors manipulated in the study, and the second includes individual differences (Olejnik & Algina, 2003; as Gillett, 2003 [p. 421] noted, individual difference factors are also called *stratified* by Glass, McGaw, & Smith, 1981; *organismic* by Cohen & Cohen, 1983; and *classification* by Winer, Brown, & Michaels, 1991). As Olejnik and Algina note, individual differences (i.e., measured factors) can be attributable to stable or transient characteristics of participants, such as an individual's gender or motivational state, and research designs can vary in the extent to which sources of individual differences are estimated or controlled. However, for comparability across between-subjects and within-subjects designs, variation due to individual differences (which was not created by the investigator) should remain in the denominator, whereas variation due to manipulated factors (which was created by the investigator) should be removed when effect sizes are being estimated.

The effect size parameter defined by Olejnik and Algina (2003) is

$$\frac{\sigma_{\text{effect}}^2}{\delta \times \sigma_{\text{effect}}^2 + \sigma_{\text{measured}}^2}, \quad (4)$$

where $\sigma_{\text{measured}}^2$ includes variance due to individual differences and where $\delta = 0$ if the effect involves one or more measured factors, either singly or combined in an interaction (they would already be included in $\sigma_{\text{measured}}^2$) and $\delta = 1$ if the effect involves only manipulated factors. Thus, the denominator for the effect size parameter in-

cludes all sources of variance that involve a measured factor and excludes all that involve only manipulated factors. The exceptions are cases in which the effect involved only manipulated factors, in which case its associated variance is also included in the denominator. η_G^2 is then estimated as

$$\eta_G^2 = \frac{SS_{\text{effect}}}{\delta \times SS_{\text{effect}} + \sum_{\text{measured}} SS_{\text{measured}}} \quad (5)$$

Specifically, for designs that include repeated measures, not just sums of squares for subjects, but also those for all subjects \times repeated measures factor interactions would be included in the denominator (see Tables 1–3).

Olejnik and Algina (2003) introduced a useful convention. They represented manipulated factors with uppercase letters (A, B, C, etc.) and individual difference factors with lowercase letters (a, b, c, etc.). I have introduced the additional convention that within-subjects factors be represented with the letters P, Q, R, and onward. These would not appear as lowercase letters; because they are associated with times or situations that the investigator has chosen for assessment and not with individual measures on which participants might be blocked, repeated measures factors are regarded as manipulated by definition. The subjects factor, whose interactions serve as error terms in within-subjects designs, is not manipulated, and so appears as a lowercase *s*. Thus, not just designs, but all their component sums of squares can be represented economically with the correct combination of upper- and lowercase letters. Several examples of simple designs that include repeated measures are given in Tables 1–3, along with the component sums of squares

that are included in each design and the formulas for the *F* ratios, η_p^2 s, and η_G^2 s associated with their effects.

The uppercase manipulated, lowercase measured convention permits Equation 5 to be restated as follows: The denominator for η_G^2 contains all sums of squares whose representations contain a lowercase letter or letters (alternatively, the denominator is the total sum of squares reduced by any sums of squares whose representation consists only of uppercase letters), with the added proviso that the sum of squares for the effect must be included in the denominator (again, see Tables 1–3). An implication of Equation 5 is that, when designs include repeated measures, η_G^2 will be smaller than η_p^2 ; as was noted earlier, the denominator for η_G^2 is larger because it includes sums of squares for subjects and for all subjects \times repeated measures factor interactions. For example, with two repeated measures (a PQ design), the denominator for the P effect is $SS_P + SS_{P_s}$ for η_p^2 but the sum of SS_P , SS_s , SS_{P_s} , SS_{Q_s} , and SS_{PQ_s} (i.e., $SS_T - SS_Q - SS_{PQ}$) for η_G^2 (see Table 1). η_p^2 removes other sources of individual differences, which, as Olejnik and Algina (2003) argue, renders it not directly comparable across studies with between-subjects and within-subjects designs.

η_G^2 is easily computed from information given by standard statistical packages such as SPSS. For example, the general linear-model repeated measures procedure in SPSS provides sums of squares for any between-subjects effect, any between-subjects error, and all within-subjects effects, including error (i.e., interactions with the subjects factor), from which η_G^2 can be computed (specify Type I, or sequential, sum of squares if any between-subjects groups have unequal *ns*). For example, Adamson, Bake- man, and Deckner (2004) analyzed percentages of times

Table 1
Computation of Partial Eta Squared (η_p^2) and Generalized Eta Squared (η_G^2) for A, P, AP, aP, and PQ Designs

Design	Effect	<i>F</i> Ratio	η_p^2	η_G^2
A	SS_A	$MS_A/MS_{s/A}$	$SS_A/(SS_A + SS_{s/A})$	$SS_A/(SS_A + SS_{s/A})$
	$SS_{s/A}$	—	—	—
P	SS_s	—	—	—
	SS_P	MS_P/MS_{P_s}	$SS_P/(SS_P + SS_{P_s})$	$SS_P/(SS_P + SS_s + SS_{P_s})$
	SS_{P_s}	—	—	—
AP	SS_A	$MS_A/MS_{s/A}$	$SS_A/(SS_A + SS_{s/A})$	$SS_A/(SS_A + SS_{s/A} + SS_{P_s/A})$
	$SS_{s/A}$	—	—	—
	SS_P	$MS_P/MS_{P_s/A}$	$SS_P/(SS_P + SS_{P_s/A})$	$SS_P/(SS_P + SS_{s/A} + SS_{P_s/A})$
	$SS_{P_s/A}$	$MS_{P_s/A}/MS_{P_s/A}$	$SS_{P_s/A}/(SS_{P_s/A} + SS_{P_s/A})$	$SS_{P_s/A}/(SS_{P_s/A} + SS_{s/A} + SS_{P_s/A})$
aP	$SS_{P_s/A}$	—	—	—
	SS_a	$MS_a/MS_{s/a}$	$SS_a/(SS_a + SS_{s/a})$	$SS_a/(SS_a + SS_{s/a} + SS_{P_s/A} + SS_{P_s/a})$
	$SS_{s/a}$	—	—	—
	SS_P	$MS_P/MS_{P_s/a}$	$SS_P/(SS_P + SS_{P_s/a})$	$SS_P/(SS_P + SS_a + SS_{s/a} + SS_{P_s/A} + SS_{P_s/a})$
	$SS_{P_s/a}$	$MS_{P_s/a}/MS_{P_s/a}$	$SS_{P_s/a}/(SS_{P_s/a} + SS_{P_s/a})$	$SS_{P_s/a}/(SS_{P_s/a} + SS_a + SS_{s/a} + SS_{P_s/A} + SS_{P_s/a})$
PQ	$SS_{P_s/A}$	—	—	—
	SS_s	—	—	—
	SS_P	MS_P/MS_{P_s}	$SS_P/(SS_P + SS_{P_s})$	$SS_P/(SS_P + SS_s + SS_{P_s} + SS_{Q_s} + SS_{PQ_s})$
	SS_{P_s}	—	—	—
	SS_Q	MS_Q/MS_{Q_s}	$SS_Q/(SS_Q + SS_{Q_s})$	$SS_Q/(SS_Q + SS_s + SS_{P_s} + SS_{Q_s} + SS_{PQ_s})$
	SS_{Q_s}	—	—	—
	SS_{PQ}	MS_{PQ}/MS_{PQ_s}	$SS_{PQ}/(SS_{PQ} + SS_{PQ_s})$	$SS_{PQ}/(SS_{PQ} + SS_s + SS_{P_s} + SS_{Q_s} + SS_{PQ_s})$
SS_{PQ_s}	—	—	—	

Note—“A” represents a manipulated between-subjects factor, “a” represents a measured between-subjects factor, “P” and “Q” represent within-subjects factors, and “s” represents the subjects factor.

Table 2
Computation of Partial Eta Squared (η_p^2) and Generalized Eta Squared (η_G^2) for ABP, aBP, and abP Designs

Design	Effect	F Ratio	η_p^2	η_G^2	
ABP	SS_A	$MS_A/MS_{s/AB}$	$SS_A/(SS_A + SS_{s/AB})$	$SS_A/(SS_A + SS_{s/AB} + SS_{Ps/AB})$	
	SS_B	$MS_B/MS_{s/AB}$	$SS_B/(SS_B + SS_{s/AB})$	$SS_B/(SS_B + SS_{s/AB} + SS_{Ps/AB})$	
	SS_{AB}	$MS_{AB}/MS_{s/AB}$	$SS_{AB}/(SS_{AB} + SS_{s/AB})$	$SS_{AB}/(SS_{AB} + SS_{s/AB} + SS_{Ps/AB})$	
	$SS_{s/AB}$	—	—	—	
	SS_P	$MS_P/MS_{Ps/AB}$	$SS_P/(SS_P + SS_{Ps/AB})$	$SS_P/(SS_P + SS_{s/AB} + SS_{Ps/AB})$	
	SS_{PA}	$MS_{PA}/MS_{Ps/AB}$	$SS_{PA}/(SS_{PA} + SS_{Ps/AB})$	$SS_{PA}/(SS_{PA} + SS_{s/AB} + SS_{Ps/AB})$	
	SS_{PB}	$MS_{PB}/MS_{Ps/AB}$	$SS_{PB}/(SS_{PB} + SS_{Ps/AB})$	$SS_{PB}/(SS_{PB} + SS_{s/AB} + SS_{Ps/AB})$	
	SS_{PAB}	$MS_{PAB}/MS_{Ps/AB}$	$SS_{PAB}/(SS_{PAB} + SS_{Ps/AB})$	$SS_{PAB}/(SS_{PAB} + SS_{s/AB} + SS_{Ps/AB})$	
	$SS_{Ps/AB}$	—	—	—	
	aBP	SS_a	$MS_a/MS_{s/aB}$	$SS_a/(SS_a + SS_{s/aB})$	$SS_a/(SS_a + SS_{aB} + SS_{s/aB} + SS_{Pa} + SS_{PaB} + SS_{Ps/aB})$
		SS_B	$MS_B/MS_{s/aB}$	$SS_B/(SS_B + SS_{s/aB})$	$SS_B/(SS_B + SS_a + SS_{aB} + SS_{s/aB} + SS_{Pa} + SS_{PaB} + SS_{Ps/aB})$
		SS_{aB}	$MS_{aB}/MS_{s/aB}$	$SS_{aB}/(SS_{aB} + SS_{s/aB})$	$SS_{aB}/(SS_{aB} + SS_a + SS_{s/aB} + SS_{Pa} + SS_{PaB} + SS_{Ps/aB})$
$SS_{s/aB}$		—	—	—	
SS_P		$MS_P/MS_{Ps/aB}$	$SS_P/(SS_P + SS_{Ps/aB})$	$SS_P/(SS_P + SS_a + SS_{aB} + SS_{s/aB} + SS_{Pa} + SS_{PaB} + SS_{Ps/aB})$	
SS_{Pa}		$MS_{Pa}/MS_{Ps/aB}$	$SS_{Pa}/(SS_{Pa} + SS_{Ps/aB})$	$SS_{Pa}/(SS_{Pa} + SS_a + SS_{aB} + SS_{s/aB} + SS_{PaB} + SS_{Ps/aB})$	
SS_{PB}		$MS_{PB}/MS_{Ps/aB}$	$SS_{PB}/(SS_{PB} + SS_{Ps/aB})$	$SS_{PB}/(SS_{PB} + SS_a + SS_{aB} + SS_{s/aB} + SS_{Pa} + SS_{PaB} + SS_{Ps/aB})$	
SS_{PAB}		$MS_{PAB}/MS_{Ps/aB}$	$SS_{PAB}/(SS_{PAB} + SS_{Ps/aB})$	$SS_{PAB}/(SS_{PAB} + SS_a + SS_{aB} + SS_{s/aB} + SS_{Pa} + SS_{Ps/aB})$	
$SS_{Ps/aB}$		—	—	—	
abP		SS_a	$MS_a/MS_{s/ab}$	$SS_a/(SS_a + SS_{s/ab})$	$SS_a/(SS_a + SS_b + SS_{ab} + SS_{s/ab} + SS_{Pa} + SS_{Pb} + SS_{Pab} + SS_{Ps/ab})$
		SS_b	$MS_b/MS_{s/ab}$	$SS_b/(SS_b + SS_{s/ab})$	$SS_b/(SS_b + SS_a + SS_{ab} + SS_{s/ab} + SS_{Pa} + SS_{Pb} + SS_{Pab} + SS_{Ps/ab})$
		SS_{ab}	$MS_{ab}/MS_{s/ab}$	$SS_{ab}/(SS_{ab} + SS_{s/ab})$	$SS_{ab}/(SS_{ab} + SS_a + SS_b + SS_{s/ab} + SS_{Pa} + SS_{Pb} + SS_{Pab} + SS_{Ps/ab})$
	$SS_{s/ab}$	—	—	—	
	SS_P	$MS_P/MS_{Ps/ab}$	$SS_P/(SS_P + SS_{Ps/ab})$	$SS_P/(SS_P + SS_a + SS_b + SS_{ab} + SS_{s/ab} + SS_{Pa} + SS_{Pb} + SS_{Pab} + SS_{Ps/ab})$	
	SS_{Pa}	$MS_{Pa}/MS_{Ps/ab}$	$SS_{Pa}/(SS_{Pa} + SS_{Ps/ab})$	$SS_{Pa}/(SS_{Pa} + SS_a + SS_b + SS_{ab} + SS_{s/ab} + SS_{Pb} + SS_{Pab} + SS_{Ps/ab})$	
	SS_{Pb}	$MS_{Pb}/MS_{Ps/ab}$	$SS_{Pb}/(SS_{Pb} + SS_{Ps/ab})$	$SS_{Pb}/(SS_{Pb} + SS_a + SS_b + SS_{ab} + SS_{s/ab} + SS_{Pa} + SS_{Pab} + SS_{Ps/ab})$	
	SS_{Pab}	$MS_{Pab}/MS_{Ps/ab}$	$SS_{Pab}/(SS_{Pab} + SS_{Ps/ab})$	$SS_{Pab}/(SS_{Pab} + SS_a + SS_b + SS_{ab} + SS_{s/ab} + SS_{Pa} + SS_{Pb} + SS_{Ps/ab})$	
	$SS_{Ps/ab}$	—	—	—	

Note—"A" and "B" represent manipulated between-subjects factors, "a" and "b" represent measured between-subjects factors, "P" represents a within-subjects factor, and "s" represents the subjects factor.

infants and mothers spent in a state they call *symbol-infused supported joint engagement* using a mixed sex \times age design. For this aP design (see Table 1), where a = infant's sex and P = infant's age, the sums of squares computed by SPSS were 23.7 and 195.8 for the a and s/a between-subjects terms, respectively, and 310.0, 4.0, and 131.6 (rounded) for the P, Pa, and Ps/A within-subjects terms, respectively. Thus, for the sex effect $\eta_p^2 = 23.7/(23.7 + 195.8) = .11$ and $\eta_G^2 = 23.7/(23.7 + 195.8 + 4.0 + 131.6) = .07$, whereas for the age effect $\eta_p^2 = 310.0/(310.0 + 131.6) = .70$ and $\eta_G^2 = 310.0/(310.0 + 23.7 + 195.8 + 4.0 + 131.6) = .47$. As was expected, η_G^2 gives smaller values than η_p^2 , but both indexes indicate that the age effect was more than six times stronger than the sex effect.

It is more difficult to recover effect size statistics from already published articles, a situation that rightly troubles meta-analysts. Usually, investigators provide *F* ratios, from which η_p^2 but not η_G^2 can be computed:

$$\eta_p^2 = \frac{F \times df_{\text{effect}}}{F \times df_{\text{effect}} + df_{\text{error}}} \quad (6)$$

If all factors are manipulated between subjects, there is no problem. In such cases, η_G^2 is identical to η_p^2 , which can be computed from *F* ratios. If all factors are between subjects but only some are manipulated, then computa-

tion of η_G^2 requires that the investigator provide an ANOVA of source table—a practice that, unfortunately, seems less common now than it was in the past.

The greatest number of difficulties occur when designs include one repeated measure or more. Even when investigators include a source table, often the table includes only within-subjects and not between-subjects information. A total sum of squares could be computed if a standard deviation for all scores were given, but almost always means and standard deviations are given separately for the repeated measures, if they are provided at all. In such cases, investigators who wish to compare their effect sizes with those of other studies in the literature may have no choice but to rely on η_p^2 computed from *F* ratios. In some cases—and assuming properly qualified interpretation—this may not be a bad alternative if it is the only one available.

After all, there is no absolute meaning associated with either η_p^2 or η_G^2 (except for zero values); their values gain meaning in relation to the findings of other, similar studies. If we focus on a pair of variables (e.g., the effect of age on receptive language or externalizing behavior) and compare effect sizes across studies with similar designs, it would not much matter which effect size statistic we used. However, if we want to apply common guidelines or if designs vary across studies, then η_G^2 is the effect size statistic of choice (S. Olejnik, personal communication,

Table 3
Computation of Partial Eta Squared (η_p^2) and Generalized Eta Squared (η_G^2) for APQ and aPQ Designs

Design	Effect	F Ratio	η_p^2	η_G^2
APQ	SS_A	$MS_A/MS_{s/A}$	$SS_A/(SS_A + SS_{s/A})$	$SS_A/(SS_A + SS_{s/A} + SS_{P/s/A} + SS_{Q/s/A} + SS_{PQ/s/A})$
	$SS_{s/A}$	—	—	—
	SS_P	$MS_P/MS_{P/s/A}$	$SS_P/(SS_P + SS_{P/s/A})$	$SS_P/(SS_P + SS_{s/A} + SS_{P/s/A} + SS_{Q/s/A} + SS_{PQ/s/A})$
	$SS_{P/A}$	$MS_{P/A}/MS_{P/s/A}$	$SS_{P/A}/(SS_{P/A} + SS_{P/s/A})$	$SS_{P/A}/(SS_{P/A} + SS_{s/A} + SS_{P/s/A} + SS_{Q/s/A} + SS_{PQ/s/A})$
	$SS_{P/s/A}$	—	—	—
	SS_Q	$MS_Q/MS_{Q/s/A}$	$SS_Q/(SS_Q + SS_{Q/s/A})$	$SS_Q/(SS_Q + SS_{s/A} + SS_{P/s/A} + SS_{Q/s/A} + SS_{PQ/s/A})$
	$SS_{Q/A}$	$MS_{Q/A}/MS_{Q/s/A}$	$SS_{Q/A}/(SS_{Q/A} + SS_{Q/s/A})$	$SS_{Q/A}/(SS_{Q/A} + SS_{s/A} + SS_{P/s/A} + SS_{Q/s/A} + SS_{PQ/s/A})$
	$SS_{Q/s/A}$	—	—	—
	SS_{PQ}	$MS_{PQ}/MS_{PQ/s/A}$	$SS_{PQ}/(SS_{PQ} + SS_{PQ/s/A})$	$SS_{PQ}/(SS_{PQ} + SS_{s/A} + SS_{P/s/A} + SS_{Q/s/A} + SS_{PQ/s/A})$
	$SS_{PQ/A}$	$MS_{PQ/A}/MS_{PQ/s/A}$	$SS_{PQ/A}/(SS_{PQ/A} + SS_{PQ/s/A})$	$SS_{PQ/A}/(SS_{PQ/A} + SS_{s/A} + SS_{P/s/A} + SS_{Q/s/A} + SS_{PQ/s/A})$
	$SS_{PQ/s/A}$	—	—	—
aPQ	SS_a	$MS_a/MS_{s/a}$	$SS_a/(SS_a + SS_{s/a})$	$SS_a/(SS_a + SS_{s/a} + SS_{P/a} + SS_{P/s/a} + SS_{Q/a} + SS_{Q/s/a} + SS_{PQ/a} + SS_{PQ/s/a})$
	$SS_{s/a}$	—	—	—
	SS_P	$MS_P/MS_{P/s/a}$	$SS_P/(SS_P + SS_{P/s/a})$	$SS_P/(SS_P + SS_a + SS_{s/a} + SS_{P/a} + SS_{P/s/a} + SS_{Q/a} + SS_{Q/s/a} + SS_{PQ/a} + SS_{PQ/s/a})$
	$SS_{P/a}$	$MS_{P/a}/MS_{P/s/a}$	$SS_{P/a}/(SS_{P/a} + SS_{P/s/a})$	$SS_{P/a}/(SS_{P/a} + SS_a + SS_{s/a} + SS_{P/s/a} + SS_{Q/a} + SS_{Q/s/a} + SS_{PQ/a} + SS_{PQ/s/a})$
	$SS_{P/s/a}$	—	—	—
	SS_Q	$MS_Q/MS_{Q/s/a}$	$SS_Q/(SS_Q + SS_{Q/s/a})$	$SS_Q/(SS_Q + SS_a + SS_{s/a} + SS_{P/a} + SS_{P/s/a} + SS_{Q/a} + SS_{Q/s/a} + SS_{PQ/a} + SS_{PQ/s/a})$
	$SS_{Q/a}$	$MS_{Q/a}/MS_{Q/s/a}$	$SS_{Q/a}/(SS_{Q/a} + SS_{Q/s/a})$	$SS_{Q/a}/(SS_{Q/a} + SS_a + SS_{s/a} + SS_{P/a} + SS_{P/s/a} + SS_{Q/s/a} + SS_{PQ/a} + SS_{PQ/s/a})$
	$SS_{Q/s/a}$	—	—	—
	SS_{PQ}	$MS_{PQ}/MS_{PQ/s/a}$	$SS_{PQ}/(SS_{PQ} + SS_{PQ/s/a})$	$SS_{PQ}/(SS_{PQ} + SS_a + SS_{s/a} + SS_{P/a} + SS_{P/s/a} + SS_{Q/a} + SS_{Q/s/a} + SS_{PQ/a} + SS_{PQ/s/a})$
	$SS_{PQ/a}$	$MS_{PQ/a}/MS_{PQ/s/a}$	$SS_{PQ/a}/(SS_{PQ/a} + SS_{PQ/s/a})$	$SS_{PQ/a}/(SS_{PQ/a} + SS_a + SS_{s/a} + SS_{P/a} + SS_{P/s/a} + SS_{Q/a} + SS_{Q/s/a} + SS_{PQ/a} + SS_{PQ/s/a})$
	$SS_{PQ/s/a}$	—	—	—

Note—“A” represents a manipulated between-subjects factor, “a” represents a measured between-subjects factor, “P” and “Q” represent within-subjects factors, and “s” represents the subjects factor.

April 1, 2004). Common guidelines can be useful; for example, it is now routine to characterize zero-order correlations of .1 to .3 as small, those of .3 to .5 as medium, and those .5 and over as large, as Cohen (1988) recommended. Cohen (1988, pp. 413–414), who did not consider repeated measures designs explicitly, defined an η^2 (which is not the same as Cohen’s f^2) of .02 as small, one of .13 as medium, and one of .26 as large. It seems appropriate to apply the same guidelines to η_G^2 as well. Still, within an area of research in which designs do not vary, investigators can develop their own common guidelines for η_p^2 , and so it can still prove an effective, if local, effect size statistic.

An additional qualification should be stated. Like R^2 , η^2 overestimates the population proportion of variance explained. Still, reports often rely on the sample statistics R^2 and η^2 because their formulas are simple and straightforward and because the overestimation becomes fairly negligible as sample size increases. Nonetheless, just as adjusted R^2 corrects for the overestimation of R^2 , ω^2 corrects for the overestimation of η^2 (Hays, 1963; Keppel, 1991). Interested readers may wish to consult Olejnik and Algina (2003) regarding the computation of ω^2 , but before they conclude that ω^2 is preferable to η^2 they should read Keppel and Wickens (2004), who wrote that although “the partial ω^2 statistics are straightforward to define theoretically, they are difficult to use practically” (p. 427).

For some simple designs, there are no differences between η_G^2 , η_p^2 , and η^2 . For a one-way between-subjects ANOVA, no matter whether the factor is manipulated or

measured (A or a), values for η_G^2 are the same as those for η_p^2 and η^2 . For factorial designs with between-subjects manipulated factors (AB, ABC, etc.), η_G^2 and η_p^2 are the same—but if any between-subjects measured factors are included, η_G^2 is less than η_p^2 . Likewise, although η_G^2 and η^2 are the same for a within-subjects ANOVA with a single factor (P), in this case and for all other within-subjects or mixed designs, η_G^2 is less than η_p^2 .

In summary, investigators who report results of ANOVAs, including those with one or more repeated measures factors, are encouraged to report η_G^2 as defined by Olejnik and Algina (2003). This statistic is easily computed, as we have demonstrated here. Unlike η^2 or η_p^2 , its value is comparable across studies that incorporate the factor and outcome of interest, regardless of whether the factor is between or within subjects; thus, common guidelines can be applied to the size of the effect of the designated factor on the specific outcome. This is only the first step. More important than simply reporting effect size statistics is their discussion. As Schmidt (1996) has reminded us, for our science to accumulate we need to pay attention to effect sizes and, with time, develop a sense of the typical strength with which a pair of variables is associated in a given area of research. Too often, current articles, when summarizing the literature, merely present effects as present or absent (usually meaning statistically significant or not). Discussion of their typical size, meaning, and practical significance, which η_G^2 affords, would be a significant advance, and for that reason is a method that deserves to be more widely known and utilized.

REFERENCES

- ADAMSON, L. B., BAKEMAN, R., & DECKNER, D. F. (2004). The development of symbol-infused joint engagement. *Child Development*, *75*, 1171-1187.
- AMERICAN PSYCHOLOGICAL ASSOCIATION (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- BAKEMAN, R. (1992). *Understanding social science statistics: A spreadsheet approach*. Hillsdale, NJ: Erlbaum.
- BAKEMAN, R., & ROBINSON, B. F. (2005). *Understanding statistics in the behavioral sciences*. Mahwah, NJ: Erlbaum.
- BOND, C. F., JR., WIITALA, W. L., & RICHARD, F. D. (2003). Meta-analysis of raw mean differences. *Psychological Methods*, *8*, 406-418.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- COHEN, J., & COHEN, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- FIDLER, F., THOMASON, N., CUMMING, G., FINCH, S., & LEEMAN, J. (2004). Editors can lead researchers to confidence intervals, but can't make them think. *Psychological Science*, *15*, 119-126.
- GILLETT, R. (2003). The metric comparability of meta-analytic effect-size estimators from factorial designs. *Psychological Methods*, *8*, 419-433.
- GLASS, G. V., MCGAW, B., & SMITH, M. M. L. (1981). *Meta-analysis in social science research*. Newbury Park, CA: Sage.
- HAYS, W. L. (1963). *Statistics*. New York: Holt, Rinehart & Winston.
- KEPPEL, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- KEPPEL, G., & WICKENS, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Prentice-Hall.
- 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.
- ROSENTHAL, R., ROSNOW, R. L., & RUBIN, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge: Cambridge University Press.
- SCHMIDT, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, *1*, 115-129.
- TABACHNICK, B. G., & FIDELL, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.
- WILKINSON, L., & THE TASK FORCE ON STATISTICAL INFERENCE, AMERICAN PSYCHOLOGICAL ASSOCIATION BOARD OF SCIENTIFIC AFFAIRS (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594-604.
- WINER, B. J., BROWN, D. R., & MICHAELS, K. M. (1991). *Statistical principles in experimental design*. New York: McGraw-Hill.

(Manuscript received October 13, 2004;
revision accepted for publication March 25, 2005.)