EPSILON-SQUARED SHOULD BE PREFERRED TO ETA-SQUARED*

Hans-Magne Eikeland, University of Bergen

Introduction.

An increasing interest of recent years in variance-accounting procedures in the analysis of experimental and ex post facto data has forcibly brought forth that ANOVA (analysis of variance) in addition to being a strong inference-making device is also a powerful correlational technique, applicable to data not meeting the requirements of variables in traditional partial, semipartial, and multiple correlation.

However, for the research worker the prolific exploration of ANOVA as a variance-accounting method accompanied by ambiguously vague guidance in the use of it, has probably at present resulted in some confusion as to which variance ratio to choose for various types of data and different research problems. There is a genuinely felt need, I think, for further and deeper penetration into the nature and informative value of constructs like eta-squared, epsilon-squared, omega-squared, and ratios of variance components. Along with this should go a more systematic study of the previous literature to make it clear to what extent seemingly new constructs in this field to day are rediscoveries of constructs already conceived some 50 or 60 years ago (see, for example, Isserlis 1919, Pearson 1923, Wishart 1932).

The discussion presented in this paper is a report on a project in which the author is presently engaged with the purpose of collecting and integrating, historically and systematically, the scattered and piecemeal treatments of the different topics and

* The draft finished August 1975

issues concerning the variance-accounting aspects of ANOVA.

As I see it, information on"variation accounted for"in ANOVA designs can be extracted from three levels: the level of sum of squares, the level of mean squares, and the level of variance components which is the deep, latent structure of mean squares. There should be general agreement that the ratio eta-squared is a sums of squares ratio, that epsilon-squared and omega-squared[1)] are ratios on the mean squares level, and that intraclass types of correlation are variance components ratios.

1) No distinction will be made in the present paper between epsilon-squared and omega-squared. They are in principle identical measures of association (see Glass and Hakstian 1969), the distinguishing feature being a slight difference in the definition of total variance.The author's personal preference is for epsilon-squared.

Now, in a very general sense sums of squares ratios, mean square ratios, and components ratios are structurally alike. They all give the proportion of variation accounted for. Thus, Hays (1963),325 maintains that "the index omega-squared (ω^2) is almost identical to two other indices,...the intraclass correlation and the correlation ratio" (the last one called eta or eta-squared in this paper). Haggard(1958),6 says, "The coefficient of intraclass correlation is the measure of the relative homogeneity of the scores within the classes in relation to the total variation among all the scores in the table"..."More specifically, we may wish to know to what extent the variation of scores within classes (persons, traits, etc.) is less than the variation of scores between classes" (p7). Haggard's description of the intraclass correlation is so general and therefore so vague that it applies as well to the other ANOVA measures of association, like eta-squared and epsilon-squared.

To me it now seems important to put an emphasis to what may be said to be the distinguishing features between eta, epsilon, an alpha (an alternative name for intraclass correlation or components ratios) in order that one should be able to see what can be considered a sound and differentiated application of the various measures of association.

According to my own conception a fundamental distinction should be made between eta and epsilon on the one hand and alpha (intraclass correlation) on the other. While eta and epsilon are ratios of manifest, observed measures of variation, alpha is a ratio of inferred measures of variation, implying a theoretical structure of the measures.

In the subsequent discussion no further attention will be paid to the construction of alpha as distinct from the construction of eta and epsilon. Rather, the emphasis will be put on an argument for the convincing reason why the research worker should choose epsilon-squared before eta-squared as a general recommendation when intraclass correlation is judged out of question.

Uneasiness about the application of eta-squared.

Recent treatments of eta-squared (see, for example, Cohen 1968, Overall and Spiegel 1969, Kennedy 1970, Eikeland 1971, Cohen 1973) have been mostly concerned with describing the neat formal properties of a general eta construct. Implicitly so to speak according to these treatments, the research worker in substantive fields may feel free to an almost unrestricted use of eta-squared, since warnings for not using eta-squared are almost lacking.

In my own thinking and in the role as a consultant to research workers in different fields I have become somewhat uneasy about an unreserved application of eta-squared, and some dramatic experiences with spuriously high ratios for no good reason have forced me to seek for a more definite answer to why artificial results so easily obtain. The vague feeling that it had something to do with degrees of freedom made me more disquiet than quiet as long as an intuitive understanding of what was at work could not be provided.

Most dramatically I experienced how deceptive and untrustworthy eta-squared can be in analyzing a methods experiment one of my consultees made, using a repeated measures design with 48 subjects and 2 replications (pre- and posttest). What struck me as unreasonable was that the eta-squared for differences between subjects across replications was so unexpectedly large. I therefore decided to perform a random experiment with the 96 actual scores gained from the methods experiment. From the pool of 96 scores I randomly picked observations to put in the 48 by 2 cells in the design table. Certainly, the logical expectation of eta-squared for subjects should be zero. However, I got $\eta^2 = 0.52$.

The expectation of eta-squared in random experiments.

When a null condition exists in a data matrix, variation of scores within groups should be equal to variation of scores between groups. From elementary sampling statistics one knows that the expected standard deviation of group means based on random samples of equal size drawn from the same population is a function of the population standard deviation and sample size.

For gaining an intuitive understanding of how to find the expected eta-squared under null condition, the standard error of the mean and its basic meaning should prove an advantageous point of departure.

Recalling that for a simple ANOVA design eta-squared can be defined as the ratio of between groups sum of squares to total sum of squares, the expected eta-squared can be derived the following way:

$$E(\eta^2) = \frac{E(SS_B)}{E(SS_T)} = \frac{E\left\{(n\frac{\sigma^2}{n})(k-1)\right\}}{E(SS_T)} = \frac{E\left\{\left(n\frac{\left(\frac{E(SS_T)}{N-1}\right)}{n}\right)(k-1)\right\}}{E(SS_T)}$$

$$= \frac{E(SS_T)(k-1)}{E(SS_T)(N-1)} = \frac{k-1}{N-1} \qquad (1)$$

in which $E(\eta^2)$ is expected eta-squared, $E(SS_B)$ the expected sum of squares for groups, $E(SS_T)$ expected total sum of squares, n number of observations within groups, σ the population standard deviation, k the number of groups, and N total number of observations.

In deriving (k-1)/(N-1) as the expected eta-squared under null condition some explanation in the development of formula (1) is in order. To obtain the expected MS_B in ANOVA from the standard deviation in the population one needs to multiply the variance of the means, σ^2/n, by n because in ANOVA the MS_B is the variance of the kn scores when the respective group means have been substituted for observed scores. Thus, $n\sigma^2/n$ is the $E(MS_B)$, and multiplying by (k-1) gives $E(SS_B)$. Further, one should note that $E(SS_T/(N-1))$ can be substituted for σ^2. In manipulating the expression it reduces to $(k-1)/(N-1)$: The expected eta-squared under null condition is the ratio of degrees of freedom for between groups to degrees of freedom for the total population of the sample. Of course, the total sample as a population should be thought of as extremely large.

It is thought that the derivation performed above should have an intuitive appeal and that it is approximately correct, statistically viewed. For those who are well versed in the logic of the variance estimates in ANOVA it would even be meaningful to derive expected eta-squared more directly by just multiplying σ^2 by $(k-1)$ to obtain $E(SS_B)$ and σ^2 by $(N-1)$ to obtain $E(SS_T)$ and taking the ratio of the two. Thus,

$$E(\eta^2) = \frac{E(SS_B)}{E(SS_T)} = \frac{\sigma^2(k-1)}{\sigma^2(N-1)} = \frac{k-1}{N-1}$$

Expected eta-squared under null condition has been derived on a strictly mathematical basis by Pearson (1923) and Wishart (1932). Wishart's result is the same as obtained in the present derivation, but in Wishart's derivation there is not much intuitive logic to be discerned for the mathematically uninitiated. Kelley (1935) and Peters and Van Voorhis (1940) both mention that expected eta-squared is $(k-1)/(N-1)$ when the population eta-squared is zero. But their derivations are indirect through the derivation of epsilon-squared.

How to understand the spuriousness in eta-squared.

Even with a proof that expected eta-squared under null condition is a function of degrees of freedom, it is somewhat difficult to grasp what kind of artificial effect is at work. An insight into the seeming mysteries of why spurious ratios obtain can be provided by becoming aware of the fact that the between groups variance can be shown to be equal to the covariance between observed and predicted scores (Eikeland 1971). Now, the hazard here is that each observed score participates in its own prediction as the predicted score is the mean of the defined group's scores. There is thus an inherent contamination in the covariance between observed and predicted scores. The magnitude of the spuriousness is a question of the influence an observed score has in its own prediction. The less the number of observations within groups, the more contamination will arise. With only two observations per group as a basis for prediction the expected eta-squared under null condition will have 0.50 as a limit when the number of groups increases. In looking at ratios of sums of squares this way, my own dramatic experience of an extremely high eta-squared coefficient when a zero one was expected, can be explained by the fact that in obtaining the eta-squared for differences between 48 persons I had only 2 observed scores for estimating the predicted score for the selfsame 2 observed scores. It goes without saying that the contamination must be appreciable. In fact, I had to expect an eta-squared coefficient of magnitude $(k-1)/(N-1) = (48-1)/(96-1) = 0.49$. Recalling that my random experiment generated an eta-squared of 0.52, the result can be considered a probable event from a sampling point of view.

In research where multivariable designs are used with relatively small samples it should by now be clear that eta-squared can be a quite treacherous measure of strength of association between independent classificatory variables and a dependent quantitative dependent variable. By knowing that spurious results is dependent on the relation of number of groups to total number of observations,one can compare observed eta-squared with its expectation under null condition and take account of this in the interpretation. It is of course a much better situation than being naive and ignorant in this respect. But there is an even better solution.

Epsilon-squared and its expectation under null condition.

Fortunately, there is another choice for a measure of strength of association which will correct for the dependency of eta-squared on degrees of freedom. Kelley (1935) was aware of the bias in eta-squared and developed its unbiased companion,epsilon-squared,ε^2, where sums of squares were substituted for mean squares,

$$\varepsilon^2 = 1 - \frac{MS_W}{MS_T} = 1 - \frac{SS_W/(N-k)}{SS_T/(N-1)} = 1 - \frac{SS_W}{SS_T}\left\{\frac{N-1}{N-k}\right\} \qquad (2)$$

where k is the number of groups and N total number of observations.

Formula (2) can easily be manipulated into another form by writing SS_W/SS_T as its complementary value, i.e.$(1 - \eta^2)$,

$$\varepsilon^2 = 1 - (1 - \eta^2)\left(\frac{N-1}{N-k}\right) \qquad (3)$$

which shows that epsilon-squared has just the same form as the shrinkage formula in traditional multiple correlation. This can be found in Peters and Van Voorhis (1940), Cohen (1965) and (1968) Cureton (1966), Glass and Hakstian (1969) and McNemar (1969).

What is of considerable interest in our context is the expectation of ε^2 under null condition. By substituting $(k-1)/(N-1)$ for η^2 in (3) we get,

$$\varepsilon^2 = 1 - (1 - (\frac{k-1}{N-1}))(\frac{N-1}{N-k})$$

$$= 1 - (\frac{N-1-k+1}{N-1})(\frac{N-1}{N-k}) = 1 - (\frac{N-k}{N-1})(\frac{N-1}{N-k}) = 1 - 1 = 0 \qquad (4)$$

Thus, epsilon-squared has a chance value of zero when zero association exists in the population between the independent and the dependent variable, which shows that we are better off with epsilon-squared than with eta-squared.

In applying (3) to my own dramatic example, the spuriously high eta-squared of 0.52 will be corrected to 0.05, which under the null condition (random experiment) is quite a plausible result for a measure of association.

Partial, semipartial, and multiple eta-squared and epsilon-squared.

The demonstrations presented above have all been for the one-way ANOVA design. Most likely, in practical research work the more fruitful application of measures of strength of relationship will prove to be with multivariable designs, i.e. with more than one independent variable. Generalizing to more complex designs will not be too difficult. In the multivariable case one should be careful to recognize the options the researcher has in choosing variants of measures of association. For non-orthogonal multiways ANOVA designs, i.e. where a correlation exists between the independent variables, the relation between independent variables and the dependent variable can be explored by way of four types of association which for conceptual purposes should be distinguished as principally different.

a) The relationship between one independent variable, uninfluenced by other independent variables in the design, and the intact dependent variable.

b) The relationship between one independent variable, uninfluenced by other independent variables in the design, and a reduced dependent variable where the other independent variables have also been partialled out.

c) The relationship between a combination of orthogonalized independent variables and the intact dependent variable.

d) The relationship between a combination of orthogonalized independent variables and a reduced dependent variable where the independent variables not included in the combination are partialled out.

The categories of relationship listed above correspond to descriptive, statistical constructs well known from more traditional correlational analysis. In ANOVA designs the resulting measures of association could appropriately be named,

a) semipartial, bivariate eta-squared or epsilon-squared

b) partial, bivariate eta-squared or epsilon-squared

c) semipartial, multiple eta-squared or epsilon-squared

d) partial, multiple eta-squared or epsilon-squared

Semipartial correlation (see, for example, Nunnally 1967) is here used as a synonym for the more commonly used part correlation (see, for example, McNemar 1969).

Which type of measure of association to choose is for the research worker to decide depending only on the research problem he seeks an answer to. Thus, there can be no general recommendation that either a partial or a semipartial approach should be a best procedure (cfr. Kennedy 1970 and Cohen 1973).

It proves almost prohibitive to work out a set of formulas to be applicable to any kind of complex ANOVA designs when these measures of association are sought for. Mechanical rules will not do. Insightful thinking is necessary to be able to construct what Cohen (1973),111 calls a "custom-tailored partial η^2". In the system of categories presented above one would like to say "custom-tailored partial and semipartial η^2 or ε^2".

In order to be more concrete and specific as to what the different measures of association (a - d above) mean and how they can be worked out and interpreted, a set of hypothetical data is presented in the matrix of Table 1. An experiment is performed to assess the effect of IQ group membership and socio-economic group membership, SE, separately for each variable, included the interaction, and also in combination, on school achievement. IQ group membership is obtained by having teachers rate pupils as above or below median intelligence, and two socio-economic sub-populations are deliberately chosen so as to possibly maximize hypothesized effects. Let us assume proportionate, stratified random sampling, and a total sample of N = 20.

Table 1. Hypothetical data matrix.

	SE_{low} (B_1)	SE_{high} (B_2)	
IQ_{high} (A_1)	24 22 20	28 26 24 22 20 18 16	22,0
IQ_{low} (A_2)	22 20 18 16 14 12 10	22 20 18	17,2
	17,8	21,4	19,6

As can be seen from the data matrix the design is made non-orthogonal. There is a correlation between IQ and SE, or between A and B, but the interaction AB is uncorrelated with both A and B. (This is a deliberate simplification in order not to complicate matters too much in the analytic procedure.)

In reading the ANOVA table, Table 2, one should note that the sums of squares in column SS_1 are not additive, i.e. they do not sum to SS_T, the total sum of squares. While the observed SSs for A, B, and AB add to 196,8, the correct SS for combined groups is 148,8. This discrepancy is a consequence of the correlation between A and B. In column SS_2 the influence of intelligence on socioeconomic groups has been partialled out, and

Table 2. ANOVA table for hypothetical data.

Source	df	SS_1	SS_2	SS_3
A (IQ)	1	115,2	115,2	67,2
B (SE)	1	64,8	16,8	64,8
AB	1	16,8	16,8	16,8
Within cells	16	240,0	240,0	240,0
Total	19	388,8	388,8	388,8

in column SS_3 the influence of SE on IQ has been partialled out. The partialized SSs will be symbolized as $SS_{B.A}$ for column SS_2 and $SS_{A.B}$ for column SS_3. By the partializing procedure columns SS_2 and SS_3 have been made additive. This is accomplished for column SS_2, for example, by subtracting A's and AB's contributions from SS for combined groups which is 148,8. Thus, $SS_{B.A} = SS_G - SS_A - SS_{AB} = 148{,}8 - 115{,}2 - 16{,}8 = 16{,}8$ which is B's contribution independent of both A's and AB's contributions to the group variation.

Now, let us see how the four categories of measures of association related to complex ANOVA designs, the a-d categories pp.10-11 above, can be applied in our example. Say that the research problem concerns the effect of SE on school achievement with the influence of IQ controlled.

a) If the intention is to asses the strength of association between SE with IQ partialled out and the intact scores on school achievement, i.e. when all other systematic variables are partialled out of the independent variable of concern but not out of the dependent variable, then a semipartial eta-squared or epsilon-squared is called for.

First, the semipartial, bivariate eta-squared will be symbolized, defined, and computed. The research problem as posed here will be concerned with column SS_2 in Table 2 because it is a question of controlling IQ. Thus,

$$\eta^2_{Y(B.A,AB)} = \frac{SS_{B.A,AB}}{SS_T} = \frac{16,8}{388,8} = 0.0432 \qquad (5)$$

In (5) the subscript to eta-squared should be noted. It signifies that eta-squared is between intact Y (dependent variable) and B controlled for A and AB. In our example it is unnecessary to control B for AB since B and AB are uncorrelated already by design. However, for the purpose of covering the more general case of nonorthogonal design, AB is included as if controlled for statistically. The result in (5) is commonly described as the correlation between Y and B.A,AB which is the square of 0.0432, i.e. 0,208. There is not much gained by sticking to this convenvention since the squared coefficient lends itself so much more easily to a meaningful interpretation.

Of even more interest in our context is to develop the degrees-of-freedom-corrected eta-squared of (5). Formula (2) with a slight modification will be used. In stead of SS_W we now had better change to SS_R, meaning sum of squares for residual. By SS_R we shall mean the left-over SS when SS for the systematic source of interest is subtracted from the defined total SS, which can be either the unreduced total SS, SS_T, or a reduced SS total, $SS_{T'}$. In effect, SS_R will be a new "error" term including a genuine error term SS_W +systematic sources. More in keeping with the logic of semipartial,bivariate eta-squared or epsilon-squared would be, I think, to regard the systematic variation not included in the systematic variation of interest as ignored variation. By taking one sytematic source of variation at a time, ignoring the other systematic sources, one behaves as if no more information were at hand than that contained in the source of particular interest right now. Therefore, the ignored systematic variation will temporarily go to the noise category of variation and in a way reduce the signal by signal + noise ratio.(Notice, this will not happen when partial eta-squared or epsilon-squared are used.) After this, a more general definition of epsilon-squared for the semipartial, bivariate category can be written,

$$\varepsilon^2_{Y(B.A,AB)} = 1 - \frac{SS_R}{SS_T}\left(\frac{df_T}{df_R}\right) \tag{6}$$

$$= 1 - \frac{372,0}{388,8}\left(\frac{19}{18}\right) = 1 - 1,009945 = -0,009945$$

Epsilon-squared of (6) applied to our hypothetical data has a value of zero.This means that there is no association between school achievement and SE when the SE effect is taken as an average across the two IQ groups, and when the general IQ effect on school achievement across the two SE groups has been controlled for. (A possible differential SE effect for the two IQ groups on school achievement,with general IQ effect controlled for, has to do with the strength of association between the AB interaction and Y, school achievement.)

Formula (6) can be manipulated into another form, yielding

$$\varepsilon^2_{Y(B.A,AB)} = \frac{SS_{B.A,AB} - (B-1)MS_R}{SS_T} \quad (7)$$

where B-1 is degrees of freedom for B.

b) If one is interested in the relationship between SE,uninfluenced by IQ and the SE/IQ interaction, and school achievement, also uninfluenced by IQ and the SE/IQ interaction, then a partial, bivariate eta-squared or epsilon-squared is called for.

$$\eta^2_{YB.A,AB} = \frac{SS_{B.A,AB}}{SS_{B.A,AB} + SS_R} = \frac{SS_{B.A,AB}}{SS_{T'}} \quad (8)$$

$$= \frac{16,8}{16,8 + 240,0} = 0,0654$$

Partial epsilon-squared, by adapting formula (6) to the present condition, will become,

$$\varepsilon^2_{YB.A,AB} = 1 - \frac{SS_R}{SS_{T'}}\left(\frac{df_{T'}}{df_R}\right) \quad (9)$$

$$= 1 - \frac{240,0}{256,8}\left(\frac{17}{16}\right) = 0,0070$$

A particular attention should be paid to the number of degrees of freedom going with partial epsilon-squared. In the present case, two sources have been partialled out, each with df = 1. Thus, $df_{T'}$ will be 2 less than df_T, and df_R will be 1 less than $df_{T'}$ since the systematic source of interest in the measure of association has df=1. The alternative form to (9) will be,

$$\varepsilon^2_{YB.A,AB} = \frac{SS_{B.A,AB} - (B-1)MS_R}{SS_{T'}} \quad (10)$$

It should be noted that MS_R in (7) and MS_R in (10) are not the same definitions of the residual variation (see p 14, bottom).

Formula (10) can be shown to be identical to formula (186) in Peters and Van Voorhis (1940), p.354, with a slight modification made for the case of intercorrelated independent variables in our formula (10).

c) Multiple correlation is the correlation between a combination of orthogonalized independent variables with a criterion. It is based on a semipartial correlation procedure in that an independent variable is partialled out of another independent variable but not out of the dependent variable. In our case the multiple etq-squared is given by taking the ratio of the between groups sum of squares, which is 148,8, and total sum of squares, which is 388,8. By this procedure maximum variation accounted for is taken out. No redundancy will occur even if the independent variables are correlated. Thus,

$$\eta^2_{Y(A+B.A+AB.A,B)} = 1 - \frac{SS_W}{SS_T} = \frac{SS_B}{SS_T} = \frac{148,8}{388,8} = 0.3827 \qquad (11)$$

The parallel to the shrunken multiple correlation squared in traditional multiple correlation procedures is multiple epsilon-squared. By using formula (2),

$$\varepsilon^2_{Y(A+B.A+AB.A,B)} = 1 - \frac{SS_W}{SS_T}\left(\frac{N-1}{N-k}\right) = 1 - \frac{240,0}{388,8}\left(\frac{19}{16}\right) = 0,2670 \qquad (12)$$

The estimate of the bias in multiple eta-squared is obtained by taking the $(k-1)/(N-1)$ ratio, i.e. $3/19 = 0.1579$, which is the expectation of (11) when no substantive association exists in data

The alternative formula, adapted for the case of multiple epsilon-squared, will have the following form,

$$\varepsilon^2_{Y(A+B.A+AB.A,B)} = \frac{SS_G - (G-1)MS_W}{SS_T} \tag{13}$$

$$= \frac{148,8 - (4-1)15.0}{368,8} = 0,2670$$

where SS_G is the sum of squares for the four groups in the data matrix of Table 1 and G total number of groups in the design.

d) In c) above where multiple eta-squared and epsilon-squared were decribed, all three independent variables, A, B, and AB, were used as predictors. By so doing, the influence of IQ together with SE and the SE/IQ interaction was observed. Now, returning to the research problem as sketched previously (see p. 13), we might be interested in seeing to what extent a combined general and differential SE effect influences achievement scores when intelligence is controlled for both in the independent variables and the dependent variable.The general SE effect is shown if there is a difference in average achievment score for the two SE groups across the two IQ groups. A differential SE effect is present if the difference in achievement between the SE groups is different for the two intelligence groups. The problem set forth here asks for a partial, multiple eta-squared or epsilon-squared which can be obtained the following way, first eta-squared,

$$\eta^2_{Y(B+AB.B).A} = \frac{SS_{B.A} + SS_{AB.A,B}}{SS_{B.A} + SS_{AB.A,B} + SS_W} \tag{14}$$

$$= \frac{16,8+16,8}{16,8+16,8+240,0} = \frac{33,6}{273,6} = 0,1228$$

Next, epsilon-squared for the same problem,

$$\varepsilon^2_{Y(B+AB.B).A} = 1 - \frac{SS_R}{SS_{T'}}\left(\frac{df_{T'}}{df_R}\right) = 1 - \frac{240,0}{273,6}\left(\frac{18}{16}\right) = 0,0132 \quad (9)$$

The form of the partial, multiple epsilon-squared as given above is equal to formula (9) but the content is somewhat different, as can be seen by comparing (8) and (14). In using (9), it should be clear that the problem posed defines $SS_{T'}$ and SS_R in (9).

The alternative form to the partial, multiple epsilon-squared will be,

$$\varepsilon^2_{Y(B+AB.B).A} = \frac{SS_{B.A} + SS_{AB.A,B} - (df_B + df_{AB})MS_R}{SS_{T'}} \quad (15)$$

The custom-tailored forms given to eta-squared and epsilon-squared above for specific questions put to data should be a reminder to the research worker that it is difficult to give quite general formulas for complex designs because so many possibilities exist for specific problems to seek an answer to. The presentation above is thought to be of considerable help in showing that a conceptualization of the problem is necessary in order for the research worker to be able to find a solution to how to generate the correct measures of association.

The presentation also has shown that for every eta-squared there has been a companion epsilon-squared at hand.

<u>The reason why epsilon-squared should be preferred.</u>

Kelley (1935) in describing the properties of epsilon-squared did not make a very strong case for generally preferring epsilon-squared to eta-squared. Peters and Van Voorhis (1940) seem more intent on the use of epsilon-squared than any other some 35 to 40 years ago, but later on the interest in both measures dwindled, and they were almost forgotten. Bolles and Messick (1958), Gaito(1958), and Diamond (1959) did not succeed in raising a new interest.

Hays's (1963) introduction of omega-squared (in fact, a re-introduction of epsilon-squared) has caught much attention and led to extended use of measures of strenght of association in ANOVA contexts. But Hays did not compare eta-squared and omega-squared. Cureton (1966) presented a very interesting and illuminating categorizing of correlation coefficients where the distinguishing feature may be said to be whether the coefficients were corrected or not corrected for bias because of degrees of freedom. But he does not take a stand as to application for one category in preference to another. To him the choice is a matter of personal preference. The present author (Eikeland 1971) in his description of how general the eta concept was, paid no attention to epsilon-squared. The same is the case with Kennedy (1970), and Cohen's (1973) reply to Kennedy does not point to epsilon-squared as a more preferable choice than eta-squared.

Thus, research workers in the substantive fields do not seem to have been well guided by methodological papers in the journals to make what to me now looks as the most reasonable choice,i.e. applying epsilon-squared is generally speaking the safest choice.

My own experience and thinking leave no doubt about that any longer. Although in principle eta-squared and epsilon-squared convey the same information from data, eta-squared has the built-in bias that happens to lead to quite deceptive and misleading results. In using epsilon-squared one need not be too wary about small samples in research work as no bias is introduced for that reason. Certainly, small samples should make us cautious in deciding what should be regarded as signal and what as noise because sampling fluctuations will be more predominant in the statistics, but this is not bias.

The research worker, being ignorant of the spuriousness in eta-squared, is likely to be deceived , for example, in exploratory investigations with a fixed data set and possibilities of splitting it up in more and more categories. Each new category will almost certainly seem to account for variation since eta-squared most probably will go up by sheer artificial reasons. (For an example of this kind of application, see Solstad 1973.) In such cases the informed researcher knows that an automatic increase in eta-squared is likely to happen because the numerator in expected eta-squared (k-1) goes up while the denominator (N-1) remains constant. In this regard epsilon-squared is safe, and it is a convincing reason for recommending it to be used

Eta-squared and epsilon-squared - how meaningful are they?

Some years ago Glass and Hakstian (1969) brought forth what might seem a devastating argument against the use of measures of strength of association for fixed ANOVA designs. According to them one should rather not use, for example, epsilon-squared.

The heart of the matter, as they see it, is that it is not meaningful to describe results in terms of a (squared) correlation coefficient when the levels of one of the variables (the independent one) are arbitrarily or purposely chosen by the investigator, and more often than not ill defined. Certainly, there are precautions to observe in this respect, but Glass and Hakstian's argument applies as well to the use of a fixed effects ANOVA model at all as to the use of measures of association.

There can be good reasons for not using the term correlation in this context. Hays (1963) distinguishes between regression problems and correlation problems (approximately equivalent to the distinction between fixed effects and random effects models in ANOVA). Glass and Hakstian argue forcefully for an approach to problems that fits the random model, i.e. one should be more concerned with drawing levels randomly to achieve representative designs. Nobody will disagree, but there can be no doubt that fixed effects models are needed in seeking answers to research questions. In Hay's terms, regression problems are relevant.

In my view the measures of strength of association related to fixed effects ANOVA designs might well be named differentiation ratios (see Diamond 1959) to avoid the mixing up with correlation coefficients in a more narrow sense.

Glass and Hakstian's discussion is a reminder not to interpret such differentiation ratios in any absolute sense. But that is even the case with coefficients in more traditional correlation problems. The interpretation of such ratios will always have to be made in a comparative and relative context, depending on the chosen levels of the fixed independent variables, the sample provided, previous result with the same kind of problem, and so on.

With such precautions in mind I can see no reason not to make more extensive use of epsilon-squared.

References.

Bolles, R. and Messick, S. 1958. Statistical utility in experimental inference. Psychological Reports, 4, 223-227.

Cohen, J. 1965. Some statistical issues in psychological research. In Wolman, B. B. (Ed). Handbook of Clinical Psychology. New York: McGraw-Hill.

Cohen, J. 1968. Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cohen, J. 1973. Eta-squared and partial eta-squared in fixed ANOVA designs. Educational and Psychological Measurement, 33, 107-112.

Cureton, E. E. 1966. On correlation coefficients. Psychometrika, 31, 605-607.

Diamond, S. 1959. Information and Error. New York: Basic Books.

Eikeland, H. M. 1971. On the generality of univariate eta. Scandinavian Journal of Educational Research, 15, 149-167.

Gaito, J. 1958. The Bolles-Messick coefficient of utility. Psychological Reports, 4, 595-598.

Glass, G. V. and Hakstian, A. R. 1969. Measures of association in comparative experiments: Their development and interpretation. American Educational Research Journal, 6, 403-414.

Haggard, E. A. 1958. Intraclass Correlation and the Analysis of Variance. New York: Dryden Press.

Hays, W. L. 1963. Statistics. New York: Holt, Rinehart and Winston.

Isserlis, L. 1919. On the partial correlation ratio. Biometrika, 10, 391-411.

Kelley, T. L. 1935. An unbiased correlation ratio measure. Proceedings of the National Academy of Sciences, 21, 554-559.

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

McNemar, Q. 1969. Psychological Statistics. Fourth Edition. New York: John Wiley.

Nunnally, J. C. 1967. Psychometric Theory. New York: McGraw-Hill.

Overall, J. E. and Spiegel, D. K. 1969. Concerning least squares analysis of experimental data. Psychological Bulletin, 72,

Pearson, K. 1923. On the correction necessary for the correlation ratio. Biometrika, 14, 412-417.

Peters, C. C. and Van Voorhis, W. R. 1940. Statistical Procedures and their Mathematical Bases. New York: McGraw-Hill.

Solstad, K. J. 1973. School transportation and physical development. Scandinavian Journal of Educational Research, 17, 117-126.

Wishart, J. 1932. A note on the distribution of the correlation ratio. Biometrika, 24, 441-456.