CONCERNING NEGATIVE VARIANCE

COMPONENTS

IN

REPEATED MEASURES DESIGNS

Hans-Magne Eikeland

UNIVERSITY OF OSLO

1973

CONCERNING NEGATIVE VARIANCE COMPONENTS IN REPEATED MEASURES DESIGNS

Hans-Magne Eikeland, University of Oslo

Among measures of association, variance components analysis has attracted increasingly more attention in social science literature as a means of assessing latent variance structures in data. (For recent references, see, for example, Hays 1963, Medley and Mitzel 1963, Gleser, Cronbach, and Rajaratnam 1965, Endler 1966, Vaughan and Corballis 1969.)

This report is concerned with a particular problem that may arise when variance components analysis is employed: The occurrence of negative components.

Intuitively, negative variance components seem to make no sense, either syntactically, or semantically. It is obviously clear that variances can not take negative values, as they are functions of the square of deviation scores. By analogy, it seems reasonable to think that components can not do so either. Further, no meaningful interpretation of negative components seems possible.

The issue has not been much discussed in the literature. Most of what is written about negative variance components can be found in technical literature. (See, for example, Nelder 1954, Scheffe 1959, Thomas and Moore 1963, Hill 1965.) In social science literature, the discussion is mostly restricted to making recommendations of what to do when negative components occur. As a recent example, say Vaughan and Corballis:

"It is possible in practice for a computed variance component to assume a <u>negative quantity</u>. This seemingly paradoxical result would occur in a one-way design, for example, when the value of MS(w) exceeds that of MS(a). In such cases, the most plausible estimate in <u>that instance</u> would be zero (Cf. Hays 1963, p. 383). However, replacing a negative estimate by zero introduces a positive bias, and the experimenter is best advised to report the negative value (Scheffe 1959, p. 229), particularly if the estimate is to be considered in conjunction with estimates from other experiments." (Vaughan and Corballis 1969, 212)

Nevertheless, there are cases where negative components occur that one reluctantly and with some hesitation yields to the recommendations commonly given that negative components should be set to zero. Reflecting on the fact that variances frequently can be looked upon as sums of variances and covariances, one wonders whether it should not be possible for variance components, at least when considered covariance components, to assume negative values, both as far as syntax and semantics are concerned.

The author has recently discussed this particular problem in connection with the Hoyt analysis of a homogeneous test and coefficient alpha. (Hoyt 1941) The Hoyt design is an N persons by k tests design, i.e. a repeated measures design, where it is undoubtedly true that the variance component of interest is a covariance component. It is an average covariance. (Eikeland 1970) Certainly, as a covariance, a component may well take a negative value.

The following discussion will be limited to repeated measures designs. At least for this type of design, one can be fairly confident in the validity of converging covariances on the variance component construct.

A simple case of negative variance components.

Consider a simple repeated measures design, say, a 5 persons by 2 tests design. The results in TABLE 1 are deliberately made extreme, in order to emphasize what is here at issue.

TABLE 1

Hypothetical data

	T_1	T_2	Sum	Dif
a	5	1	6	4
b	4	2	6	2
c	3	3	6	0
e	2	4	6	-2
d	1	5	6	-4
$\overline{X}$	3	3	6	0

Source	SS	df	MS
Rows	0	4	0
Columns	0	1	0
R x C	20	4	5
Total	20	9	

	Model	MS	VC
MS(row)	$\sigma^2_{rc} + 2\sigma^2_r$	$0 = 5 + 2(-2,5)$	-2,5
MS(col)	$\sigma^2_{rc} + 5\sigma^2_c$	$0 = 5 + 5(-1,0)$	-1,0
MS(rc)	σ^2_{rc}	$5 = 5$	5,0

The observed results of zero variances for rows and columns in TABLE 1 should not be left by just reporting this seemingly uninteresting outcome. In fact, they are challenging results when the underlying structure of these zero variances is looked for. By writing the ANOVA model for rows and columns, one learns that the variances are composed of two components each, one positive, the other negative, such that the positive and the negative values balance.

The variance of rows is the sum score variance across the two tests. One can easily see that the undifferentiated sum scores are obtained by summing two perfectly negatively correlated sets of scores. This is reflected in the negative component for rows.

The variance of columns is the sum score variance across the five persons. It is more difficult to see that the undifferentiated test scores are a function of negatively correlated persons. On the average, persons are negatively correlated, although some persons correlate positively. The negative component for columns reflects this fact.

It is extremely important to realize that the models for rows and columns are functions of two different variance-covariance matrices. The model for rows is a function of the 2 x 2 variance-covariance matrix for tests, while the model for columns is a function of the 5 x 5 variance-covariance matrix for persons.

TABLE 2
Variance-covariance matrices
for tests and persons

	T_1	T_2
T_1	2,5	-2,5
T_2	-2,5	2,5

V = 2,5   Cov = -2,5

	a	b	c	d	e
a	+8	+4	0	-4	-8
b	+4	+2	0	-2	-4
c	0	0	0	0	0
d	-4	-2	0	+2	+4
e	-8	-4	0	+4	+8

$\overline{V} = 4$
$\overline{Cov} = -1$

The variance component for rows is the covariance between tests, -2,5. The variance component for columns is the average covariance among persons, -1,0. The residual, or the row by column interaction, in TABLE 1 is obtained by simple subtraction, $\overline{v} - \overline{cov}$. The same result is obtained either by using the variance-covariance matrix of tests, or the variance-covariance matrix of persons: 2,5-(-2,5)= 5,0, or 4,0 - (-1) = 5.

One should also realize that the difference score variance of tests in TABLE 1 is equal to the row by column interaction. The interaction variance can also be written as a variance-covariance matrix. In the matrix for the variance of difference scores the sign for test T_2 will be reflected, and the sign for the covariance between tests in TABLE 2 becomes positive. Thus the sum of the variance-covariance matrix for the difference scores will be 10 which is k times greater than the variance obtained in ANOVA. The general relationship between the sum of the variance-covariance matrix, either the sum score variance or the difference score variance, and the variance obtained in analysis of variance is, V(sum) = kMS(p), and V(dif) = kMS(res).

The relation between columns and among rows can be expressed as intraclass correlations by way of the model in TABLE 1. The intraclass correlation is here called alpha.

$$\text{alpha}_{\text{row}} = \frac{\sigma_r^2}{\sigma_{rc}^2 + \sigma_r^2} = \frac{-2,5}{5 - 2,5} = -1,0$$

$$\text{alpha}_{\text{col}} = \frac{\sigma_c^2}{\sigma_{rc}^2 + \sigma_c^2} = \frac{-1,0}{5 - 1,0} = -0,25$$

It is indeed uncommon to have negative intraclass correlations. One may reasonably ask whether they are correct and what they mean. There is no doubt that the values of the components are arithmetically correct. And so are the intraclass correlations. It should also be clear what they mean. The intraclass correlation of -1,0 for rows is just the perfect negative correlation between tests. That this is correct, can be verified by inspecting TABLE 1. The intraclass correlation for columns is the average correlation among persons. This is not convincingly clear by just inspecting TABLE 1 to see how persons go together. However, the variance-covariance matrix for persons in TABLE 2 shows that there are more negative covariances than positive ones among persons.

One thing should be clear about the intraclass correlations above: $\text{Alpha}_{\text{row}}$ is the correlation between columns, and $\text{alpha}_{\text{col}}$ is the correlation among rows. The variance of rows is concerned with the differentation of sum scores across columns. This sum score variance is to a very great extent dependent on the correlation between the tests. The more positively correlated columns, the more differentiation among rows in terms of sum score variance. The opposite is true for the difference score variance:The more negatively correlated columns,the larger the difference score variance for rows.The same reasoning holds for the sum score variance for tests,and the difference score variance for tests. These variances are dependent on the correlation among persons.

The hypothetical data above has been so chosen as to be as simple and as extreme as possible in order to make it easily seen that negative variance components may be tenable,both syntactically and semantically.

It is much more difficult to conceive of negative variance components as meaningful in more complex designs where one gets involved in variance components for interactions,say. The following discussion is concerned with reanalyzing and reinterpreting more complex data where negative components have been found,in order to explore the possibility of making meaningful interpretations of such components.

A reanalysis of Bock's (1960) data.

In a paper concerned with an ANOVA method for a structural and a discriminal analysis of psychological tests, Bock (1960) presented data where a negative component occurs for one of the interactions. As is commonly done in such cases, Bock ignores the negative value. A consequence of ignoring the negative variance component is that one of Bock's intraclass correlations is not arithmetically correct. Another possible consequence may be that interesting substantive results are lost by setting the negative component to zero.

Bock analyzes four tests that are classifiable a priori in a factorial design. They are identified according to two modes of classification, stimulus and response. There are two stimuli, picture and word; two responses, oral and graphic. The tests in the battery constitute an experimental form of the Language Modalities Survey to be used with aphasic subjects.

The test design is a 2 x 2 factorial design as shown in FIGURE 1. It is a repeated measures design in that there are more than one observation for each person.

Stimulus:	Picture(b_1)		Word(b_2)	
Response:	Oral(c_1)	Graphic(c_2)	Oral(c_1)	Graphic(c_2)
Persons : 1 2 . . . 49 50				

FIGURE 1. Bock's (1960) test design.

For Bock's purpose the only interest is in individual differences in common and more specific test effects. Thus, only sources of variance including persons are involved in the analysis. This means that the data matrix in FIGURE 1 will be column-centered. By so doing, all variances connected with the tests are partialled out. These sources are stimulus, response, and stimulus by response interaction.

The ANOVA results are presented in TABLE 3. One simplification has been made in the result table: The replication variance, or error, has been ignored, leaving the design as if it were an un-

TABLE 3

Bock's (1960) ANOVA results

Source	df	MS	VC
A Persons	49	26,313	4,552
B Person x stimulus	49	3,120	-1,391
C Person x response	49	10,887	2,492
D P x S x R	49	5,903	5,903
Total'	196	46,223	11,556

replicated one. In effect, this means that the average of two observations is used in each of the 50 x 4 cells in the design. For the purpose of the present discussion, this simplification does not in any way influence the points subsequently to be made.

The variance components in TABLE 3 are obtained by solving for the unknowns in the models for three of the MS's,

$$MS_{p} = \sigma^2_{psr} + 2\sigma^2_{pr} + 2\sigma^2_{ps} + 4\sigma^2_{p}$$
$$MS_{ps} = \sigma^2_{psr} + 2\sigma^2_{ps}$$
$$MS_{pr} = \sigma^2_{psr} + 2\sigma^2_{pr}$$
$$MS_{psr} = \sigma^2_{psr}$$

It is important to realize that in TABLE 3 the MS's can be added to yield the sum of the variances of the four tests, which is 46,223. This sum implies the concept of trace from multivariate statistics.Trace is the sum of the variances in the principal diagonal of the 4 x 4 variance-covariance matrix of the four tests.By the ANOVA technique the trace has been split into variances of four orthogonal linear combinations of the four tests.Source A in TABLE 3 is the sum score variance,the sum being made across all four tests.Source B is the variance of the difference scores of the two stimuli,keeping response constant. Source C is the variance of the difference scores for the two responses,keeping stimuli constant.Lastly,the D source is the triple interaction PSR,or the variance of the difference of difference scores,like the difference between the differences of response scores within each of the two stimuli,or the difference between the differences of the stimulus scores within each of the two responses. The four orthogonal linear combinations exhaust the variance in the four original tests.

Also,it should be realized that the components add to the average trace,or the average test variance,which is 46,223/4 = 11,556.

By setting the PS component to zero,as did Bock,the sum of components will be too large;12,948 in stead of 11,556.One consequence of this is that Bock's intraclass correlation for A,for persons,will be too low,since the component for A has been divided by a too large average test variance.Bock's intraclass correlation is obtained by 4,552/12,948 = 0,351; while the arithmetically correct coefficient should be 4,552/11,556 = 0,394.

In order to explore into how negative variance components can be viewed in the present design,it may prove highly useful to look at the MS's in TABLE 3 as functions of the variance-covariance matrix of the four tests involved in the Bock study. It will then be clear that the variance components are complex functions [redacted] of the same variance-covariance matrix.

First, the variances of the four linear combinations of the four tests, mentioned above, are developed. The sum score variance is obtained by adding and squaring the deviation scores of the four tests:

$$V_a = \frac{N}{N-1} \sum (x_{11} + x_{12} + x_{21} + x_{22})^2 \qquad (1)$$

By expanding (1), 16 variances and covariances are obtained. There are 4 variances, one for each of the tests, and 12 covariances among tests. The covariances can be grouped into 3 different classes of covariances: (a) Heterostimulus-heteroresponse covariances. These are the covariances among tests that differ both in stimulus and response. (b) Heterostimulus-monoresponse covariances. These are the covariances among tests that differ in stimulus, but not in response. (c) Monostimulus-heteroresponse covariances. These are the covariances among tests that differ in

TABLE 4

Categorized variance-covariance matrix

	T_{11}	T_{12}	T_{21}	T_{22}
T_{11}	$\overline{v}$	$\overline{cov}_c$	$\overline{cov}_b$	$\overline{cov}_a$
T_{12}	$\overline{cov}_c$	$\overline{v}$	$\overline{cov}_a$	$\overline{cov}_b$
T_{21}	$\overline{cov}_b$	$\overline{cov}_a$	$\overline{v}$	$\overline{cov}_c$
T_{22}	$\overline{cov}_a$	$\overline{cov}_b$	$\overline{cov}_c$	$\overline{v}$

response, but not in stimulus. The different categories of covariances (variance included) are presented in TABLE 4. The covariances are averages of respective category of covariance, while the variance is the average test variance.

It should be clear that the sum of the variance-covariance matrix in TABLE 4 equals the sum score variance, i.e. the sum score of persons across the four tests. Also, it should be understood that MS_a in TABLE 3 is 1/k of V_a, or 1/k of the variance-covariance matrix. While V_a is the sum score variance, MS_a is the variance of the average sum score across tests when these averages are substituted for the observed scores. Still another variance would be $V_{\bar{a}}$, the average sum score variance. Thus the following relationship obtains for the three variances: $V_a = kMS_a = k^2 V_{\bar{a}}$. (2)

The variance of the linear combination in (1), which is the sum of the tests, after expanding and ordering, can be witten in terms of average variance and average covariances as,

$$V_a = 4\bar{v} + 4\overline{cov}_c + 4\overline{cov}_b + 4\overline{cov}_a \tag{3}$$

According to (2), the variance for persons in the ANOVA table, TABLE 3, can be written,

$$MS_a = \bar{v} + \overline{cov}_c + \overline{cov}_b + \overline{cov}_a \tag{4}$$

As mentioned above, the decomposition of total test variance brought about by the ANOVA technique, has generated four orthogonal linear combinations of the tests, of which the sum score is one of them. The other three linear combinations are difference scores based on the a priori design on the tests.

The variance of these three linear combinations are obtained the following way:

$$V_{ps} = V_b = \frac{N}{N-1}\Sigma(x_{11} + x_{12} - x_{21} - x_{22})^2 \tag{5}$$

$$V_{pr} = V_c = \frac{N}{N-1}\Sigma(x_{11} - x_{12} + x_{21} - x_{22})^2 \tag{6}$$

$$V_{psr} = V_d = \frac{N}{N-1}\Sigma(x_{11} - x_{12} - x_{21} + x_{22})^2 \tag{7}$$

By expanding formulas (5), (6), and (7), and by grouping and averaging covariances as was done with the sum score variance, the three variances can be given as functions of the average test variance and averages of different categories of covariances. In going from the variances of the linear combinations in (5),(6), and (7),one should recall that those variances have to be divided by k,or for this particular case,by 4,in order to obtain the correct MS's.

$$MS_{ps} = MS_b = \bar{v} + \overline{cov}_c - \overline{cov}_b - \overline{cov}_a \quad (8)$$

$$MS_{pr} = MS_c = \bar{v} - \overline{cov}_c + \overline{cov}_b - \overline{cov}_a \quad (9)$$

$$MS_{psr} = MS_d = \bar{v} - \overline{cov}_c - \overline{cov}_b + \overline{cov}_a \quad (10)$$

From (4),(8),(9),and (10) one can see that the MS's in TABLE 3 are all functions of the same terms,but the configuration of signs differ from source to source.

The development of the mean squares as functions of variances and covariances has been necessary in order to take a next step that is crucial for the present discussion.This further step is the development of the model in terms of variances and covariances. What is aimed at is to express the components as functions of the variance-covariance matrix.

TABLE 5

Variance components
in terms of variances and covariances

	Observed	Inferred	Variance components
MS_a	$= \bar{v}+\overline{cov}_c+\overline{cov}_b+\overline{cov}_a$	$=\sigma_d^2+2\sigma_c^2+2\sigma_b^2+4\sigma_a^2$	$\sigma_a^2=\overline{cov}_a$
MS_b	$= \bar{v}+\overline{cov}_c-\overline{cov}_b-\overline{cov}_a$	$=\sigma_d^2+2\sigma_b^2$	$\sigma_b^2=\overline{cov}_c-\overline{cov}_a$
MS_c	$= \bar{v}-\overline{cov}_c+\overline{cov}_b-\overline{cov}_a$	$=\sigma_d^2+2\sigma_c^2$	$\sigma_c^2=\overline{cov}_b-\overline{cov}_a$
MS_d	$= \bar{v}-\overline{cov}_c-\overline{cov}_b+\overline{cov}_a$	$=\sigma_d^2$	$\sigma_d^2=\bar{v}-\overline{cov}_c-\overline{cov}_b+\overline{cov}_a$
Sum	$4\bar{v}$	$4\sigma_d^2+4\sigma_c^2+4\sigma_b^2+4\sigma_a^2$	$\sigma_d^2+\sigma_c^2+\sigma_b^2+\sigma_a^2 = \bar{v}$

From TABLE 5 it is apparent that the MS's of the four linear combinations, given as functions of the average variance and the average of different categories of covariances, add to $4\bar{v} = \Sigma v$, which is the trace. The sum of the unweighted components is $\bar{v}$. What is crucial to notice in the present context, is the way the components can be written in terms of average covariances.

That the triple interaction component also can be conceived of as a residual variance, besides being the variance of a particular linear combination of the four tests, can be observed by subtracting the components for A, B, and C (P, PS, PR):

$$\sigma^2_{psr} = \sigma^2_{res} = \bar{v} - \overline{cov}_a - (\overline{cov}_c - \overline{cov}_a) - (\overline{cov}_b - \overline{cov}_a)$$
$$= \bar{v} - \overline{cov}_c - \overline{cov}_b + \overline{cov}_a$$

The fact that variance components can be expressed as functions of covariances, should make it convincingly clear that negative values obtain quite naturally. This is certainly obvious for the component of A, or the P component.

In the case of the components for PS and PR it is not so easily seen how a negative component should be judged, although it is clear how they are generated. In order to obtain negative components for PS and PR, the $\overline{cov}_c$ and the $\overline{cov}_b$ have to be smaller than $\overline{cov}_a$. This can certainly easily happen.

The ANOVA model commonly used, implies a hierarchical structure in that the general component A is assumed to be partialled out of the sources of variance on less general levels, like the interactions. (For a discussion of this, see Overall and Spiegel 1969.) It therefore seems sound to regard the components for PS and PR as partial covariances. They are covariances among residual scores after the general component has been partialled out. Thus, there seems to be a rationale for interpreting negative components as meaningful,also on the level of interactions.While the negative component for A can be directly interpreted to mean that observed scores on the average are negatively correlated, the negative components for interactions would imply negative partial correlations,i.e.partial scores are,on the average, negatively correlated.

What might seem to cause some trouble to a meaningful interpretation of components in TABLE 5,as far as the PS and the PR components are concerned,is the seemingly paradox in the fact that to obtain the components for PS,the PR covariance is used. Likewise,in obtaining the component for PR,the PS covariance is being used.On a more simple level,this is what happened in the introductory hypothetical data presented in TABLE 1,where the row component was the covariance of the columns,and the column component the average row covariance.

Returning now to the Bock data,it should be clear that the MS's in TABLE 3 can be given in terms of the variance-covariance matrix of the four tests.

TABLE 6

Average variance-covariance matrix of Bock's data

	11	12	21	22
11	11,556	3,161	7,044	4,552
12	3,161	11,556	4,552	7,044
21	7,044	4,552	11,556	3,161
22	4,552	7,044	3,161	11,556

By computing the average variance of the four tests and the average values of the heterostimulus-heteroresponse,the heterostimulus-monoresponse,and the monostimulus-heteroresponse covariances,the average variance-covariance matrix as presented in TABLE 6 is obtained.

The variance-covariance matrices for the other three linear combinations will have just the same values in the cells as found in TABLE 6. What differs from TABLE 6 is the signs of the tests. As can be seen from (5),(6),and (7), two of the tests are reflected in each of the linear combinations, the reflection implying that two tests get negative signs.

The four variance-covariance matrices representing the variances of the four linear combinations of the four tests, are all observed matrices. The structures revealed by these matrices are manifest structures.

An inferred variance-covariance structure can be uncovered by introducing the variance components into the variance-covariance matrix. According to TABLE 5, the average variance and the various average covariances can be written as linear combinations of the components. Such a latent variance-covariance matrix is constructed in TABLE 7 for the sum score variance.

TABLE 7

Latent variance-covariance matrix
of Bock's data

	11	12	21	22
11	4,552 +2,492 -1,391 +5,903	4,552 -1,391	4,552 +2,492	4,552
12	4,552 -1,391	4,552 +2,492 -1,391 +5,903	4,552	4,552 +2,492
21	4,552 +2,492	4,552	4,552 +2,492 -1,391 +5,903	4,552 -1,391
22	4,552	4,552 +2,492	4,552 -1,391	4,552 +2,492 -1,391 +5,903

The intuitive logic of the latent variance-covariance matrix is that the general ability in the four tests,represented by the average covariance among heterostimulus-heteroresponse tests, is imposed on the covariances of the heterostimulus-monoresponse and monostimulus-heteroresponse types.The average covariance between T_{11} and T_{21},and between T_{12} and T_{22},is construed to be a sum of two additive components,the hetero-hetero component,σ_a^2, and the hetero-mono component,σ_b^2.The covariance component proper for stimulus means that when the general component is partialled out of the stimulus scores across responses,the residual stimulus scores are still positively correlated,as shown by the positive partial covariance component.

The average covariance between T_{11} and T_{12},and between T_{21} and T_{22},is also construed to be a sum of two additive components,the hetero-hetero component,σ_a^2,and the mono-hetero component,σ_c^2.The covariance component proper for response may be taken to mean that when the general component is partialled out of the response scores across stimuli,the residual response scores are negatively correlated,as shown by the negative partial covariance component.

The average test variance is construed to be a sum of four additive components,the three covariance components plus a residual component,or the triple interaction component.

By collecting components of the same categories in TABLE 7,and summing,the latent structure of the sum score variance is revealed:

$$V_a = V_p = 4.\ 5,903 + 8.\ -1,391 + 8.\ 2,492 + 16.\ 4,552 = 105,252$$

$$MS_a = MS_p = 5,903 + 2.\ -1,391 + 2.\ 2,492 + 4.\ 4,552 = 26,313$$

It is interesting to note that the composition of MS_a,as derived from the latent variance-covariance matrix,has just the same structure as the theoretical model derived from ANOVA.

The inclusion of the negative component in the model should cause no trouble as it is clear that it is a covariance component. Negatively correlated partial scores included in a sum score means that the sum score variance will be attenuated by the negative correlation. Adding negatively correlated scores means a smaller sum score variance as compared to a sum score variance obtained by adding positively correlated scores.

The weighted components, including the negative PS component, add to the sum score variance,

$$5{,}903 + 4{,}984 - 2{,}782 + 18{,}208 = 26{,}313$$
$$0{,}224 + 0{,}189 - 0{,}106 + 0{,}692 = 1{,}000$$

In setting the variance to unit variance, it might be difficult to regard the weighted components as proportions with a negative value included. But the values and the signs tell how much and in what direction the various components affect the sum score variance.

Also, the relative contribution of unweighted components can be found. The unweighted components sum to the average test variance,

$$5{,}903 + 2{,}492 - 1{,}391 + 4{,}552 = 11{,}556$$
$$0{,}511 + 0{,}216 - 0{,}120 + 0{,}394 = 1{,}000$$

By setting the average variance to unit variance, an interesting convergence is seen. The general component value, 0,394, is approximately the average heterostimulus-heteroresponse intertest correlation in Table III in Bock's paper ($\bar{r}_{hh} = 0{,}394$). The general component + the PS component, $0{,}394 - 0{,}120 = 0{,}274$, is approximately equal to the average monostimulus-heteroresponse intertest correlation ($\bar{r}_{mh} = 0{,}284$). Lastly, the general component + the PR component, $0{,}394 + 0{,}216 = 0{,}610$, is approximately equal to the heterstimulus-monoresponse intertest correlation ($\bar{r}_{hm} = 0{,}603$).

The reason why the compared values are not exactly the same, is that the present values are generated from the variance-covariance matrix, while the Bock values are generated from the correlation matrix. The slight discrepancy in values is caused by the fact that the tests do not have exactly the same variances. What is evident in the comparison made here, is generally the case when intraclass correlations are compared to interclass correlations.

It should be clear from the derivations made that an analysis of variance of Bock's data in standardized scores, should make it possible to reconstruct the correlation matrix with average values from the components thus derived.

As the variance of the sum score can be represented as the sum of the latent variance-covariance matrix, so can also the variance of the other three linear combinations be given as sums of the latent variance-covariance matrix. What will be different in these matrices, compared to the matrix of TABLE 7, is the reflection of some of the tests, with the consequence of changing the signs of some of the components. By summing these three matrices, each of them reduces to a structure equal to the ANOVA model.

From all of the four latent variance-covariance matrices it would be possible to get an ostensive conception of the different intraclass correlations connected to the P, the PS, and the PR variances.

The intraclass correlation for P would be the ratio of the average heterostimulus-heteroresponse intertest covariance to average test variance,

$$\text{alpha}_p = \frac{4,552}{11,556} = 0,394$$

The intraclass correlation for PS is the ratio of the average partial covariance for monostimulus-heteroresponse tests to the average partial variance of the partial response scores. This can be done by using the inferred structure, or the model, for the PS, or the B, interaction in TABLE 5. Unweighted components are commonly used for intraclass correlation coefficients.

$$\text{alpha}_{ps} = \frac{\sigma_b^2}{\sigma_d^2 + \sigma_b^2} = \frac{-1,391}{4,512} = -0,308$$

Likewise, the intraclass correlation for PR will be the ratio of the average partial covariance for heterostimulus-monoresponse tests to the average partial variance of the partial stimulus scores. According to the model for PR in TABLE 5, the partial intraclass correlation should be,

$$\text{alpha}_{pr} = \frac{\sigma_c^2}{\sigma_d^2 + \sigma_c^2} = \frac{2,492}{8,395} = 0,297$$

At some length it has been shown that a negative variance component can be easily integrated with other components to account for complex variance structures. As far as syntax is concerned, there should be no doubt that a negative component, when observed, is a necessary part of the structure. Semantically, a negative component is interpreted as zero according to conventional practice. The contention of this discussion is that negative variance components can be substantively meaningful too, at least for the type of design considered here.

To what extent the negative partial intraclass correlation in Bock's data is meaningful in the context of the Language Modalities Survey, is difficult to judge for one who is not well versed in that context. It might seem, though, that/the negative component could be an interesting challenge for substantive interpretation.

A reinterpretation of Osnes' (1971) data.

In an experiment intending to assess the influence of handwriting neatness and composition errors on the marking of essays, Osnes (1971) got negative variance components for two interactions.As is conventionally done,Osnes ignored the negative components by setting them to zero.In the following an effort is made to interpret the negative components as highly meaningful in the context of the substantive problem.

As Osnes' experimental design in crucial aspects,as far as this discussion is concerned, equals Bock's test design,the explorations made into that design may facilitate a re-interpretation of the results obtained by Osnes.

Osnes picked 24 essays of varying quality written by university students in educational psychology at an external examination. Of ecah essay he made four experimental versions.The four versions are combinations of handwriting neatness and composition errors. Handwriting neatness is a dichotomous variable,neat versus poor handwriting. Likewise,error is dichotomized,no composition error versus many composition errors. Thus,the four essay versions are,(11) neat handwriting and no composition error,(12) neat handwriting and many composition errors,(21)poor handwriting and no composition error,and (22) poor handwriting and many composition errors.

Each of the four essay versions was then evaluated and marked by three raters,each essay thus being marked 12 times.As there were only 12 raters altogether,each rater did not mark each of the 24 essays.In order to reduce costs,a counterbalancing arrangement was made in allocating raters to essays.(This counterbalancing arrangement is of no concern for the present discussion.)

The Osnes design is thus a 24 essays by 2 handwriting levels by 2 composition error levels by 3 raters design. It is a factorial repeated measures design, as can be seen from TABLE 8. 288 marks are given.

TABLE 8

The Osnes experimental design

	B_1						B_2					
	C_1			C_2			C_1			C_2		
A_1	X	X	X	X	X	X	X	X	X	X	X	X
A_2	X	X	X	X	X	X	X	X	X	X	X	X
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
A_{23}	X	X	X	X	X	X	X	X	X	X	X	X
A_{24}	X	X	X	X	X	X	X	X	X	X	X	X

Note.- A = essays, B = handwriting neatness, C = composition errors, X = rater within cell.

A descriptive variance components analysis of the Osnes data will now be performed. It is called a descriptive analysis as no account is taken of considering factors as random or fixed in writing the models for the various sources. What is of concern here is to make a structural analysis of variances in order to see where the contributions to variance come from according to the theoretical model, which is an inference of how the data are generated. In a descriptive variance component analysis no statistical generalization is implied at all.

As is evident from the result table, the present design is somewhat more complex than the Bock design as presented in this paper. While the replication variance was ignored in the analysis of Bock's data, the among raters variance within cells is included in the present analysis. This difference between the two designs will have no consequence for the interpretation of the negative components to be considered in the following discussion of the Osnes experiment.

As can be seen from TABLE 9, presenting the ANOVA results of the Osnes data, there are two negative variance components, one

TABLE 9
Osnes' (1971) ANOVA results

Source	SS	df	MS	VC
A Essays	4081,1	23	177,439	14,558
B Handwriting	76,1	1	76,100	
C Errors	224,0	1	224,000	
AB	524,1	23	22,787	-1,224
AC	231,9	23	10,084	-3,341
BC	1,1	1	1,100	
ABC	693,0	23	30,130	5,304
D(ABC) Raters	2730,0	192	14,219	14,219
Total	8561,3	287		

for the essay by handwriting neatness interaction (AB), the other for the essay by composition error interaction (AC).

As the sources of variance including essay is of most interest in the present problem, the variance components are given for those sources only. The components have been derived from the following structural models:

$$MS_a = \sigma^2_{d(abc)} + 3\sigma^2_{abc} + 6\sigma^2_{ac} + 6\sigma^2_{ab} + 12\sigma^2_a$$
$$MS_{ab} = \sigma^2_{d(abc)} + 3\sigma^2_{abc} + 6\sigma^2_{ab}$$
$$MS_{ac} = \sigma^2_{d(abc)} + 3\sigma^2_{abc} + 6\sigma^2_{ac}$$
$$MS_{abc} = \sigma^2_{d(abc)} + 3\sigma^2_{abc}$$
$$MS_{d(abc)} = \sigma^2_{d(abc)}$$

According to the conclusion reached concerning the negative component in Bock's data, which was an interaction component of the same type as the present ones, the interpretation of th e negative variance components in the Osnes data should be that the partial covariances of residualized handwriting neatness scores (error constant) and of residualized composition error scores (neatness constant) are negative, the partial negative covariances being generated after the component due to the general essay quality is partialled out of the original scores, or marks. This can be taken to mean that when the 24 essays are given in two versions, one neatly, the other poorly written; or one with no error, the other with many errors; there is a negative covariance between the partial marks for the two versions, [redacted] general essay quality being partialled out. Because the effect of the general essay quality is running through all marks, whatever the version of the essays are, to such a great extent, the observed correlations among marks are always positive, but on the average less for essays of different handwriting neatness (error constant) and for essays of differing amount of errors (handwriting constant). What would be observable from the manifest correlation matrix, a 4 x 4 matrix, is that the average correlation between heterohandwriting-monoerror marks and between monohandwriting-heteroerror marks are smaller than the average heterohandwriting-heteroerror marks correlation. The model imposes two effects on the hetero-mono and the mono-hetero covariances, one that tends to increase the covariance, the general essay quality effect; and another effect that tends to attenuate, or decrease, the covariance. This attenuating effect materializes in a slight tendency for raters to rank an essay lower when poorly written, higher when neatly written; and to rank an essay lower with many errors, higher with no errors.

How much the residualized marks correlate can be found by computing the partial intraclass correlations for essay by handwriting interaction and essay by composition error interaction, respectively.

The partial intraclass correlations for the two interactions can be found by using the models for MS_{ab} and MS_{ac} on page 22.

$$alpha_{ab} = \frac{\sigma^2_{ab}}{\sigma^2_{d(abc)}+\sigma^2_{abc}+\sigma^2_{ab}} = \frac{-1,224}{14,219+5,304-1,224} = -0,067$$

$$alpha_{ac} = \frac{\sigma^2_{ac}}{\sigma^2_{d(abc)}+\sigma^2_{abc}+\sigma^2_{ac}} = \frac{-3,341}{14,219+5,304-3,341} = -0,206$$

The intraclass correlation for the essay by handwriting interaction should be interpreted as the value of the negative correlation between partial marks for no error/many errors essays. Likewise, the intraclass correlation for the essay by composition error interaction should be interpreted as the value of the negative correlation between partial marks for neat handwriting versus poor handwriting essays.

Concluding remarks.

While negative variance components seem impossible to integrate rationally in a probabilistic ANOVA model, it is reasonable to believe that such components can be naturally integrated in a deterministic ANOVA model, at least for certain types of design.

In repeated measures designs, it can be shown that some of the variance components are covariance components. As components are average values, it should cause no difficulty in accepting negative variance components of this particular type.

A descriptive variance components analyses draws on the deterministic ANOVA model, and purports to decompose data into hypothetical sources according to the theoretical structure implied in the ANOVA model. In fact, the structural model of ANOVA is a theory of how data are generated in terms of the factors included in the design.

The illustrations of repeated measures designs, both hypothetical and realistic, presented in this paper, should make it plausible to accept negative variance components as meaningful, syntactically as well as semantically. This is even the case for negative components generated from mean squares for interactions, which involves partial scores.

For the Osnes data, the negative variance components seem so reasonable that they might have been postulated as a consequence of a serious research hypothesis. The effects of handwriting neatness and composition errors on the marks assigned to essays are within reasonable limits probably consciously defined away by the raters as non-existing. However, to err is humane, and so it might be that what is consciously defined away, can appear as crypto-effects in negative variance components.

References

Eikeland,H.M.Coefficient alpha and the expected variance-covariance matrix of random composite measurements. Oslo 1970. Mimeographed.

Endler,N.S.Estimating variance components from mean squares for random and mixed effects analysis of variance models.Perceptual and Motor Skills,1966,22,559-570.

Gleser,G.C.,Cronbach,L.J.,and Rajaratnam,N.Generalizability of scores influenced by multiple sources of variance.Psychometrika, 1965,30,395-418.

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

Hill, .Inference about variance components in the one-way model.Journal of the American Statistical Association,1965,60, 806-825.

Medley,D.M.and Mitzel,H.E.Measuring classroom behavior by systematic observation.In Gage,N.L.(Editor) Handbook of Research on Teaching.Chicago:Rand McNally,1963.

Nelder, .The interpretation of negative components of variance. Biometrika,1954,41,544-548.

Overall,J.E.,and Spiegel,D.K.Concerning least squares analysis of experimental data.Psychological Bulletin,1969,72,311-322.

Scheffe,H.The Analysis of Variance.New York:John Wiley & Sons. 1959.

Thomson, and Moore, .Non-negative estimates of variance components.Technometrics,1963,5,441-4510.

Vaughan,G.M.and Corballis,M.C.Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. Psychological Bulletin,1969,72,204,213.