Educational and Psychological Measurement
1977, 37

# THE RELIABILITY OF DIFFERENCE SCORES WHEN ERRORS ARE CORRELATED

RICHARD H. WILLIAMS
University of Miami
DONALD W. ZIMMERMAN
Carleton University

The usual formulas for the reliability of differences between two test scores $X$ and $Y$ are based on the assumption that the error scores $E_X$ and $E_Y$ are uncorrelated. In modern developments of test score theory, such as that of Lord and Novick, a true score is defined as the expected value of an individual's observed score. This definition implies that true scores on any test are uncorrelated with error scores on any test, but it does not imply that error scores on distinct tests $X$ and $Y$ are uncorrelated. A zero correlation between the errors can be obtained only by introducing an additional assumption of "experimental independence" that does not follow from the other axioms in the model. This assumption restricts severely the class of random variables to which the usual formulas for reliability of differences will apply. The present paper investigated the reliability of difference scores in more general cases where it is not assumed that error scores on distinct tests are uncorrelated. The formulas derived are relatively simple, and they reduce to the usual ones of the classical model when $E_X$ and $E_Y$ are uncorrelated.

THE reliability of difference scores has been studied extensively in test theory. Formulas for the reliability of differences between two test scores apparently are quite useful, since scores of this type arise frequently in practice. However, as Lord and Novick (1968), Stanley (1967), and others have pointed out, these formulas depend on certain restrictive assumptions and sometimes are difficult to apply correctly in testing situations.

A well-known formula for the reliability of difference scores is

$$\rho_{DD'} = \frac{(\rho_{XX'} + \rho_{YY'})/2 - \rho_{XY}}{1 - \rho_{XY}} \tag{1}$$

(see, for example, Gulliksen, 1950). Here, $\rho_{XY}$ denotes the correlation between observed scores $X$ and $Y$, $\rho_{XX'}$ and $\rho_{YY'}$ the reliability coefficients of the two tests, and $\rho_{DD'}$ the reliability of the difference scores, $D = X - Y$. As several writers have pointed out, this formula is limited in usefulness because it depends on the assumption that $\sigma_X = \sigma_Y$, i.e., that the standard deviations of the two tests are equal. Stanley (1967) has pointed out that even when the tests used have been scaled so that their standard deviations are equal, the actual computed standard deviations will not necessarily be equal. In any case, equation (1) is inappropriate in a situation where $\sigma_X$ and $\sigma_Y$ differ greatly.

A formula that does not require equality of $\sigma_X$ and $\sigma_Y$ is

$$\rho_{DD'} = \frac{\rho_{XX'} \operatorname{Var} X + \rho_{YY'} \operatorname{Var} Y - 2\rho_{XY}\sigma_X\sigma_Y}{\operatorname{Var} X + \operatorname{Var} Y - 2\rho_{XY}\sigma_X\sigma_Y} \tag{2}$$

(see, for example, Lord, 1963, Stanley, 1967). Note that when $\sigma_X = \sigma_Y$ (2) reduces to (1).

Both of the above formulas for reliability of difference scores depend on another assumption that has been made routinely in test theory when scores on two tests $X$ and $Y$ are considered. This is the assumption that $\rho(E_X, E_Y) = 0$, that is, that the error scores $E_X$ and $E_Y$ on the two tests are uncorrelated. In the test theory model developed by Lord and Novick (1968) and others, in which a true score is identified with the expected value of an individual's observed score, it follows from the basic definitions that true scores on any test and error scores on any test are uncorrelated. However, it does not follow from the expected-value concept that error scores on arbitrary tests $X$ and $Y$ are uncorrelated, unless an axiom of "experimental independence" or "linear experimental independence" is added to the theory.

Although relatively little attention has been paid to this assumption, it severely restricts the class of random variables to which formulas such as the above apply, and these formulas may not realistically describe testing situations. The purpose of the present paper is to derive more general formulas for the reliability of difference scores which apply in situations where errors of measurement are correlated. We will begin with essentially the same approach to test reliability developed by Lord and Novick and others, but will then proceed without making the usual assumption of linear experimental independence of scores on distinct tests $X$ and $Y$. Our concern, then, is somewhat similar to that of Guttman (1953) and of Rozeboom (1966), both of whom had reservations about the concept of experimental independence in reliability theory. The results, we will see, are not as complicated as might be expected, and the formulas obtained reduce to the usual ones when $\rho(E_X, E_Y) = 0$.

*Test Reliability and the Reliability of Difference Scores*

We consider observed score random variables $X$ and $Y$ defined on an arbitrary but fixed probability space. The corresponding true scores $T_X$ and $T_Y$ and the error scores $E_X = X - T_X$ and $E_Y = Y - T_Y$ are constructed as in a previous paper (Zimmerman, 1975).

Our concept of true score, on which the derivations in the sequel are based, is essentially the same as that of Lord and Novick (1968). As in the Lord-Novick model, it follows from the construction of these random variables that true scores on any test are uncorrelated with error scores on any test—that is, $\rho(T_X, E_Y) = 0$, $\rho(T_X, E_X) = 0$, and so on. But, as mentioned above, it does not follow from these axioms that $\rho(E_X, E_Y) = 0$, and in the present paper we do not make this assumption. That is, we do not introduce an assumption of linear experimental independence, as done in the Lord-Novick model.

The usual formulas for test reliability in the classical test-theory model can be obtained from these axioms. We summarize briefly some relations needed for derivations in the present paper (see also Zimmerman, 1975). From the concept of true score mentioned above it follows that, for any observed score random variables $X$ and $Y$, and for all real numbers $a$ and $b$,

$$T_{aX+bY} = aT_X + bT_Y, \tag{3}$$

and, in particular, $T_{X-Y} = T_X - T_Y$. The reliability coefficient $\rho_{XX'}$ is defined as follows

$$\rho_{XX'} = \frac{\text{Var } T_X}{\text{Var } X}. \tag{4}$$

It can be proved that $\rho_{XX'} = \rho^2(X, T_X)$, that is, reliability is the squared correlation between observed scores and true scores and also that $\rho_{XX'}$ is the correlation between two random variables $X$ and $X'$ that are *parallel* as usually defined in test theory, a fact that justifies the above notation. Also,

$$\sigma_{E_X} = \sigma_X \sqrt{1 - \rho_{XX'}}, \tag{5}$$

which is the usual expression for the standard error of measurement. A relation not found in the classical test theory is

$$\text{Cov } (X, Y) = \text{Cov } (T_X, T_Y) + \text{Cov } (E_X, E_Y) \ . \tag{6}$$

If Cov $(E_X, E_Y) = 0$, this equation reduces to a familiar result. In the present paper, however, the more general form given above is used.

Now, let $X$ and $Y$ be two observed score random variables defined on the same probability space. A *difference score* is just the random variable $D = X - Y$, the pointwise difference of $X$ and $Y$. The *reliability of difference scores* is defined in a way analogous to $\rho_{XX'}$. That is,

$$\rho_{DD'} = \frac{\text{Var } T_D}{\text{Var } D} = \frac{\text{Var } T_{X-Y}}{\text{Var }(X - Y)}. \tag{7}$$

As before $\rho_{DD'} = \rho^2(D, T_D)$, and, furthermore, $\rho_{DD'}$ is the correlation between parallel random variables D and D′—or, otherwise expressed, it is the correlation between $X - Y$ and $X' - Y'$, where $X$ and $X'$ are parallel and $Y$ and $Y'$ are parallel.

We can now prove the following theorem, which provides a formula for the reliability of difference scores for the most general case, where $\rho(E_X, E_Y)$ is not necessarily zero and $\sigma_X$ and $\sigma_Y$ are not necessarily equal.

*Theorem 1.*

Let $X$ and $Y$ be observed-score random variables with finite non-zero variance, with corresponding true scores $T_X$ and $T_Y$, such that the random variables $D = X - Y$, $E_X = X - T_X$, and $E_Y = Y - T_Y$ have non-zero variance, and let $\lambda = \sigma_X/\sigma_Y$. Then,

$$\rho_{DD'} = \frac{(\lambda\rho_{XX'} + \lambda^{-1}\rho_{YY'})/2 - \rho_{XY}}{(\lambda + \lambda^{-1})/2 - \rho_{XY}} + \frac{\rho(E_X, E_Y)\sqrt{(1 - \rho_{XX'})(1 - \rho_{YY'})}}{(\lambda + \lambda^{-1})/2 - \rho_{XY}}. \tag{8}$$

*Proof.* Since, by (7), $\rho_{DD'} = \text{Var } T_{X-Y}/\text{Var }(X - Y)$, by (3), $\text{Var } T_{X-Y} = \text{Var }(T_X - T_Y) = \text{Var } T_X + \text{Var } T_Y - 2 \text{ Cov }(T_X, T_Y)$ and $\text{Var }(X - Y) = \text{Var } X + \text{Var } Y - 2 \text{ Cov }(X, Y)$, and by (6), $\text{Cov }(T_X, T_Y) = \text{Cov }(X, Y) - \text{Cov }(E_X, E_Y)$, we obtain

$$\rho_{DD'} = \frac{\text{Var } T_X + \text{Var } T_Y - 2 \text{ Cov }(X,Y) + 2 \text{ Cov }(E_X, E_Y)}{\text{Var } X + \text{Var } Y - 2 \text{ Cov }(X,Y)}, \tag{9}$$

or, using (4) and (5),

$$\rho_{DD'} = \frac{\rho_{XX'} \text{Var } X + \rho_{YY'} \text{Var } Y - 2\rho_{XY}\sigma_X\sigma_Y + 2\rho(E_X, E_Y)\sigma_X\sigma_Y\sqrt{(1 - \rho_{XX'})(1 - \rho_{YY'})}}{\text{Var } X + \text{Var } Y - 2\rho_{XY}\sigma_X\sigma_Y}$$

Dividing numerator and denominator by $2\sigma_X\sigma_Y$, substituting $\lambda = \sigma_X/\sigma_Y$, and rearranging gives (8). This completes the proof.

Equation (8) expresses the reliability of difference scores as a function of the reliability coefficients of the two tests, the correlation between the observed scores of the two tests, and the correlation between error scores, under general conditions where $\rho(E_X, E_Y)$ is not necessarily zero and $\sigma_X$ and $\sigma_Y$ are not necessarily equal. If $\sigma_X = \sigma_Y$, so that $\lambda = \lambda^{-1} = 1$, (8) reduces to

$$\rho_{DD'} = \frac{(\rho_{XX'} + \rho_{YY'})/2 - \rho_{XY}}{1 - \rho_{XY}} + \frac{\rho(E_X, E_Y)\sqrt{(1 - \rho_{XX'})(1 - \rho_{YY'})}}{1 - \rho_{XY}} . \tag{10}$$

If $\rho(E_X, E_Y) = 0$, (8) reduces to

$$\rho_{DD'} = \frac{(\lambda\rho_{XX'} + \lambda^{-1}\rho_{YY'})/2 - \rho_{XY}}{(\lambda + \lambda^{-1})/2 - \rho_{XY}} ,$$

which is equivalent to (2). And if $\sigma_X = \sigma_Y$ and $\rho(E_X, E_Y) = 0$, (8) reduces to (1).

The above derivation makes clear that only the *ratio* of the standard deviations influences the reliability of difference scores, and not the specific values of the standard deviations of the two tests. It can be seen by inspecting equation (8) that, if the error scores on the two tests are positively correlated, the reliability of difference scores will be greater than the value given by equation (2), and if error scores are negatively correlated, the reliability of difference scores will be less than the value given by (2). Investigators using formulas such as (1) and (2) have usually taken for granted that "the difference between two fallible measures is frequently much more fallible than either (Lord, 1963, p. 132)." An interesting fact that emerges from the above development, however, is that, if errors are positively correlated, it is possible for the reliability of difference scores to be higher than the reliability of either of the original tests. Examples of this paradoxical result are considered in the sequel.

### *The Standard Error of Measurement of Difference Scores*

When errors are correlated, the standard error of measurement of difference scores is modified in the following way. Since

$$\text{Var } E_D = \text{Var }(E_X - E_Y) = \text{Var } E_X + \text{Var } E_Y - 2 \text{ Cov }(E_X, E_Y),$$

substitutions using (4) and (5) give the result

$$\text{Var } E_D = (\text{Var } X)(1 - \rho_{XX'}) + (\text{Var } Y)(1 - \rho_{YY'}) - 2\rho(E_X, E_Y)\sigma_X\sigma_Y \sqrt{(1 - \rho_{XX'})(1 - \rho_{YY'})}, \tag{11}$$

and the square root of (11) is the standard error of measurement of difference scores. If $\rho(E_X, E_Y) = 0$, this reduces to the classical formula

$$\sigma_{E_D} = \sqrt{(\text{Var } X)(1 - \rho_{XX'}) + (\text{Var } Y)(1 - \rho_{YY'})}.$$

When error scores are positively correlated, confidence intervals constructed on difference scores will be smaller than those obtained from the classical formula.

### *Some Comments On the Nature of Error Score Covariance*

The extent to which these more general equations differ from equations such as (1) or (2) depends on the magnitude of Cov $(E_X, E_Y)$. We now examine the nature of this covariance term. It is well-known that the unconditional variance of a random variable equals the variance of the conditional expectation plus the expectation of the conditional variance. A similar relation holds for covariance (see, for example, Lord and Novick, p. 51, exercise 2.16). In the present context this means that

$$\text{Cov}\,(X, Y) = \text{Cov}\,(T_X, T_Y) + \mathcal{E}\ \text{Cov}\,(X \mid f, Y \mid f),$$

where $X \mid f$ and $Y \mid f$ denote observed scores restricted to individuals, and $\mathcal{E}$ is the expected value. Comparing this expression to equation (6), we see that

$$\text{Cov}\,(E_X, E_Y) = \mathcal{E}\ \ \text{Cov}\,(X \mid f, Y \mid f). \tag{12}$$

Hence, Cov $(E_X, E_Y)$ is an average of the covariances between the observed score random variables $X \mid f$ and $Y \mid f$ for each individual taking the test (just as total error variance is an average of individual error variances). Note that *individual* error score variances and observed score variances are equal since they differ by a constant true score, and similarly for covariances.

Any factor which would make an individual's scores on two tests more alike than usual would tend to produce a positive value of Cov $(X \mid f, Y \mid f)$ for the individual and thereby contribute to a positive value of Cov $(E_X, E_Y)$. For example, if two tests are given at the same sitting and if a person is sick, temporarily indisposed, unusually sensitive to noise in the test environment, if cheating occurs, and so on, the scores on both tests could be affected together in a non-random way (see also Rozeboom, 1966). We will provide a numerical example designed to illustrate the possible influence of such factors below (see Example 2).

### *Computer Generation of Values of the Reliability of Difference Scores*

A computer program written in FORTRAN IV was used to generate values of $\rho_{DD'}$ for various combinations of $\rho_{XX'}$, $\rho_{YY'}$, $\rho_{XY'}$, and $\rho(E_X, E_Y)$. Each of the latter four coefficients were assigned values ranging from .20 to .80 with increments of .20 so that there were 256

TABLE 1
*Values for the Reliability of Difference Scores**

| | $\rho(X, Y)$ | .20 | | | | | .40 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho(E_X, E_Y)$ | $\rho_{YY'}$ \ $\rho_{XX'}$ | .20 | .40 | .60 | .80 | $\rho_{YY'}$ \ $\rho_{XX'}$ | .20 | .40 | .60 | .80 |
| .20 | .20 | .20 | .30 | .39 | .48 | .20 | — | .06 | .19 | .30 |
| | .40 | | .40 | .50 | .59 | .40 | | .20 | .33 | .45 |
| | .60 | | | .60 | .70 | .60 | | | .47 | .59 |
| | .80 | | | | .80 | .80 | | | | .73 |
| .40 | .20 | .40 | .47 | .53 | .58 | .20 | .20 | .30 | .38 | .43 |
| | .40 | | .55 | .62 | .67 | .40 | | .40 | .49 | .56 |
| | .60 | | | .70 | .77 | .60 | | | .60 | .69 |
| | .80 | | | | .85 | .80 | | | | .80 |
| .60 | .20 | —** | .64 | .67 | .68 | .20 | .47 | .53 | .57 | .57 |
| | .40 | | .70 | .74 | .76 | .40 | | .60 | .66 | .68 |
| | .60 | | | .80 | .84 | .60 | | | .73 | .78 |
| | .80 | | | | .90 | .80 | | | | .87 |
| .80 | .20 | — | — | .82 | .78 | .20 | — | .76 | .75 | .70 |
| | .40 | | .85 | .86 | .85 | .40 | | .80 | .82 | .80 |
| | .60 | | | .90 | .91 | .60 | | | .87 | .88 |
| | .80 | | | | .95 | .80 | | | | .93 |

* The values for $\rho_{DD'}$ shown in these sub-tables were computed from equation (10)

** The value for $\rho_{DD'}$ is omitted for all combinations not satisfying inequality (13).

possible combinations in all. Equation (10) was used for the computations. As the following development will show, not all of the 256 possibilities are meaningful.

By the Cauchy-Schwartz inequality,

$$| \operatorname{Cov}(T_X, T_Y) | \leq \sigma_{T_X}\sigma_{T_Y}.$$

We again use equation (6) to write

$$| \operatorname{Cov}(X, Y) - \operatorname{Cov}(E_X, E_Y) | \leq \sigma_{T_X}\sigma_{T_Y}.$$

Dividing the entire inequality by $\sigma_X\sigma_Y$ produces

$$\left| \frac{\operatorname{Cov}(X, Y)}{\sigma_X\sigma_Y} - \frac{\operatorname{Cov}(E_X, E_Y)}{\sigma_X\sigma_Y} \right| \leq \frac{\sigma_{T_X}\sigma_{T_Y}}{\sigma_X\sigma_Y},$$

TABLE 1 (*Continuation*)

| | | $\rho(X, Y)$ = .60 | | | | $\rho(X, Y)$ = .80 | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\rho(E_X, E_Y)$ | $\rho_{YY'}$ \ $\rho_{XX'}$ | .20 | .40 | .60 | .80 | .20 | .40 | .60 | .80 |
| .20 | .20 | — | — | — | — | — | — | — | — |
| | .40 | | — | — | .17 | | — | — | — |
| | .60 | | | .20 | .39 | | | — | — |
| | .80 | | | | .60 | | | | .20 |
| .40 | .20 | — | — | — | — | — | — | — | — |
| | .40 | | .10 | .24 | .35 | | — | — | — |
| | .60 | | | .40 | .53 | | | — | .07 |
| | .80 | | | | .70 | | | | .40 |
| .60 | .20 | .20 | .29 | .35 | .35 | — | — | — | — |
| | .40 | | .40 | .48 | .52 | | — | — | — |
| | .60 | | | .60 | .67 | | | .20 | .35 |
| | .80 | | | | .80 | | | | .60 |
| .80 | .20 | .60 | .64 | .63 | .55 | .20 | .27 | — | — |
| | .40 | | .70 | .73 | .69 | | .40 | .46 | .39 |
| | .60 | | | .80 | .82 | | | .60 | .63 |
| | .80 | | | | .90 | | | | .80 |

or, in terms of correlation coefficients, we have

$$\left|\rho_{XY} - \frac{\rho(E_X, E_Y)\sigma_{E_X}\sigma_{E_Y}}{\sigma_X\sigma_Y}\right| \leq \sqrt{\rho_{XX'}\rho_{YY'}}\ .$$

Applying (5) then gives

$$\left|\ \rho_{XY} - \rho(E_X, E_Y)\sqrt{(1 - \rho_{XX'})(1 - \rho_{YY'})}\ \right| \leq \sqrt{\rho_{XX'}\rho_{YY'}}. \quad (13)$$

This inequality now contains just the four correlation terms that appear in the right-hand side of equation (10). Only combinations which satisfy this inequality will yield a meaningful value for $\rho_{DD'}$. It is interesting to note that if $\rho(E_X, E_Y)$ is set equal to zero in (13), the result is the more familiar inequality

$$|\ \rho_{XY}\ | \leq \sqrt{\rho_{XX'}\rho_{YY'}}.$$

The results generated by using the computer program are shown in

TABLE 2
*Sample Space For Example 1*

| $\omega$ | $p(\omega)$ | $f(\omega)$ | $X(\omega)$ | $Y(\omega)$ | $(X - Y)(\omega)$ | $T_X(\omega)$ | $T_Y(\omega)$ | $T_{X-Y}(\omega)$ | $E_X(\omega)$ | $E_Y($ |
|---|---|---|---|---|---|---|---|---|---|---|
| $\omega_1$ | 1/3 | $\alpha_1$ | 0 | 0 | 0 | 1 | 1/2 | 1/2 | −1 | −1/2 |
| $\omega_2$ | 1/3 | $\alpha_1$ | 2 | 1 | 1 | 1 | 1/2 | 1/2 | 1 | 1/2 |
| $\omega_3$ | 1/3 | $\alpha_2$ | 0 | 1 | −1 | 0 | 1 | −1 | 0 | 0 |

Table 1. Since each of the sub-tables are symmetrical, the lower off-diagonal triangular portions have been omitted. Even a cursory inspection of the values for $\rho_{DD'}$ in Table 1 indicates that the reliability of difference scores is as large as or larger than the corresponding values for $\rho_{XX'}$ or $\rho_{YY'}$ for many of the admissible combinations. It is seen that even when the correlation between the error scores is as small as .20, the values of $\rho_{DD'}$ are somewhat larger than the values which would be obtained if equation (1) were used instead of equation (10), the discrepancies being greatest for large values of $\rho_{XY}$.

## *Some Examples*

*Example 1.* We now introduce a numerical example which provides checks on equations (9) and (12). The sample space shown in Table 2, which displays a fundamental probability set $\Omega = \{\omega_1, \omega_2, \omega_3\}$, a discrete probability density $p$, a random point $f$, and random variables $X$ and $Y$, together with other random variables obtained subsequently from this information, has been constructed so that the errors are correlated. When variances, covariances, and correlations are computed for the random variables in Table 2 and substituted into equation (9), the resulting value for $\rho_{DD'}$ is .75. The same value for $\rho_{DD'}$ is obtained when Var $T_{X-Y}$/Var $(X - Y)$ is evaluated. When the appropriate values are substituted into equation (2), however, the value produced for the reliability of difference scores is meaningless (viz., $\rho_{DD'} = -.25$). Making a second check from Table 2, we have Cov $(E_X, E_Y) = \mathcal{E}$ Cov $(X \mid f, Y \mid f) = 1/3$, which substantiates equation (12).

TABLE 3
*First Sample Space For Example 2*

| $\omega$ | $p(\omega)$ | $f(\omega)$ | $X(\omega)$ | $Y(\omega)$ | $T_X(\omega)$ | $T_Y(\omega)$ | $E_X(\omega)$ | $E_Y(\omega)$ |
|---|---|---|---|---|---|---|---|---|
| $\omega_1$ | 1/4 | $\alpha_1$ | 1 | 0 | 1 | 1 | 0 | −1 |
| $\omega_2$ | 1/4 | $\alpha_1$ | 1 | 1 | 1 | 1 | 0 | 0 |
| $\omega_3$ | 1/4 | $\alpha_1$ | 1 | 2 | 1 | 1 | 0 | 1 |
| $\omega_4$ | 1/4 | $\alpha_2$ | 2 | 2 | 2 | 2 | 0 | 0 |

TABLE 4
*Second Sample Space For Example 2*

| $\omega$ | $p(\omega)$ | $f(\omega)$ | $X(\omega)$ | $Y(\omega)$ | $T_X(\omega)$ | $T_Y(\omega)$ | $E_X(\omega)$ | $E_Y(\omega)$ |
|---|---|---|---|---|---|---|---|---|
| $\omega_1$ | 1/4 | $\alpha_1$ | 1 | 0 | 2/3 | 2/3 | 1/3 | −2/3 |
| $\omega_2$ | 1/4 | $\alpha_1$ | 0 | 0 | 2/3 | 2/3 | −2/3 | −2/3 |
| $\omega_3$ | 1/4 | $\alpha_1$ | 1 | 2 | 2/3 | 2/3 | 1/3 | 4/3 |
| $\omega_4$ | 1/4 | $\alpha_2$ | 2 | 2 | 2 | 2 | 0 | 0 |

*Example 2.* This example provides further insight into equation (12) and to the nature of the correlation between error scores. Notice that for the sample space shown in Table 3 the test score for person $\alpha_1$ on both tests on occasion $\omega_2$ is 1. For this sample space we also have Cov $(E_X, E_Y) = 0$. Now suppose instead that person $\alpha_1$ is ill when the tests are administered on occasion $\omega_2$ so that his scores on both test $X$ and test $Y$ are reduced by one point. The sample space displayed in Table 4 reflects this situation. For this latter sample space we now obtain Cov $(E_X, E_Y) = 1/6$. Factors may exist, then, in an actual testing situation, which make an individual's observed score covariance positive, and which thereby elevate both the error covariance and the reliability of the difference scores.

*Example 3.* Let $\sigma_X = \sigma_Y$, $\rho_{XX'} = \rho_{YY'}$, and $\rho_{XY} = \rho(E_X, E_Y)$. Then, applying (10) gives

$$\rho_{DD'} = \frac{\rho_{XX'} - \rho_{XY}}{1 - \rho_{XY}} + \frac{\rho_{XY}(1 - \rho_{XX'})}{1 - \rho_{XY}} = \rho_{XX'},$$

so that, under these conditions, the reliability of the difference scores is the same as the reliability of the original tests.

The present results suggest, therefore, that difference scores may be more reliable than test theorists have previously believed, and can even be more reliable than either of the original tests, if errors are positively correlated. Evidence provided in this paper from some computer calculations over a range of values of the correlation terms, from generation of "toy" sample spaces, and from mathematical considerations indicates that one may expect a positive correlation between error scores in many actual testing situations. It would appear, then, that the correlation between error scores on two distinct tests $X$ and $Y$ is not just a "pathological case" but rather has definite practical significance.

## REFERENCES

Gulliksen, H. *Theory of mental tests.* New York: Wiley, 1950.
Guttman, L. Reliability formulas that do not assume experimental independence. *Psychometrika,* 1953, 18, 123–130.

Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison, Wisconsin: University of Wisconsin Press, 1963.

Lord, F. M. and Novick, M. *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley, 1968.

Rozeboom, W. *Foundations of the theory of prediction*. Homewood, Illinois: Dorsey Press, 1966.

Stanley, J. C. General and special formulas for the reliability of differences. *Journal of Educational Measurement*, 1967, 4, 249–252.

Zimmerman, D. W. Probability spaces, Hilbert spaces, and the axioms of test theory. *Psychometrika*, 1975, 40, 395–412.