PSYCHOMETRIKA—VOL. 32, NO. 1
MARCH, 1967

COEFFICIENT ALPHA AND THE RELIABILITY OF COMPOSITE MEASUREMENTS

MELVIN R. NOVICK
EDUCATIONAL TESTING SERVICE*

AND

CHARLES LEWIS
PRINCETON UNIVERSITY

Following a general approach due to Guttman, coefficient α is rederived as a lower bound on the reliability of a test. A necessary and sufficient condition under which equality is attained in this inequality and hence that α is equal to the reliability of the test is derived and shown to be closely related to the recent redefinition of the concept of parallel measurements due to Novick. This condition is then also shown to be closely related to the unit rank assumption originally adopted by Kuder and Richardson in the derivation of their formula 20. The assumption later adopted by Jackson and Ferguson and the one adopted by Gulliksen are shown to be related to the necessary and sufficient condition derived here. It is then pointed out that the statement that "coefficient α is equal to the mean of the split-half reliabilities" is true only under the restricted condition assumed by Cronbach in the body of his derivation of this result. Finally some limitations on the uses of any function of α as a measure of internal consistency are noted.

In the study of the reliability of composite measurements within the context of the classical test theory model, the function *coefficient* α, so named and extensively studied by Cronbach [1] and in its general or special forms studied by Cureton [3], Dressel [4], Guttman [6], Hoyt [7] Jackson and Ferguson [8], Kuder and Richardson [9], Lyerly [11], Rulon [13], and others, plays a most central role. In the matrix sampling model, (Lord [10]), this function again finds application. Considering the large number of papers that have been devoted to the study of this function, the relative completeness of Cronbach's [1] classic treatment of the properties of α and the length of time that it has been subject to very close scrutiny, it is surprising to find that no really complete and satisfactory general mathematical treatment of the *derivation* of this function exists. The major gap in the mathematical

*Research reported herein was supported in part by the Logistics and Mathematical Statistics Branch of the Office of Naval Research under contract Nonr-4866(00), NR 042-249, Melvin R. Novick, principal investigator. Reproduction, translation, publication, use and disposal in whole or in part by or for the United States Government is permitted. The authors are indebted to Frederic M. Lord, Michael Browne and an unknown referee for constructive criticism of earlier drafts of this manuscript.

treatment is the specification of a meaningful, *necessary and sufficient* condition under which α is equal to the reliability of the test. The major purpose of the present paper is to close that gap.

Within the framework of the classical mental test theory model, as recently reformulated by Novick [12], we begin by adopting Guttman's approach in which coefficient α is derived as a lower bound on the reliability of a composite test. However, by adopting the definition of parallelism due to Novick [12] and introducing related definitions of τ-equivalence and essential τ-equivalence, we are able to extend Guttman's work by noting a necessary and sufficient condition under which equality is attained (and hence that coefficient α is equal to the reliability of the test) and to interpret this condition in meaningful test theoretic terms. This resolves (though only obliquely) a recent point of contention (Cronbach [2] and Cureton [3]) concerning the relationship between two sets of assumptions from which α may be derived as the reliability of the composite test. This simple analysis further permits us to incorporate some of the major results concerning coefficient α into one very simple and straightforward development and to relate this development to previous derivations of coefficient α. For further interpretations, applications, and counsel the reader may refer to Cronbach's [1] classic treatment and discussion. On several points, however, we are able to supplement and clarify some of Cronbach's discussion.

1. *Definitions and Preliminary Results*

Let $Y_1, Y_2, \cdots, Y_n$ be observed-score random variables, taking values $y_1, y_2, \cdots, y_n$, and having finite variance. Let $\mathrm{T}_1, \mathrm{T}_2, \cdots, \mathrm{T}_n$ and $E_1, E_2, \cdots, E_n$, taking values $\tau_1, \tau_2, \cdots, \tau_n$ and $e_1, e_2, \cdots, e_n$, be the associated true- and error-score random variables *constructed*, as detailed by Novick [12], so that the classical test theory model is obtained. Then, with sampling *over persons*, and denoting an expectation by $\mathcal{E}$ and a correlation by ρ, we have, for the population $\mathcal{P}$ under consideration or any proper subpopulation thereof,

$$(1.1) \qquad Y_g = \mathrm{T}_g + E_g, \qquad g = 1, 2, \cdots, n$$

$$(1.2) \qquad \mathcal{E}[E_g \mid \mathrm{T}_g = \tau_g] = 0, \quad \text{for all} \quad \tau_g, \qquad g = 1, 2, \cdots, n.$$

A further assumption of linear experimental independence yields

$$(1.3) \qquad \rho(E_g, E_h) = 0, \quad \text{for all} \quad g, h = 1, 2, \cdots, n, \quad \text{and} \quad g \neq h;$$

and

$$(1.4) \qquad \rho(E_g, \mathrm{T}_h) = 0, \quad \text{for all} \quad g, h = 1, 2, \cdots, n, \quad \text{and} \quad g \neq h.$$

Equation (1.2) implies the more conventional assumptions:

$$(1.2a) \qquad \rho(E_g, \mathrm{T}_g) = 0$$

and

(1.2b) $\mathcal{E}(E_g) = 0$, in $\mathcal{P}$ or any proper subpopulation thereof.

Let

$$X = \sum_{g=1}^{n} Y_g \tag{1.5}$$

be the *composite* observed (test) score with *component* observed (item) scores $Y_1, Y_2, \cdots, Y_n$. Then it is easily seen that

$$\mathrm{T} = \sum_{g=1}^{n} \mathrm{T}_g \tag{1.6}$$

and

$$E = \sum_{g=1}^{n} E_g \tag{1.7}$$

are the classical true- and error-score random variables corresponding to X and satisfying (1.1) and (1.2). Denoting a variance by σ^2, the reliability of the composite measurement X is defined as the variance ratio $\rho^2(X, \mathrm{T}) \equiv \sigma_{\mathrm{T}}^2/\sigma_X^2$. We shall be concerned with the relationship of this quantity to the variances, covariances and reliabilities of the component measurements $Y_1, Y_2, \cdots, Y_n$. The results which we derive in sec. four and five depend heavily on the following definition:

Definition 1.1: Measurements g and g' are *parallel* if, for every person a in the fixed population $\mathcal{P}$ under consideration, the true scores τ_{ga} and $\tau_{g'a}$ of person a on measurements g and g' are equal and the error variances *over replications* for fixed person a, which we denote by $\sigma^2(E_{ga})$ and $\sigma^2(E_{g'a})$, are also equal.

This definition states, in a very strong sense, that measurements g and g' measure the same thing and, in fact, do so with equal precision for every person. Immediate consequences of this definition are the results

$$\mathrm{T}_g \equiv \mathrm{T}_{g'} \tag{1.8}$$

and

$$\sigma^2(E_g) = \sigma^2(E_{g'}) \tag{1.9}$$

over random sampling of persons from *any subpopulation of* $\mathcal{P}$, where T_g and $\mathrm{T}_{g'}$ are the true-score random variables and E_g and $E_{g'}$ are the error-score random variables for measurements g and g'. Equation (1.8) states that the true-score random variables T_g and $\mathrm{T}_{g'}$ (varying over persons) are identical. Equation (1.9) states that the error variances for randomly selected people on measurements g and g' are equal. In the context of the classical test theory model, which deals only with first- and second-order moments, parallel measurements are completely interchangeable. However, in making inferences about higher-order moments of the error distributions an even more stringent equivalence relation is required.

It is often convenient to consider measurements which "measure the same thing", though perhaps not equally well. Hence we have, using the notation of Definition 1.1,

Definition 1.2: Measurements g and g' are *τ-equivalent* if, for every person a in the fixed population $\mathcal{P}$, $\tau_{ga} = \tau_{g'a}$.
An immediate consequence of Definition 1.2 is that (1.8) holds in every subpopulation of $\mathcal{P}$. Also for sect. three we will need

Definition 1.3: Measurements g and g' are *essentially τ-equivalent* if for every person a in the fixed population $\mathcal{P}$ $\tau_{ga} = a_{gg'} + \tau_{g'a}$, where $a_{gg'}$ is a constant dependent on the measurements g and g'.

The principal test theory results that we shall require are contained in the following two lemmas.

Lemma 1.1: For arbitrary measurements Y_g and Y_h with corresponding true scores T_g and T_h the covariance between the observed scores is equal to the covariance between the true scores. Denoting a covariance by $\sigma(\cdot\,,\cdot)$ we have

$$\sigma(Y_g\,, Y_h) = \sigma(T_g\,, T_h). \tag{1.10}$$

Proof:

$$\begin{aligned}\sigma(Y_g, Y_h) &= \sigma[(T_g + E_g), (T_h + E_h)]\\ &= [\sigma(T_g\,, T_h) + \sigma(T_g\,, E_h) + \sigma(T_h\,, E_g) + \sigma(E_g\,, E_h)].\end{aligned}$$

But $\sigma(E_g\,, E_h) = 0$ by (1.3); and $\sigma(T_g\,, E_h) = \sigma(T_h\,, E_g) = 0$ by (1.4). Hence the result follows. Also we shall require the well-known result

Lemma 1.2: If X_g and $X_{g'}$ are parallel measurements, then

$$\sigma(T_g\,, T_{g'}) = \sigma^2(T_g).$$

Proof: This follows immediately on noting that $T_g \equiv T_{g'}$.

2. *Derivation of Coefficient Alpha as a Lower Bound on the Reliability of the Composite Measurement*

We begin with the obvious statement

$$[\sigma(T_g) - \sigma(T_h)]^2 \geq 0 \tag{2.1}$$

or

$$\sigma^2(T_g) + \sigma^2(T_h) \geq 2\sigma(T_g)\sigma(T_h).$$

The Cauchy–Schwartz inequality, in this context, states that

$$\sigma(T_g)\sigma(T_h) \geq |\sigma(T_g, T_h)| \tag{2.2}$$

(i.e., that the correlation coefficient $\rho(T_g, T_h) = \sigma(T_g, T_h)/\sigma(T_g)\sigma(T_h)$ is, in absolute value, less than or equal to one). Hence

$$\sigma^2(T_g) + \sigma^2(T_h) \geq 2\,|\sigma(T_g, T_h)| \geq 2\sigma(T_g, T_h). \tag{2.3}$$

Summing for $g \neq h$ we have

$$\sum_{g \neq h}\sum [\sigma^2(T_g) + \sigma^2(T_h)] \geq 2 \sum_{g \neq h}\sum \sigma(T_g, T_h). \tag{2.4}$$

Now

$$\sum_{g=1}^{n}\sum_{h=1}^{n} [\sigma^2(T_g) + \sigma^2(T_h)] = \sum_{g=1}^{n}\left[n\sigma^2(T_g) + \sum_{h=1}^{n}\sigma^2(T_h)\right]$$

$$= n\sum_{g=1}^{n}\sigma^2(T_g) + n\sum_{h=1}^{n}\sigma^2(T_h) = 2n\sum_{g=1}^{n}\sigma^2(T_g).$$

Also

$$\sum_{g=1}^{n}\sum_{h=1}^{n} [\sigma^2(T_g) + \sigma^2(T_h)]$$

$$= \sum_{g=h}\sum [\sigma^2(T_g) + \sigma^2(T_h)] + \sum_{g \neq h}\sum [\sigma^2(T_g) + \sigma^2(T_h)]$$

$$= 2\sum_{g=1}^{n}\sigma^2(T_g) + \sum_{g \neq h}\sum [\sigma^2(T_g) + \sigma^2(T_h)].$$

Therefore

$$\sum_{g \neq h}\sum [\sigma^2(T_g) + \sigma^2(T_h)] = 2(n-1)\sum_{g=1}^{n}\sigma^2(T_g).$$

Therefore the inequality (2.4) is *equivalent* to the inequality

$$\sum_{g=1}^{n}\sigma^2(T_g) \geq \frac{\sum_{g \neq h}\sum \sigma(T_g, T_h)}{(n-1)}. \tag{2.5}$$

Later we shall show that (2.5) *as an equality* is equivalent to the assumption made by Gulliksen [5] in his derivation of coefficient α as the reliability of the test. Now

$$\sigma^2(T) = \sum_{g=1}^{n}\sigma^2(T_g) + \sum_{g \neq h}\sum \sigma(T_g, T_h)$$

so

$$\sigma^2(T) \geq \frac{\sum_{g \neq h}\sum \sigma(T_g, T_h)}{n-1} + \sum_{g \neq h}\sum \sigma(T_g, T_h)$$

or

$$\sigma^2(\mathrm{T}) \geq \left(\frac{n}{n-1}\right) \sum_{g \neq h} \sum \sigma(\mathrm{T}_g , \mathrm{T}_h). \tag{2.6}$$

Then, using Lemma 1.1, we may write

$$\sigma^2(X) - \sum \sigma^2(Y_g) = \sum_{g \neq h} \sum \sigma(\mathrm{T}_g , \mathrm{T}_h).$$

Now substituting into (2.6) and dividing by $\sigma^2(X)$ we have

$$\rho^2(X, \mathrm{T}) \equiv \frac{\sigma^2(\mathrm{T})}{\sigma^2(X)} \geq \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum \sigma^2(Y_g)}{\sigma^2(X)}\right).$$

We state this formally as

Theorem 2.1: Let $Y_1 , Y_2 , \cdots , Y_n$ be n measurements with true scores $\mathrm{T}_1 , \mathrm{T}_2 , \cdots , \mathrm{T}_n$ and let $X = Y_1 + Y_2 + \cdots + Y_n$ have true score T. Then

$$\rho^2(X, \mathrm{T}) \geq \left(\frac{n}{n-1}\right)\left[1 - \frac{\sum \sigma^2(Y_g)}{\sigma_X^2}\right]. \tag{2.7}$$

The right-hand member of (2.7) has been called *coefficient* α by Cronbach [1]. The left-hand member of (2.7) is the reliability of the composite measurement.

In selecting a derivation of α modeled after that of Guttman we do not suggest that the only important use of this coefficient is as a lower bound on reliability. Except for cases in which tests are relatively homogeneous or long this use is of limited value since in other cases the bound will be a very bad one. Rather we have begun with this derivation of α because we are then able to obtain an understanding of the predominant importance of this function in the study of composite tests by noting the condition under which equality is attained in (2.7). A better lower bound is given by Guttman [6].

3. *Conditions under Which Coefficient Alpha Is Equal to the Reliability of the Test*

The general necessary and sufficient condition for equality in the Cauchy–Schwartz inequality (2.2) is, in our context, the condition that T_g is a linear function of T_h. In the proof of (2.7), the only other inequalities used were $[\sigma(\mathrm{T}_g) - \sigma(\mathrm{T}_h)]^2 \geq 0$ and $|\sigma(\mathrm{T}_g , \mathrm{T}_h)| \geq \sigma(\mathrm{T}_g , \mathrm{T}_h)$ for all (g, h), $g \neq h$. Given that $\mathrm{T}_g = a_{gh} + b_{gh}\mathrm{T}_h$, the first of these becomes an equality if and only if $b_{gh} = +1$ and the second, if and only if $|b_{gh}| = b_{gh}$. Together, these imply $b_{gh} = 1$. Thus we have

Theorem 3.1: The necessary and sufficient condition for (2.7) to hold as an *equality* is that $\mathrm{T}_g \equiv a_{gh} + \mathrm{T}_h$ for all (g, h), i.e., that all components are essentially τ-equivalent.

Clearly then, coefficient α is equal to the reliability of the composite measurement whenever all of the components are τ-equivalent measurements. The condition is not quite necessary since, as we have shown, they may differ, in pairs, by an additive *constant*. We may note that the assumption of τ-equivalence of items implies that the variance-covariance matrix of *true scores* has unit rank (essentially the assumption adopted by Kuder and Richardson [9]).

The converse, however, is not true, for if the variance-covariance matrix of true scores has unit rank, we may conclude only that

$$T_g = a_{gh} + b_{gh}T_h$$

and may not conclude either that T_g and T_h are τ-equivalent or that equality holds in (2.7). For equality in (2.7) we require that $b_{gh} = 1$. But this condition can be violated by a set of true-score random variables having a variance-covariance matrix with unit rank, and hence the unit rank assumption does not imply that coefficient α will be equal to the reliability of the test. Thus the unit rank assumption is a necessary but not a sufficient one for equality to hold in (2.7). The necessary and sufficient condition is just that given in Theorem 3.1, i.e., that the components are essentially τ-equivalent. The reader will no doubt feel that the condition of essential τ-equivalence is much to expect from two composite measurements. But this is the *necessary* (and sufficient) condition for α to be equal to the reliability of the test. Moreover, for our definition of τ-equivalent measurements we require additionally that $a_{gh} = 0$ for all g, h.

4. *Equivalence of the Gulliksen Assumption*

A standard derivation of coefficient α given by Gulliksen [5] as a formula for the reliability of the composite (hence implying equality in (2.7)) proceeds by considering two n-item tests that are "parallel" item for item and introduces the assumption that the average covariance between the matched (parallel) items across the two tests is equal to the average covariance between pairs of items within each form. Gulliksen states that this is "the simplest and most direct assumption" in order to derive coefficient α. We shall show, however, that this assumption is *formally equivalent* to the assumption of Theorem 3.1. The assumption adopted by Gulliksen may be written as

$$\sum_{g=g'} \rho_{gg'}\sigma_g^2 = \frac{\sum\limits_{g \neq h} \rho_{gh}\sigma_g\sigma_h}{n-1}, \tag{4.1}$$

where g and g' are matched items and g and h are items on the same test and where $\rho_{gg'}$, ρ_{gh}, σ_g, and σ_h are correlations and standard deviations of observed scores, there being n covariances between matched items and $n(n-1)$ covariances within each form. But, using Lemmas 1.1 and 1.2, we have

$$\rho_{gg'}\sigma_g^2 = \sigma(\mathrm{T}_g, \mathrm{T}_{g'}) = \sigma^2(\mathrm{T}_g) \quad \text{and} \quad \rho_{gh}\sigma_g\sigma_h = \sigma(Y_g, Y_h) = \sigma(\mathrm{T}_g, \mathrm{T}_h).$$

Hence (4.1) may be written as

$$\sum_{g=1}^{n} \sigma^2(\mathrm{T}_g) = \frac{\sum\sum_{g \neq h} \sigma(\mathrm{T}_g, \mathrm{T}_h)}{n-1}.$$

The equivalence of this assumption with the assumption $\mathrm{T}_g = a_{gh} + \mathrm{T}_h$ is contained in the proof of Theorems 2.1 and 3.1. (See equations (2.4) and (2.5).) Hence we have

Theorem 4.1:

$$\sum_{g=g'} \rho_{gg'}\sigma_g^2 = \frac{\sum\sum_{g \neq h} \rho_{gh}\sigma_g\sigma_h}{n-1}$$

if and only if $\mathrm{T}_g = a_{gh} + \mathrm{T}_h$.

5. *Equivalence of the Jackson–Ferguson Assumption*

The general method of Jackson and Ferguson [8] and the assumption employed by them can be preserved in the following rather direct derivation of coefficient α:

Let $X = \sum_{g=1}^{n} Y_g$ and $X' = \sum_{h'=1}^{n} Y'_{h'}$ be "parallel" composite measurements. It is *not* assumed that the correspondingly numbered components Y_g and $Y_{g'}$ of X and X' are either parallel or τ-equivalent. Let σ_{gh} be the covariance of components Y_g and Y_h of X. Let $\sigma_{gh'}$ be the covariance of the component Y_g of X with the component $Y_{h'}$ of X'. Let $\sigma_{gg} = \sigma_g^2$. The reliability ρ_{XT}^2 of a test may be shown to be equal to the correlation $\rho_{XX'}$ between two parallel forms of the test, i.e.,

$$\rho_{XT}^2 = \rho_{XX'} = \frac{\sigma(X, X')}{\sigma^2(X)} = \frac{\sum_g \sum_{h'} \sigma_{gh'}}{\sum_g \sum_h \sigma_{gh}}. \tag{5.1}$$

Jackson and Ferguson then assume that

$$\frac{\sum_g \sum_{h'} \sigma_{gh'}}{n^2} = \frac{\sum\sum_{g \neq h} \sigma_{gh}}{n(n-1)}, \tag{5.2}$$

i.e., that the average covariance between components on different composites is equal to the average covariance between components on the same composite. Then from (5.1) and (5.2) we have

$$\rho^2_{XT} = \left(\frac{n}{n-1}\right)\left[\frac{\sum_{g \neq h}\sum \sigma_{gh}}{\sum_g \sum_h \sigma_{gh}}\right] \tag{5.3}$$

$$= \left(\frac{n}{n-1}\right)\left[\frac{\sum_g \sum_h \sigma_{gh} - \sum_g \sigma_g^2}{\sum_g \sum_h \sigma_{gh}}\right]$$

$$= \left(\frac{n}{n-1}\right)\left[1 - \frac{\sum_g \sigma_g^2}{\sigma_X^2}\right].$$

A corresponding statement, of course, applies for $\rho^2_{X'T'}$. Hence, the Jackson–Ferguson conditions are sufficient for coefficient α to be equal to the reliability of the test. But if $\alpha = \rho^2_{XT}$, then all of the components of X must satisfy the conditions of Theorem 3.1, hence implying a variance-covariance matrix of component true scores having unit rank. Moreover if X and X' are parallel composite measurements for which α is equal to the reliability of the measurement, then (5.2) must hold. This may easily be demonstrated by equating (5.1) to (5.3). Hence if X and X' are parallel composite measurements, then a necessary and sufficient condition that α be equal to the reliability of either measurement is just condition (5.2), and the condition (5.2) is equivalent to the condition of Theorem 3.1.

Since both the Gulliksen and Jackson–Ferguson assumptions are equivalent to the condition given in Theorem 3.1, they must be equivalent to each other. The nature of this latter equivalence is well worth noting explicitly. It is made precise in the following two statements:

(i) If we have a composite measurement which satisfies the Gulliksen assumption with respect to *any* second measurement having components which are pairwise parallel to the first measurement, then the Jackson–Ferguson assumption is satisfied for the first measurement and *any* third composite measurement which is parallel to it (even though the components of these two composites are not pairwise parallel).

(ii) If we have a composite measurement which satisfies the Jackson–Ferguson assumption with respect to *any* parallel composite measurement (it not being assumed that the components are pairwise parallel), then the Gulliksen assumption is satisfied for the first measurement and *any* third composite measurement having components which are pairwise parallel to the first measurement.

Jackson–Ferguson and Gulliksen originally adopted their respective assumptions with little intuitive justification. On the other hand, the assumption of essential τ-equivalence is quite understandable though disconcertingly restrictive. That all three sets of assumptions are equivalent is illuminating. However, we must stress that this equivalence was obtained by strengthening the definition of the term parallel measurement used in both the Gulliksen, and Jackson–Ferguson derivations.

6. *Coefficient α as the "Mean of the Split-Half Reliabilities"*

Coefficient α has a further property which was established by Cronbach [1]. Suppose we have a test of $2n$ items. The test may be divided into halves in $\frac{1}{2}(2n!)/(n!)^2$ ways. For each of these split halves a coefficient alpha (for $n = 2$ in the right-hand member of (2.7)) may be defined and denoted α_2. Also the coefficient alpha for the $2n$-item test may be defined and denoted α_{2n}. These quantities may be related by

Theorem 6.1:

$$\mathcal{E}\alpha_2 = \alpha_{2n}, \tag{6.1}$$

i.e., the expected value of α_2 is equal to α_{2n}, where this expectation can be interpreted in two ways. First it may be interpreted as stating that if we compute all $\frac{1}{2}(2n!)/(n!)^2$ possible values of α_2 and take their average, this will be equal to α_{2n}. Also it states if items are assigned to the two half-tests randomly, then the expected value of α_2 over the $\frac{1}{2}(2n!)/(n!)^2$ possible splits is α_{2n}.

Proof:

$$\left(\frac{n}{n-1}\right)\left[1 - \frac{\sum \sigma^2(Y_g)}{\sigma_X^2}\right] = \left(\frac{n}{n-1}\right)\left[\frac{\sum_{g\neq h} \sigma(Y_g, Y_h)}{\sigma^2(X)}\right] \tag{6.2}$$
$$= \frac{1}{n(n-1)} \sum_{g\neq h} \sigma(Y_g, Y_h) \Big/ \frac{1}{n^2}\sigma^2(X).$$

Let Z_1 and Z_2 be arbitrary split halves where, with no loss of generality, we take

$$Z_1 = \sum_{g=1}^{n} Y_g \qquad Z_2 = \sum_{g=n+1}^{2n} Y_g .$$

For this split the coefficient α, which we shall denote by α_2, is

$$\begin{aligned}\alpha_2 &= 4\sigma(Z_1, Z_2)/\sigma_X^2\\ &= 4 \sum_{g=1}^{n} \sum_{h=n+1}^{2n} \sigma(Y_g, Y_h)/\sigma_X^2 .\end{aligned}$$

Taking the expected value over all splits (each of which because of random sampling is assumed to be equally likely), we have

$$\mathcal{E}\alpha_2 = (4/\sigma_X^2) \sum_{g=1}^{n} \sum_{h=n+1}^{2n} \mathcal{E}\sigma(Y_g, Y_h).$$

For an arbitrary split, all distinct pairs of components have an equal chance of being assigned to the labels Y_g and Y_h for $g = 1, 2, 3, \cdots, n$ and $h = n + 1, \cdots, 2n$. Hence all terms in the double sum are equal, and moreover, equal to the average covariance between distinct components of the test. Since there are n^2 terms in the sum, we may then write $\mathcal{E}\alpha_2$ as

$$\mathcal{E}\alpha_2 = \frac{4n^2}{\sigma^2(X)} \sum_{g \neq h}\sum \frac{\sigma(Y_g , Y_h)}{2n(2n - 1)}$$

where we have singled out a particular labelling of the components and hence let each of the subscripts g and h range from 1 to $2n$, for $g \neq h$. But from (6.2) the general formula for α_{2n}, coefficient α for a test of $2n$ items, may be written as

$$\alpha_{2n} = \frac{1}{2n(2n - 1)} \sum_{g \neq h}\sum \sigma(Y_g , Y_h) \Big/ \frac{1}{4n^2}\, \sigma_X^2 .$$

Hence

$$\mathcal{E}\alpha_2 = \alpha_{2n} .$$

Also we have

Corollary 6.2: If the $2n$ items of the test satisfy the condition of Theorem 3.1, then

$$\alpha_2 \equiv \alpha_{2n} .$$

This follows from Theorem 3.1 on noting that each of the possible values of α_2 is equal to the reliability of the composite test, since all possible half-tests are essentially τ-equivalent.

Our derivation of Theorem 6.1 differs little from that of Cronbach [1] who correctly made explicit in his introductory discussion and in the body of his derivation his verbal interpretation of (6.1). In reading the summary to Cronbach's paper [1] the reader should bear in mind that when Cronbach states that α is equal to the mean of the split-half reliabilities, he refers to the Rulon stepped-up split-half reliability which is in fact simply α_2, coefficient alpha for the test divided into two components and *not* to the Spearman–Brown formula for the reliability of a test of double length.

7. *The Use of Coefficient Alpha as a Measure of Internal Consistency*

The primary use of coefficient α has been not as a lower bound on the reliability coefficient, but rather as a measure of internal consistency of a test. Cronbach [1], however, has pointed out that α increases with n. Indeed typically the limiting value of α is unity as n approaches infinity. Thus Cronbach recommends the use, as a measure of internal consistency, of a function of α which is independent of test length.

However, the use of any function of α as a measure of internal consistency requires one caution. The condition for equality in (2.7) is that $T_g = a_{gh} + b_{gh}T_h$ for all (g, h) and that all $b_{gh} = 1$. Thus departures from the condition $b_{gh} = 1$ for any (g, h) will lower the value of α, i.e., α is sensitive to departures from $b_{gh} = 1$. Hence functions of α will be valid measures of internal consistency only in the sense of measuring the degree to which $T_g = a_{gh} + T_h$ for all (g, h) but not the degree to which $T_g = a_{gh} + b_{gh}T_h$. Also note that α is not sensitive to departures from $a_{gh} = 0$ for any (g, h).

8. *Relation to Previous Work*

This paper has been designed to shed some light on some long-standing problems about which conflicting statements have been made. However, it is not possible for us generally to point to specific statements of other writers and judge them either correct or incorrect. Indeed even in presenting derivations which we credit to others we usually begin with the phrase "following an approach due to . . ." or some equivalent obscuring device.

We have adopted this position simply because it is not possible to make valid direct comparisons with previous work, for the reason that the definition of parallel measurement that we have adopted here is not that which has been in general usage. However, with the present definition, which we believe satisfies the general notion that parallel measurements measure the same thing and equally well, and the closely related definitions of τ-equivalence and essential τ-equivalence, we find that long-standing problems disappear.

9. *Summary*

Adopting a new definition of parallelism due to Novick [12] and introducing corresponding definitions of τ-equivalence and essential τ-equivalence we have:

(1) following Guttman, rederived coefficient α as a lower bound on the reliability of a test;

(2) shown that the necessary and sufficient condition that this bound is attained, i.e., that α is equal to the reliability of the test, is that the true-score random variables T_g and T_h of any pair of components are related by $T_g = a_{gh} + T_h$ where a_{gh} is a constant, i.e., that they are essentially τ-equivalent;

(3) indicated that this condition implies that the variance-covariance matrix of true scores has unit rank;

(4) shown that adopting the new definition of parallelism the assumption adopted by Gulliksen in his derivation of coefficient α is equivalent to the assumption of essential τ-equivalence of the components;

(5) presented a derivation of coefficient α based on the assumption employed by Jackson and Ferguson [8] and shown that adopting the new definition of parallelism this assumption is equivalent to the assumption of essential τ-equivalence;
(6) concluded that adopting the new definition of parallelism the assumption of Gulliksen and that of Jackson and Ferguson are equivalent;
(7) pointed out that coefficient α, (α_{2n}), for a test having $2n$ items is equal to the average value of the coefficients α_2 (not the average value of the Spearman–Brown stepped-up split-half reliabilities), for all possible combinations of items into two half-tests;
(8) pointed out the limitation in interpreting any function of α as a measure of internal consistency that α is sensitive to changes in the scale of the components but not to changes in location.

REFERENCES

[1] Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, **16**, 297–334.
[2] Cronbach, L. J., Rajaratnam, Nageswari, & Gleser, Goldine C. Theory of generalizability: A liberalization of reliability theory. *Brit. J. Statist. Psychol.*, 1963, **16**, 137–163.
[3] Cureton, E. E. The definition and estimation of test reliability. *Educ. Psychol. Measmt.*, 1958, **18**, 715–738.
[4] Dressel, P. L. Some remarks on the Kuder-Richardson reliability coefficient. *Psychometrika*, 1940, **5**, 305–310.
[5] Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.
[6] Guttman, L. Reliability formulas that do not assume experimental independence. *Psychometrika*, 1953, **18**, 225–239.
[7] Hoyt, C. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **6**, 153–160.
[8] Jackson, R. W. B., & Ferguson, G. A. Studies on the reliability of tests. Bulletin 12, Department of Educational Research, University of Toronto, 1941.
[9] Kuder, G. F., & Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, **2**, 151–160.
[10] Lord, F. M. Statistical inferences about true scores. *Psychometrika*, 1959, **24**, 1–17.
[11] Lyerly, S. B. The Kuder-Richardson formula (21) as a split-half coefficient and some remarks on its basic assumption. *Psychometrika*, 1958, **23**, 267–270.
[12] Novick, M. R. The axioms and principal results of classical test theory. *J. Math. Psychol.*, 1966, **3**, 1–18.
[13] Rulon, P. J. A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educ. Rev.*, 1939, **9**, 99–103.

Manuscript Received 2/21/66
Revised Manuscript Received 5/9/66