ACT Research Report Series 93-4

Linear Models for Item Scores: Reliability, Covariance Structure, and Psychometric Inference

David Woodruff

August 1993

ACT

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243

Linear Models for Item Scores: Reliability, Covariance Structure, and Psychometric Inference

David Woodruff
American College Testing

Running Head: LINEAR MODELS

Linear Models for Item Scores:
Reliability, Covariance Structure,
and Psychometric Inference

Abstract

Two ANOVA models for item scores are compared. The first is an items by subject random effects ANOVA. The second is a mixed effects ANOVA with items fixed and subjects random. Comparisons regarding reliability, Cronbach's α coefficient, psychometric inference, and inter-item covariance structure are made between the models. When considering the inter-item covariance structures for the two ANOVA models, brief comparisons with factor analysis models are also made. It is concluded that inference from a sample of items to a population of items requires homogeneous inter-item covariances, that reliability has different meanings under the two models, and that while coefficient α is a lower bound for reliability under the second model, it is not under the first.

Key Words: Coefficient Alpha, Covariance Structure, Generalizability, Linear Models, Psychometric Inference, Reliability

Introduction

This paper compares two different ANOVA models for items. The first model is the two-way items by examinees random effects (Model II) ANOVA. The second model is the two-way items by examinees mixed effects (Model III) ANOVA. Very careful and complete statistical derivations of these models are given by Scheffe´ (1956a, 1956b, and 1959). This paper draws heavily from Scheffe´'s work. The two ANOVA models are compared to each other in detail and briefly to factor analysis models. Factor analysis models are extensively discussed by Harmon (1976) and Mulaik (1972). As considered here, the factor analysis model is statistically more similar to the mixed ANOVA model than to the random ANOVA model. Under the factor analysis model, items are considered fixed and non-random, while subjects are randomly sampled from a population of subjects. See Mulaik and McDonald (1978), Williams (1978), and McDonald and Mulaik (1979) for an alternative formulation of the factor analysis model.

All of the models under consideration are linear models. A model is defined as linear if an examinee's expected score on an item is a linear function of item characteristics. Item characteristics may be fixed parameters as in the mixed ANOVA model or random variables as in the random ANOVA model. The factor analysis model is here considered to be linear in its item parameters which are usually called factor loadings even though these linear coefficients are applied to factor scores, which are unobserved random variables associated with examinees. An example of a nonlinear model is the logistic ogive item characteristic curve model (Lord and Novick, 1968). From a theoretical viewpoint, linear models usually do not accurately describe dichotomously scored items, and most items are so scored. However, for carefully constructed tests, linear models for item scores are often sufficiently accurate to provide useful approximations. [See Feldt (1965), Hsu and Feldt (1969), Hakstian and Whalen (1976), Seeger and Gabrielsson (1968), Gabrielsson and Seeger (1976), McDonald and Ahlawat (1974), McDonald (1981, 1985), and Collins, Cliff, McCormick, and Zatkin (1986).]

The discussion of the models presented here will focus on three characteristics useful in psychometrics. The first is reliability. Under the three models reliability is defined as the squared correlation between an observed and a true score. A few relevant references regarding reliability are Gutman (1945), Novick and Lewis (1967) Bentler (1972), Jackson and Agunwamba (1977), and Bentler and Woodward (1980, 1983). Parametric expressions for reliability and Cronbach's (1951) coefficient alpha are given, and the sampling distribution for the sample alpha coefficient is discussed. The second characteristic is the inter-item covariance matrix. For each model, the assumed or resulting covariance structure is discussed and compared with factor analysis models. Finally, psychometric inference is discussed. Psychometric inference is considered as statistical inference to a population of items from a sample of items randomly drawn from the population. The more general term generalizability is not used since it connotes statistical inference for a wide array of facets, not just items. There is a large body of literature on psychometric inference. A few references are Hotelling (1933), Tryon (1957), Lord and Novick (1968), Cronbach, Gleser, Nanda, and Rajaratnam (1972), Mulaik (1972), Kaiser and Michael (1975), Rozeboom (1978), McDonald (1978), and Brennan (1983). Both the approach and results presented here, while most similar to, differ in part from those developed by Lord and Novick (1968) and Cronbach et al. (1972).

Brief descriptions of seven conclusions original to this paper are:

1. Conditional variances for interaction effects may be heterogeneous in the random ANOVA model.

2. The random ANOVA model requires the inter-item covariance matrix to have homogeneous off-diagonal elements, while the mixed ANOVA model places no restrictions on the inter-item covariance matrix except positive semi-definiteness. Hence, any factor analysis model may be subsumed under the mixed ANOVA model but not the random ANOVA model.

3. Interaction effects in the random ANOVA model are analogous to specific factors in a certain single common factor factor analysis model, while the examinee main effect is analogous to the single common factor.

4. The squared correlation between observed scores and true scores is a useful definition of reliability under the random ANOVA model as well as under the mixed ANOVA model, but the definition of true score differs under the two models.

5. Reliability as defined in 4. has different meanings under the two models. In the mixed ANOVA model, interaction (specific) variance is included in true score variance, while in the random ANOVA model it is not.

6. The parametric value of Cronbach's alpha coefficient is a lower bound to the parametric value of reliability (as defined in 4) under the mixed ANOVA model but not under the random ANOVA model.

7. Given certain normality assumptions, a transformation of the sample alpha coefficient has an F distribution under the random ANOVA model. For the mixed ANOVA model, the F distribution only holds if in addition to certain normality assumptions there are either no interactions or the inter-item covariance matrix has special restricted forms.

The practical implications of these conclusions for the analysis of test data will be discussed in the last section of this paper.

The Items by Examinees Random ANOVA Model

The model presented here is essentially the same model developed by Scheffe´ (1959, chap. 7). It assumes that a random sample of n items chosen from a countably infinite population of items is administered to a random sample of N examinees chosen from a countably infinite population of examinees. The sampling of items and examinees is assumed to be completely independent. Let x_{ij} represent subject j's observed score on item i. A preliminary form of the model is

$$x_{ij} = t_{ij} + e_{ij} \quad i = 1,\ldots,n \quad j = 1,\ldots,N \, . \tag{1}$$

The quantities t_{ij} and e_{ij} are, respectively, the true score and the error score of examinee j on item i. Different definitions for true and error scores under the random ANOVA model will be admitted later. Within the present context, true and error scores are not absolutes; their definitions may vary depending on the inferences being made. The various true and error scores considered in this paper are not necessarily an exhaustive set of possible true and error scores under the models presented.

If examinee j responds independently and repeatedly to item i, these replications are indexed by the subscript k. For cognitive tests such random replications are rarely available, though they occasionally may be obtained for affective scales. The present development assumes that such replications are not available from the data. In the theoretical development of the model, these replications are allowed to be present. In particular, the model assumes that for the sequences of independent random variables $e_{ij1}, e_{ij2}, \ldots, e_{ijk}, \ldots$: $E(e_{ijk}) = 0$ for all i, j, and k, and that $Var(e_{ijk}) = E(e^2_{ijk}) = \sigma^2(e_{ij})$, i.e., that the error variances are heterogeneous over the domains of i and j. For notational simplicity, the subscript k will usually be suppressed, since for the remainder of the paper it will usually take the value of one.

The above imply that $E_i(e_{ij}) = 0$ and that $E_j(e_{ij}) = 0$, where notation such as E_i and Var_i means that the expectation and variance are taken over the population whose members are indexed by the subscript i. When no subscript is present the expectation is over random replications. The above also imply that the true and error scores are uncorrelated, i.e., $Cov_i(t_{ij},e_{ij'}) = Cov_j(t_{ij},e_{i'j}) = 0$ for all j,j′ and i,i′ , respectively. It is further assumed that all errors are independent within and across all populations.

Scheffe´ (1959, chap. 10) shows that the expressions for expected mean squares, to be presented later, are valid under the heterogeneity of error variances indicated above. He also shows that the F distribution theory invoked later is exactly valid only when the error variances are homogeneous, but holds approximately when the error variances are mildly heterogeneous if the design is balanced. This paper assumes that the error variances are only mildly heterogeneous and that each examinee responds to each item once and only once. Hence, the design is balanced and the F distribution theory will be assumed to hold when the appropriate normality assumptions, discussed later, are invoked.

The following quantities will be used in later developments:

$$E_j(e_{ij}^2) = E_jE(e_{ij}^2) = E_j(\sigma^2(e_{ij})) = \sigma^2(e_i) \ ,$$

$$E_i(e_{ij}^2) = E_iE(e_{ij}^2) = E_i(\sigma^2(e_{ij})) = \sigma^2(e_j) \ , \text{ and}$$

$$E_iE_jE(e_{ij}^2) = E_i(\sigma^2(e_j)) = E_j(\sigma^2(e_i)) = \sigma^2(e) \ .$$

The model is further specified by writing

$$t_{ij} = \mu + a_i + b_j + c_{ij} \qquad (2)$$

where $\mu = E_iE_j(t_{ij})$, $a_i = E_j(t_{ij}) - \mu$, $b_j = E_i(t_{ij}) - \mu$, and $c_{ij} = t_{ij} - E_i(t_{ij}) - E_j(t_{ij}) + \mu$. The overall mean is denoted by μ , while a_i and b_j denote the main effects due to item i and examinee j, respectively. The interaction effect due to item i and examinee j is denoted by c_{ij}. These definitions implicitly assume that all items are similarly scored and hence on the same scale. Scheffe´ (1959) shows that the above definitions imply that the model components: a_i, b_j, and c_{ij} have unconditional and for the c_{ij} also conditional expectations of zero.

For what follows, it is important to note that the subscripts i and j do double duty; they are both subscript indices and random variables. Furthermore, the a_i, b_j, and c_{ij} are functions of the random variables i and j. Scheffe´ introduces additional notation to avoid these double meanings for the subscripts, but the present paper sacrifices Scheffe´'s conceptual clarity for notational economy.

Scheffe´ (1959, pp 240-241) shows that certain marginal covariances among the model components are zero. His derivations are presented here in detail because of their importance. Scheffe´ shows that

$$\sigma(a_i, c_{ij}) = E_i E_j (a_i * c_{ij})$$

$$= E_i [a_i * E_j (c_{ij}) | i]$$

$$= E_i (a_i * c_{i\cdot}) = 0 \quad \text{because} \quad c_{i\cdot} = 0 \quad \text{for all } i,$$

$$\sigma(b_j, c_{ij}) = E_j [b_j * E_i (c_{ij}) | j]$$

$$= E_j (b_j * c_{\cdot j}) = 0 \quad \text{because} \quad c_{\cdot j} = 0 \quad \text{for all } j,$$

$$\sigma(c_{ij}, c_{i'j}) = E_i E_{i'} E_j (c_{ij} * c_{i'j}) \qquad i \neq i'$$

$$= E_j [E_i E_{i'} (c_{ij} * c_{i'j}) | j]$$

$$= E_j [E_i (c_{ij} | j) * E_{i'} (c_{i'j} | j)]$$

$$= E_j(c_{\cdot j}*c_{\cdot j}) = 0 \quad \text{because } c_{\cdot j} = 0 \text{ for all } j,$$

and

$$\sigma(c_{ij},C_{ij'}) = E_i[E_j(c_{ij}|i)*E_{j'}(c_{ij'}|i)] \qquad j \neq j'$$

$$= E_i(c_{i\cdot}*c_{i\cdot}) = 0 \quad \text{because } c_{i\cdot} = 0 \text{ for all } i.$$

In the above, the notation $E_iE_{i'}$ refers to the expectation over the bivariate distribution obtained from sampling pairs of items from the population of items where the members of each pair are distinct.

Scheffe´ (1959) does not discuss the following model component conditional covariances:

$$\sigma(a_i,c_{ij}|j) = E_i(a_i*c_{ij}|j),$$

$$\sigma(b_j,c_{ij}|i) = E_j(b_j*c_{ij}|i),$$

$$\sigma(c_{ij},c_{i'j}|i,i') = E_j(c_{ij}*c_{i'j}|i,i'), \quad \text{and}$$

$$\sigma(c_{ij},c_{ij'}|j,j') = E_i(c_{ij}*c_{ij'}|j,j') .$$

These conditional covariances are of considerable concern because as will be seen later their values determine the inter-item covariance matrix.

Though a formal proof will not be given, it is asserted here that the above conditional covariances are also zero under Scheffe´'s (1959) model. Four considerations lead to this conclusion. First it does not appear possible to generate model component data such that Scheffe´'s marginal covariances are zero but the above conditional covariances are not.

Second, Scheffe´'s proof that the above marginal covariances are zero depends on the order in which the conditional covariances are taken. If the order is switched the same result must be found. This implies that the above conditional covariances must have expected values of zero, and this can occur only if all are zero or some are positive and some negative such that their average is zero. Because, as will be shown, these conditional covariances determine the inter-item covariances, and tests are usually constructed of items that all intercorrelate positively, it appears more reasonable in a testing context to assume that the conditional covariances are zero rather than some positive and some negative. Third, Scheffe´ (1959, pp 242-243) considers the two-way random model interaction components as analogous to the error terms in a two-way fixed effects model and these later have all conditional covariances as zero. Fourth, Cornfield and Tukey (1956) consider several covariances in the derivation of expected mean squares for factorial designs, but in the two-way random model these covariances are all zero.

Scheffe´ (1959) defines the variance components of the model as: $\sigma^2(a) = E_i(a_i^2)$, $\sigma^2(b) = E_j(b_j^2)$, and $\sigma^2(c) = E_iE_j(c_{ij}^2)$. In defining $\sigma^2(c)$, Scheffe´ does not consider the interaction conditional variances $\sigma^2(c_i) = E_j(c_{ij}^2)$ and $\sigma^2(c_j) = E_i(c_{ij}^2)$. Though i and j are assumed to be statistically independent variables, c_{ij} is a function of both these variables and for this reason the conditional interaction variances need not be homogeneous. If it is assumed that the model components have a multivariate normal distribution as Scheffe´ sometimes does, then the model components are mutually statistically independent and this forces the interaction conditional variances to be homogeneous. Here they will be considered heterogenous unless otherwise specified. Scheffe´'s (1959, chap. 10) demonstration that his formulas for expected mean squares are valid under heterogeneity of error variances implies the same under heterogeneity of interaction conditional variances.

Of particular interest in the random model ANOVA are the mean squares for examinees and the mean squares for items by examinees which are denoted MS_b and MS_c, respectively. Scheffe´ (1959) derives the following expressions for the expected value of these mean squares: $E_{nN}(MS_b) = n\sigma^2(b) + \sigma^2(c) + \sigma^2(e)$ and $E_{nN}(MS_c) = \sigma^2(c) + \sigma^2(e)$, where E_{nN} denotes that these expectations are the means of an infinite number of bivariate random samples consisting of n items and N subjects.

These mean squares are of interest because Hoyt (1941) has shown that the sample value of Cronbach's (1951) coefficient α, denoted $\hat{\alpha}$ herein, is given by $\hat{\alpha} = [(MS_b - MS_c)/MS_b] = 1 - (MS_c/MS_b)$. The parametric counterpart of $\hat{\alpha}$ depends upon the statistical model used to describe the data. For the random ANOVA model this parameter is denoted α_{RA} , the subscript RA denoting that this definition is specific to the random model ANOVA. The parameter α_{RA} is defined by

$$\alpha_{RA} = \frac{E_{nN}(MS_b) - E_{nN}(MS_c)}{E_{nN}(MS_b)} = \frac{\sigma^2(b)}{\sigma^2(b) + \sigma^2(c)/n + \sigma^2(e)/n} . \qquad (3)$$

The rationale for this definition is that $\hat{\alpha}$ converges in probability to α_{RA} under the RA model. This is discussed further below. Since α_{RA} is defined in terms of $E_{nN}(MS_b)$ and $E_{nN}(MS_c)$ whose definitions in turn depend upon the RA model, the definition of α_{RA} is tied to the RA model and hence the RA subscript. Feldt (1965) has shown that under the additional assumptions of independent normal distributions for the $\{a_i\}$, $\{b_j\}$, $\{c_{ij}\}$, and $\{e_{ij}\}$, $(1 - \alpha_{RA})/(1 - \hat{\alpha})$ is distributed as $F[N-1, (n-1)(N-1)]$. Under these assumptions, the conditional variances for both the interactions and errors are considered homogeneous, but slight heterogeneity should produce at most only mild departures from the F distribution. Using the expression for the mean of an F distribution it follows that $E_{nN}(\hat{\alpha}) = [(N - 1)/N - 3)]\alpha_{RA} - [2/(N - 3)]$. This shows that $\hat{\alpha}$ is an asymptotically (as $N \to \infty$) unbiased estimator of α_{RA} . Even without the normality assumptions, $\hat{\alpha}$ is still a consistent estimator for α_{RA} since it is a method of moments estimator for α_{RA} (Serfling, 1983), and equivalently converges in probability to α_{RA} .

The random ANOVA (RA) model has been presented in some detail. It is now of interest to compare that model to the factor analysis (FA) model. This comparison may be made by examining the conditional covariance matrix for the n sampled items, the conditioning being on the n items selected from the infinite population of items. Let the observed scores on the n items be represented by the column vector $\underset{\sim}{x}$. The conditional covariance matrix is $\underset{\sim}{\Sigma}_{x|n} = E_j[(\underset{\sim}{x}_j - E_j(\underset{\sim}{x}_j))'(\underset{\sim}{x}_j - E_j(\underset{\sim}{x}_j))]$. The diagonal elements of this matrix are $Var_j(x_{ij}) = \sigma^2(b) + \sigma^2(c_i) + \sigma^2(e_i)$. Because it is assumed here that under the RA model $cov_j(c_{ij},c_{i'j}) = 0$ for any pair of items randomly selected from the population of items, it follows that this covariance will be zero for all pairs of items in the randomly selected sample of n items, and consequently that the off-diagonal elements of this matrix are $Cov_j(x_{ij},x_{i'j}) = \sigma^2(b)$. The rather simple form of this conditional covariance matrix may be represented as $\underset{\sim}{\Sigma}_{x|n} = \sigma^2(b)\underset{\sim}{J} + \underset{\sim}{\Delta}[\sigma^2(c_i) + \sigma^2(e_i)]$ where $\underset{\sim}{J}$ represents a matrix of all ones and $\underset{\sim}{\Delta}$ is a diagonal matrix with the indicated elements. It follows that the conditional covariance matrix for the true scores on the n items is

$$\underset{\sim}{\Sigma}_{t|n} = \sigma^2(b)\underset{\sim}{J} + \underset{\sim}{\Delta}(\sigma^2(c_i)) \ . \qquad (4)$$

Hocking (1985) presents covariance structures for a wide variety of random and mixed ANOVA models. He assumes homogeneity among the error and conditional interaction variances. Given his assumptions, his results agree with those presented here.

The RA conditional covariance structure is identical to the covariance structure of a one common factor FA model with homogeneous factor loadings and n specific factors distinct from the errors. This is Spearman's (1904) model but with the additional restriction that the items all correlate equally with the general factor. More specifically, the subject main effect variance in the RA model is analogous to the common factor variance in the FA model while the conditional interaction variances in the RA model are analogous to specific variances in the FA model. Another way to characterize this conditional covariance structure is as an essentially tau equivalent model (Lord and Novick, 1968) but with the addition of n specific factors with possibly heterogeneous variances.

If the specific factors have homogeneous variances, then the conditional covariance structure for the true scores is equivalent to the equicorrelation model (Morrison, 1976). Under the equicorrelation model, the first and largest eigenvalue of $\underline{\Sigma}_{t|n}$, denoted λ_1, is equal to $n\sigma^2(b) + \sigma^2(c)$. The second distinct eigenvalue of $\underline{\Sigma}_{t|n}$ has multiplicity n-1 and is given by $\sigma^2(c)$. It is denoted λ_2 .

The simple form of the conditional covariance matrix in the RA model results from the uncorrelatedness of the model components. Though this covariance structure is a rather restricted special case of the many more versatile covariance structures permitted by FA models, the RA model permits explicit statistical inference to a population of items. The price for this gain in "generalizability" is the assumption of a simple covariance structure among the items.

The inferential differences between considering items random and considering items fixed may be illustrated by how reliability may be defined under these conditions. For subject j, let the item domain true score be defined as $\tau_j = E_i(x_{ij}) = \mu + b_j$. This implies that the item domain error score for subject j is $\varepsilon_j = \bar{x}_{\cdot j} - \tau_j = \bar{a}_{\cdot} + \bar{c}_{\cdot j} + \bar{e}_{\cdot j}$. Note that for random replications $E(\varepsilon_j) = \bar{a}_{\cdot} + \bar{c}_{\cdot j}$, and that for examinees $E_j(\varepsilon_j) = \bar{a}_{\cdot}$. Furthermore, considering just a one-item test, $\mathrm{Cov}_i(\varepsilon_{ij}, \varepsilon_{ij'}) = \sigma^2(a)$ for all $j \neq j'$. These conditions violate the usual assumptions of classical test theory (Lord and Novick, 1968, chap. 3), because here the errors do not have means of zero and the errors are inter-correlated. However, $\mathrm{Cov}_j(\tau_j, \varepsilon_j) = 0$ and this crucial result implies that if interest focuses on the reliability of a specific test composed of n randomly selected items with respect to the item domain true scores, then a useful definition of reliability is $\mathrm{Rel}(\bar{x}_{\cdot j}, \tau_j) = [\mathrm{Cor}_j(\bar{x}_{\cdot j}, \tau_j)]^2$. Reliability so defined measures the accuracy with which relationships between observed test scores are indicative of relationships between item domain true scores. Since $\mathrm{Cov}_j(\bar{x}_{\cdot j}, \tau_j) = \sigma^2(b)$,

$$\mathrm{Var}_j(\bar{x}_{\cdot j}) = \sigma^2(b) + (1/n^2)\sum_i^n \sigma^2(c_i) + (1/n^2)\sum_i^n \sigma^2(e_i) ,$$

and $\mathrm{Var}_j(\tau_j) = \sigma^2(b)$, it follows that

$$\mathrm{Rel}(\bar{x}_{\cdot j}, \tau_j) = \frac{\sigma^2(b)}{\sigma^2(b) + (1/n^2)\sum_i^n \sigma^2(c_i) + (1/n^2)\sum_i^n \sigma^2(e_i)} \qquad (5)$$

$$= \mathrm{Var}_j(\tau_j)/\mathrm{Var}_j(\bar{x}_{\cdot j}),$$

which is the usual ratio of true score variance to observed score variance. If the error variances and the conditional interaction variances are homogeneous then $\alpha_{RA} = \mathrm{Rel}(\bar{x}_{\cdot j}, \tau_j)$, otherwise α_{RA} is only an approximation to this reliability, albeit not a bad one.

An alternative definition of reliability under the RA model which is more appropriate when concern is not with the reliability of a particular randomly constructed test but rather with the population of such tests is $E_n[\mathrm{Rel}(\bar{x}_{\cdot j}, \tau_j)]$. Here, E_n denotes that the expectation is over the population of randomly constructed tests consisting of n items. This definition of reliability is appropriate when the same test will be administered to every examinee, but concern is with the reliability of any randomly constructed test rather than a particular test that is randomly selected. The situation in which different examinees take different randomly constructed test forms is not often encountered in practice and is not addressed in this paper (but see Lord and Novick, 1968, p. 208). If the error variances and the conditional interaction variances are homogeneous, then $E_n[\mathrm{Rel}(\bar{x}_{\cdot j}, \tau_j)] = \alpha_{RA}$. This follows since $\mathrm{Rel}(\bar{x}_{\cdot j}, \tau_j) = \alpha_{RA}$ for each and every randomly constructed test consisting of n items. If homogeneity does not hold, an exact expression for $E_n(\mathrm{Rel}(\bar{x}_{\cdot j}, \tau_j))$ requires additional model specifications which will not be attempted in this paper. However, it may be shown by using the delta method of Kendall and Stuart (1977, Vol. I) that α_{RA} is a first order approximation for $E_n[\mathrm{Rel}(\bar{x}_{\cdot j}, \tau_j)]$ under heterogeneity.

If the data are accurately described by the RA model, but the usual definition of reliability (Lord and Novick, 1968, chap. 3) is adopted, then

$$\mathrm{Rel}(\bar{x}_{\cdot j},\bar{t}_{\cdot j}) = [\mathrm{Cor}_j(\bar{x}_{\cdot j},\bar{t}_{\cdot j})]^2 = \mathrm{Var}_j(\bar{t}_{\cdot j})/\mathrm{Var}_j(\bar{x}_{\cdot j}) \qquad (6)$$

$$= \frac{\sigma^2(b) + (1/n^2)\sum_i^n \sigma^2(c_i)}{\sigma^2(b) + (1/n^2)\sum_i^n \sigma^2(c_i) + (1/n^2)\sum_i^n \sigma^2(e_i)}$$

Usually, $\mathrm{Rel}(\bar{x}_{\cdot j},\bar{t}_{\cdot j}) > \alpha_{RA}$. However, if there is no item by examinee interaction and the error variances are homogeneous then $\mathrm{Rel}(\bar{x}_{\cdot j},\bar{t}_{\cdot j}) = \alpha_{RA}$.

A comparison of (6) to (5) shows that the interaction (specific) variances are included in the numerator of $\mathrm{Rel}(\bar{x}_{\cdot j},\bar{t}_{\cdot j})$ but excluded from the numerator of $\mathrm{Rel}(\bar{x}_{\cdot j},\tau_j)$. This difference is due to the difference in definitions between $\bar{t}_{\cdot j}$ and τ_j . If the true score is specific to the test, i.e., $\bar{t}_{\cdot j}$, then the interaction (specific) variances are included in the true score variance. When the true score is defined over the population of items, i.e., τ_j , then the interaction (specific) variances do not contribute to the true score variance.

Two brief observations regarding the RA model are of interest. If no interactions are present the RA model may be viewed as a linear analog of the one parameter Rasch model (Lord and Novick, 1968, p 402) with explicit item and examinee sampling. Second, the symmetry of the RA model allows consideration of not only the inter-item covariance matrix but also the similarly constrained inter-examinee covariance matrix.

This section of the paper has presented a detailed development of the RA model and a brief comparison of the RA model to the FA model. The development demonstrates that under the RA model generalization in a statistical manner over a population of items requires a simple and specialized covariance structure among the items. In the next section, the mixed ANOVA (MA) model is considered.

The Items by Examinees Mixed ANOVA Model

Hocking (1973) compares three different versions of the two-way mixed ANOVA (MA) model that have been presented in the statistical literature, and resolves the differences between their associated expressions for expected mean squares. This paper adopts the most general one of these three which is due to Scheffe´ (1959). In the mixed ANOVA model, the N examinees are randomly sampled from an infinite population of examinees, but the n items are considered fixed and non-random. Even though the items may be randomly chosen from a population of items, this fact is ignored; the MA model simply is not concerned with statistical inferences to a population of items. All statistical inferences are conditional on the n items selected, since the population of items is not defined in the MA model.

The model may be written as

$$x_{ij} = t_{ij} + e_{ij} \quad i=1,\ldots,n \quad j=1,\ldots,N$$

where $t_{ij} = \mu + \alpha_i + b_j + c_{ij}$. The model assumes that the error scores have zero means for all i and j and this implies that the true and error scores are uncorrelated. The non-random parameters μ and α_i represent the overall mean and the main effect of item i, respectively. The random variable b_j represents the main effect due to examinee j, while the random variable c_{ij} represents the effect due to the interaction of examinee j with item i. These model components are defined as $\mu = E_j[(1/n)\sum_i^n t_{ij}] = E_j(\bar{t}_{.j})$, $\alpha_i = E_j(t_{ij}) - \mu$, $b_j = \bar{t}_{.j} - \mu$, and $c_{ij} = t_{ij} - \bar{t}_{.j} - E_j(t_{ij}) + \mu$. The above definitions imply that the model components will satisfy the following conditions: $\sum_i^n \alpha_i = \sum_i^n c_{ij} = E_j(b_j) = E_j(c_{ij}) = 0$.
It is also implicitly assumed that the items are similarly scored and hence on

the same scale. Allowing for heterogeneous error variances yields the following: $\sigma^2(e_i) = E_j(e_{ij}^2)$ and $\sigma^2(e) = (1/n)\sum_i^n \sigma^2(e_i)$.
If the error variances are homogeneous, then $\sigma^2(e_i) = \sigma^2(e)$ for all i.

Let $\underset{\sim}{t}_j$ represent the n dimensional column vector of examinee j's true scores on the n items. The true score covariance matrix is
$\underset{\sim}{\Sigma} = \{\sigma_{ii'}\} = E_j[(\underset{\sim}{t}_j - E_j(\underset{\sim}{t}_j))'(\underset{\sim}{t}_j - E_j(\underset{\sim}{t}_j))]$. The only restriction placed on $\underset{\sim}{\Sigma}$ is that it be positive semi-definite. The covariance among the items may be of a very general form, including any multiple common factor model. This is quite different from the RA model where a simple specific conditional covariance structure is assumed. Removing the randomness of the items permits a much more general covariance structure among the items, but eliminates any statistical inferences concerning the population of items.

From the definitions of the random model components, the variances and covariances for these components may be expressed as functions of the $\{\sigma_{ii'}\}$. Scheffe´ (1959) shows that

$$\mathrm{Var}_j(b_j) = E_j(b_j^2) = \bar{\sigma} \tag{7}$$

$$\mathrm{Cov}_j(c_{ij}, c_{i'j}) = E_j(c_{ij} * c_{i'j}) = \sigma_{ii'} - \bar{\sigma}_{i} - \bar{\sigma}_{\ i} + \bar{\sigma} \quad , \text{ and} \tag{8}$$

$$\mathrm{Cov}_j(b_j, c_{ij}) = E_j(b_j * c_{ij}) = \bar{\sigma}_{i.} - \bar{\sigma}_{..} \ . \tag{9}$$

Scheffe´ (1959) defines the variance components as

$$\sigma^2(b) = \mathrm{Var}_j(b_j) \qquad \text{and} \tag{10}$$

$$\sigma^2(c) = [1/(n-1)]\sum_i^n \mathrm{Var}_j(c_{ij}) = [1/(n-1)]\sum_i^n(\sigma_{ii} - \bar{\sigma}_{..}) \ . \tag{11}$$

Using these definitions, he shows that MS_b and MS_c, as previously defined under the RA model, have the following expected values under the MA model: $E_N(MS_b) = n\sigma^2(b) + \sigma^2(e)$ and $E_N(MS_c) = \sigma^2(c) + \sigma^2(e)$, where E_N denotes the expectation over an infinite number of random samples of N examinees.

It is interesting to note that the random components are correlated in the MA model and that these correlations are determined by $\underset{\sim}{\Sigma}$. In the RA model the random components are uncorrelated, but the covariances among the items are required to be homogeneous. What happens to the component correlations in the MA model when the inter-item covariances are assumed to be homogeneous will be investigated shortly.

First, however, reliability and its relationship to coefficient alpha will be discussed. The sample alpha coefficient under the MA model is identical to the sample alpha for the RA model, and is given as $\hat{\alpha} = (MS_b - MS_c)/MS_b$. Its parametric counterpart under the MA model will be denoted by α_{MA} and is defined as

$$\alpha_{MA} = \frac{[E_N(MS_b) - E_N(MS_c)]}{E_N(MS_b)} = \frac{\sigma^2(b) - \sigma^2(c)/n}{\sigma^2(b) + \sigma^2(e)/n} . \tag{12}$$

The rationale for this definition is that $\hat{\alpha}$ converges in probability to α_{MA} under the MA model. This is further discussed below. If (1) the random model components including the errors are normally distributed, (2) the error variances are homogeneous (though mild heterogeneity should be acceptable), and (3) $\sigma^2(c) = 0$, then using results given by Scheffe′ (1959) it may be shown that $[(1 - \alpha_{MA})/(1 - \hat{\alpha})]$ is distributed as $F[N-1, (n-1)(N-1)]$, which is the same distribution as under the RA model. Similarly, this F distribution implies that $E_N(\hat{\alpha}) = [(N-1)/(N-3)]\alpha_{MA} - [2/(N-3)]$, and hence that $\hat{\alpha}$ is an asymptotically unbiased and consistent estimate of α_{MA} . Kristof (1963) has previously derived these results. If $\sigma^2(c) \neq 0$, then the F distribution still holds if Σ has the highly symmetric structure discussed by Scheffe´ (1959, p 264) or if Σ_x has the type H form described by Huynh and Feldt (1970); but as will be seen later α_{MA} is then a strict lower bound to reliability. However, even if the foregoing assumptions are not fulfilled, $\hat{\alpha}$ is still a consistent estimator of α_{MA} since it is a method of moments estimate for α_{MA} (Serfling, 1983), and equivalently converges in probability to α_{MA} . Finally, it should be noted that if $\sigma^2(c) = 0$, then all the $c_{ij} = 0$ and the MA model is identical to the essentially tau equivalent model discussed by Lord and Novick (1968).

Under the MA model, the mean true score of examinee j is $\bar{t}_{\cdot j} = (1/n)\sum_i^n E(x_{ijk})$ where, as discussed under the RA model, E denotes expectation over the errors associated with random replications. Let x_j denote the n dimensional column vector of the j-th examinee's observed scores on the n items. Let Σ_x denote the covariance matrix for the observed scores. It follows that $\Sigma_x = \Sigma + \Delta(\sigma^2(e_i))$ where $\Delta(\sigma^2(e_i))$ is a diagonal matrix with the error variances as its elements. Following Lord and Novick (1968, chap. 3), reliability under the MA model is defined as

$$\mathrm{Rel}(\bar{x}_{\cdot j},\bar{t}_{\cdot j}) = [\mathrm{Cor}_j(\bar{x}_{\cdot j},\bar{t}_{\cdot j})]^2 = \sigma^2(b)/[\sigma^2(b) + \sigma^2(e)/n] \qquad (13)$$

$$= \mathrm{Var}_j(\bar{t}_{\cdot j})/\mathrm{Var}_j(\bar{x}_{\cdot j}) \quad .$$

The above follows from the expressions for the variance components given in (10) and (11). Comparison of the last expression in the first line of (13) with the expression for α_{MA} given in (12) demonstrates that $\alpha_{MA} = \mathrm{Rel}(\bar{x}_{\cdot j},\bar{t}_{\cdot j})$ if and only if $\sigma^2(c) = 0$, i.e., the items are essentially tau equivalent. Otherwise, $\alpha_{MA} < Rel(\bar{x}_{\cdot j}, \bar{t}_{\cdot j})$. This agrees with the results of Guttman (1945), Novick and Lewis (1967), Bentler (1972), and Jackson and Agunwamba (1977).

Under the assumption of equivalent covariance structures for the RA and MA models, comparisons between the two models regarding variance components, reliability, and coefficient alpha will now be undertaken. The RA true score conditional covariance structure given in (4) may be reexpressed as

$$\Sigma_{t|n} = q\underset{\sim}{J} + \Delta(u_i^2) \qquad (14)$$

where $\sigma^2(b) = q$ and $\sigma^2(c_i) = u_i^2$. The following true score covariance structure will be assumed for the MA model:

$$\underset{\sim}{\Sigma} = q\underset{\sim}{J} + \underset{\sim}{\Delta}(u_i^2) \quad . \qquad (15)$$

For the above covariance structure, Table 1 displays the variance

---

Insert Table 1 about here

---

components for the RA and MA models. This paper has followed the convention of labeling the variance components the same in both models, but Table 1 shows that the variance components have different meanings under the two models. While $\sigma^2(c)$ depends only on the specific variances, though in different ways in the two models, $\sigma^2(b)$ includes common and specific variances under the MA model but only common variance under the RA model. For more complicated covariance structures than (15) under the MA model, such simple relationships between the variance components and the covariance matrix are not apparent.

The differences in variance components between the two models have ramifications for reliability and coefficient alpha under the two models.

Table 2 displays alpha and reliabilities for the two models under the

---

Insert Table 2 about here

---

indicated covariance structure. Coefficient alpha differs statistically under the two models in that expectations are used in the denominator of α_{RA} while summations are used in the denominator of α_{MA}. Nonetheless, coefficient alpha has a similar psychometric meaning under the two models since under both models the numerator and denominator depend, with slight variations, on the same elements of the covariance matrix. $Rel(\bar{x}_{.j},\bar{t}_{.j})$ is identical under the two models, but differs from $Rel(\bar{x}_{.j},\tau_j)$ under the RA model as has already been noted.

Under the RA model, the random model components are uncorrelated as was previously discussed. For the MA model under the covariance structure in (15),

$$Cov_j(b_j,c_{ij}) = [q + (u_i^2/n)] - [q + (1/n^2)\sum_i^n u_i^2]$$

$$= [u_i^2 - (1/n)\sum_i^n u_i^2]/n \quad \text{and}$$

$$Cov_j(c_{ij},c_{i'j}) = q - (q + u_i^2/n) - (q + u_{i'}^2/n) + [q + (1/n^2)\sum_i^n u_i^2]$$

$$= [(1/n)\sum_i^n u_i^2 - u_i^2 - u_{i'}^2]/n \quad .$$

If all the u_i^2 are equal, then $Cov_j(b_j,c_{ij}) = 0$ and $Cov_j(c_{ij},c_{i'j}) = -u^2/n$ where u^2 is the common value for all the u_i^2. The covariance $-u^2/n$ is due to the fact that under the MA model $\sum_i^n c_{ij} = 0$ for all j. As was noted previously for the RA model, the uncorrelatedness of the random model components results in the simple covariance structure given in (4) and (14). What has just been shown is that when a slightly simpler covariance structure is assumed for the MA model, the random model components essentially become uncorrelated. Hence, the correlations among the random components and the inter-item covariances are related in a similar fashion under both models. To obtain psychometric inference under a more complicated inter-item covariance structure than (14) requires an RA type model which permits the model components to be correlated. Such correlations would make expressions for the mean squares much more difficult to obtain.

Finally, when the u_i^2 are homogeneous and hence the equicorrelation covariance structure presented by Morrison (1976) (that is equivalent to Scheffe´'s (1959) highly symmetric covariance structure) holds, then $n\sigma^2(b) = \lambda_1$ where λ_1 is the first and largest eigenvalue of $\underset{\sim}{\Sigma}$ in the MA model. The one remaining distinct eigenvalue of Σ, λ_2, has multiplicity n-1 and is equal to $\sigma^2(\dot{c})$.

Summary and Discussion of Implications for Practice

It has been shown that coefficient alpha is approximately equal to but not necessarily a lower bound to reliability under the RA model, and that it is a lower bound to reliability under the MA and FA models (the result for the FA model having been shown previously by others). These conclusions concern the parameter values for these quantities and not necessarily their sample estimates. Under the RA model where statistical inference to a population of items from a sample of items is permitted, it was found that the inter-item covariances must be homogeneous, and that this homogeneity is due to the model components being uncorrelated. This restriction is not required under the MA model, but it does not permit psychometric inference. These conclusions are, of course, specific to the models under consideration, and other models may yield different results.

It is usually the case in education and psychology that inference from a sample of items to a population of items is a desired goal in the analysis of test data. However, this may not always be true. A situation in educational measurement where psychometric inference may not be required is when a test is divisible into well defined content heterogeneous subtests, and the subtest scores are the measurements being analyzed. In this situation, an appropriate model for the data could be a subtest by examinee two-way MA model. In psychology, if an affective scale such as a personality inventory consists of well defined psychologically distinct subscales, then a subscales by subjects two-way MA model could also be an appropriate model for the data.

If psychometric inference is desired and if the RA model presented within is going to be used to analyze the data, then it is appropriate to investigate whether or not the data satisfy the covariance structure assumed under the RA model. This covariance structure is a linear covariance structure, and Browne (1972) has derived a procedure based on the principle of generalized least squares (GLS) estimation that may be used to statistically test the fit of the data to the RA model covariance structure. Browne's (1972) method is non-iterative and hence relatively simple computationally. Jöreskog (1978) discusses statistical tests for covariance structures based on GLS and maximum likelihood (ML) estimation methods. The computer program LISREL VI (Jöreskog and Sörbom, 1986) implements those methods as well as others, and is accessible through the $SPSS^X$ ($SPSS^X$ Inc, 1986) computer program. Bentler (1983) and Browne (1984) have developed GLS test procedures with weaker distributional assumptions but more computational complexity. Bentler (1985) has also written a computer program, EQS, which implements his procedure and is available as part of the BMDP Statistical Software computer package. It is designed for easy use. If the RA model fits the data, then $\hat{\alpha}$ is an appropriate estimator for the reliability index, $\text{Rel}(\bar{x}_{.j}, \tau_j)$, which assesses how well relationships between observed scores represent relationships between item domain true scores.

If the items are dichotomously scored, then difficulties may arise in applying the above procedures to the usual sample covariance matrix or the sample matrix of phi coefficients. Mislevy (1986) discusses these problems and reviews alternative methods for testing covariance structures designed to deal with dichotomously scored items. However, the results of Collins et al. (1986) suggests that it may be appropriate to first analyze the usual matrix of sample moment covariances or correlations. If difficulties arise, then recourse may be had from the more theoretically and computationally complex methods discussed by Mislevy (1986).

If the RA model cannot be applied because the data substantially violate the requirement of homogeneous inter-item covariances, or inference to a population of items is not desired, then the MA model may be used. As was shown, α is a lower bound to reliability under the MA model and consequently under any FA model (the latter having been shown previously by many others). However, under the MA model, better lower bounds than α exist. The best is the greatest lower bound to reliability, derived independently by Jackson and Agunwamba (1977) and Bentler and Woodward (1980). Bentler and Woodward (1983) present the most efficient numerical algorithm for computing a sample estimate of the greatest lower bound to reliability. In general terms, the computation requires the solution of a nonlinear optimization problem with inequality constraints and is rather complex. For the investigator who desires a simpler estimate, even if it is less optimal, Jackson and Agunwamba (1977) suggest that Guttman's λ_6 coefficient may be advantageous "in the typical situation where the inter-item correlations are positive, modest in size, and rather similar." The computer package SPSS^X (SPSS^X Inc., 1986) has a reliability component which computes a sample estimate for λ_6 as well as several other reliability estimates.

If the test has many items, then some investigators may find it difficult or expensive to compute sample estimates for λ_6 or the greatest lower bound. These investigators may view coefficient α as an appealing reliability index for long tests because of its computational simplicity. Such investigators may find solace in the results of Green, Lissitz, and Mulaik (1977) which suggest that α increases as the number of items increases even when the test has multiple common factors and α is only a strict lower bound to the parameter value of reliability. Green et al. (1977) argue that this result makes α a poor index of test unidimensionality. Fortunately, those qualities which make α a poor index for unidimensionality increase its worth as a reliability index, and this is especially true for long tests. Nonetheless, the greatest lower bound to reliability has optimal properties which indicate that it is worth computing whenever feasible.

Finally, because coefficient alpha may be a useful estimate of reliability under both the RA and MA models, it is worthwhile to review the F distribution theory for $\hat{\alpha}$ under both models. In addition to the appropriate normality assumptions for each model, the F distribution theory requires homogeneity of error variances under both ANOVA models and homogeneity of interaction conditional variances under the RA model, but mild heterogeneity of these variances should not greatly affect the distribution theory. Under the RA model, α may equal or approximately equal reliability when the F distribution for $\hat{\alpha}$ holds, but α is not a lower bound for reliability. Under the MA model, the F distribution theory for $\hat{\alpha}$ holds and α equals reliability when there are no interactions. If interactions are present, then the F distribution theory for $\hat{\alpha}$ requires the special covariance structures of Scheffé (1959, p 264) or Huynh and Feldt (1970) and α is then a strict lower bound to reliability. If a conservative estimate of α or the parameter value of reliability under either model is desired, then Woodward and Bentler (1978) show how the F distribution theory for $\hat{\alpha}$ may be used to obtain a probabilistic lower bound to α.

References

Bentler, P. M. (1972). A lower bound estimate for the dimension-free measurement of internal consistency. Social Science Research, 1, 343-357.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. Psychometrika, 48, 493-517.

Bentler, P. M. (1985). Theory and implementation of EQS a structural equations program. Los Angeles: BMDP Statistical Software, Inc.

Bentler, P. M., & Woodward, J. A. (1980). Inequalities among lower bounds to reliability: With applications to test construction and factor analysis. Psychometrika, 45, 249-267.

Bentler, P. M., & Woodward, J. A. (1983). The greatest lower bound to reliability. In H. Wainer and S. Messick (Eds.), Principals of modern psychological measurement. Hillsdale, NJ: Lawrence Erlbaum.

Brennan, R. L. (1983). Elements of generalizability. Iowa City, IA: American College Testing Program.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. South African Statistical Journal, 8, 1-24.

Browne, M. W. (1984). Asymptotic distribution-free methods for the analysis of covariance structure. British Journal of Mathematical and Statistical Psychology, 37, 62-83.

Collins, L. M., Cliff, N., McCormick, D. J., & Zatkin, J. L. (1986). Factor recovery in binary data sets: A simulation. Multivariate Behavioral Research, 21, 377-391.

Cornfield, J., and Tukey, J. W. (1956). Average values of mean squares in factorials. The Annals of Mathematical Statistics, 27, 907-949.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. _Psychometrika_, _37_, 62-83.

Cronbach, L., Gleser, G., Nanda, H., and Rajaratnam, N. (1972). _The dependability of behavioral measurements: Theory of generalizability for scores and profiles_. New York: Wiley.

Feldt, L. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. _Psychometrika_, _30_, 357-370.

Gabrielsson, A., & Seeger, P. (1971). Tests of significance in two-way designs (mixed model) with dichotomous data. _British Journal of Mathematical and Statistical Psychology_, _24_, 111-116.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. _Educational and Psychological Measurement_, _37_, 827-838.

Guttman, L. (1945). A basis for analyzing test-retest reliability. _Psychometrika_, _10_, 255-282.

Hakstian, A., & Whalen, T. (1976). A k-sample significance test for independent alpha coefficients. _Psychometrika_, _41_, 219-231.

Harmon, H. (1976). _Modern factor analysis_ (3rd ed. revised). Chicago: University of Chicago Press.

Hocking, R. (1985). _The analysis of linear models_. Monterey: Brook/Cole.

Hotelling, H. (1933). Analysis of a complex of statistical variables unto principal components. _Journal of Educational Psychology_, _24_, 417-441, 498-520.

Hoyt, C. (1941). Test reliability estimated by the analysis of variance. _Psychometrika_, _6_, 153-160.

Hsu, T., & Feldt, L. (1969). The effect of limitations on the number of criterion score values on the significance level of the F-test. *American Educational Research Journal*, *6*, 515-527.

Huynh, H., and Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurement designs have exact F-distributions. *Journal of the American Statistical Association*, *65*, 1582-1589.

Jackson, P., & Agunwamba, C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika*, *42*, 567-578.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*, 443-477.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI user's guide*. Morrsville, IN: Scientific Software, Inc.

Kaiser, H., & Michael, W. (1975). Domain validity and generalizability. *Educational and Psychological Measurement*, *35*, 31-35.

Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics* (Vol. I, 4th ed.) London: Griffin.

Kristof, W. (1973). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika*, *28*, 221-238.

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

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum Associates.

McDonald, R. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*, 100-117.

McDonald, R. (1978). Generalizability in factorable domains: "domain validity and generalizability." Educational and Psychological Measurement, 38, 75-79.

McDonald, R., & Ahlawat, K. (1974). Difficulty factors in binary data. British Journal of Mathematical and Statistical Psychology, 27, 82-99.

McDonald, R. P. & Mulaik, S. A. (1979). Determinacy of common factors: A non-technical review. Psychological Bulletin, 86, 297-306.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. Journal of Educational Statistics, 11, 3-31.

Mulaik, S. (1972). The foundations of factor analysis. New York: McGraw-Hill.

Mulaik, S. A., & McDonald, R. P. (1978). The effect of additional variables on factor indeterminacy. Psychometrika, 43, 177-192.

Morrison, D. (1976). Multivariate statistical methods. New York: McGraw-Hill.

Novick, M., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. Psychometrika, 32, 1-13.

Rozeboom, W. (1978). Domain validity--why care. Educational and Psychological Measurement, 38, 81-88.

Scheffe´, H. (1956a). A "mixed model" for the analysis of variance. The Annals of Mathematical Statistics, 27, 23-36.

Scheffe´, H. (1956b). Alternative models for the analysis of variance. The Annals of Mathematical Statistics, 27, 251-271.

Scheffe´, H. (1959). The analysis of variance. New York: Wiley.

Seeger, P., & Gabrielsson, A. (1968). Applicability of the Cochran Q test and the F test for statistical analysis of dichotomous data for dependent samples. Psychological Bulletin, 69, 269-277.

Serfling, R. (1983). _Approximation theorems of mathematical statistics_. New York: Wiley.

Spearman, C. (1904). General intelligence, objectively determined and measured. _American Journal of Psychology_, _15_, 201-293.

SPSS Inc. (1986). _SPSS^X user's guide_, (2nd ed.). Chicago: SPSS Inc.

Tryon, R. (1957). Reliability and behavior domain validity: Reformulation and historical critique. _Psychological Bulletin_, _54_, 229-249.

Williams, J. S. (1978). A definition for the common-factor analysis model and the elimination of problems of factor score indeterminacy. _Psychometrika_, _43_, 293-306.

Woodward, J. A., & Bentler, P. M. (1978). A statistical lower bound to population reliability. _Psychological Bulletin_, _85_, 1323-1326.

Table 1

A Comparison Between Variance Components for the RA and MA Models Under the Indicated Covariance Structure for Both Models

RA Model	MA Model
$\Sigma_{t\|n} = q\underset{\sim}{J} + \underset{\sim}{\Delta}(u_i^2)$	$\underset{\sim}{\Sigma} = q\underset{\sim}{J} + \underset{\sim}{\Delta}(u_i^2)$
$\sigma^2(b) = q$	$\sigma^2(b) = q + (1/n^2)\sum_i^n u_i^2$
$Var_j(c_{ij}) = \sigma^2(c_i) = u_i^2$	$Var_j(c_{ij}) = [(n-2)u_i^2 + (1/n)\sum_i^n u_i^2]/n$
$\sigma^2(c) = E_i[\sigma^2(c_i)] = E_i(u_i^2)$	$\sigma^2(c) = [1/(n-1)]\sum_i^n Var_j(c_{ij}) = [1/n]\sum_i^n u_i^2$

Table 2

A Comparison Between Alpha Coefficients and Reliabilities for the RA and MA Models Under the Indicated Covariance Structure for Both Models

RA Model	MA Model
$\underset{\sim}{\Sigma}_{t\|n} = q\underset{\sim}{J} + \underset{\sim}{\Delta}(u_i^2)$	$\underset{\sim}{\Sigma} = q\underset{\sim}{J} + \underset{\sim}{\Delta}(u_i^2)$
$\alpha_{RA} = \frac{q}{q + (1/n)E_i(u_i^2) + (1/n)E_i[\sigma^2(e_i)]}$	$\alpha_{MA} = \frac{q}{q + (1/n^2)\sum_i^n(u_i^2) + (1/n^2)\sum_i^n[\sigma^2(e_i)]}$
$Rel(\bar{x}_{.j}, \tau_j) = \frac{q}{q + (1/n^2)\sum_i^n(u_i^2) + (1/n^2)\sum_i^n[\sigma^2(e_i)]}$	
$Rel(\bar{x}_{.j}, \bar{t}_{.j}) = \frac{q + (1/n^2)\sum_i^n(u_i^2)}{q + (1/n^2)\sum_i^n(u_i^2) + (1/n^2)\sum_i^n[\sigma^2(e_i)]}$	$Rel(\bar{x}_{.j}, \bar{t}_{.j}) = \frac{q + (1/n^2)\sum_i^n(u_i^2)}{q + (1/n^2)\sum_i^n(u_i^2) + (1/n^2)\sum_i^n[\sigma^2(e_i)]}$