

*Center for Advanced Studies in  
Measurement and Assessment*

*CASMA Technical Note*

*Number 4*

**Notes about Nominal Weights in  
Multivariate Generalizability Theory**

*Robert L. Brennan*

February 2009

Center for Advanced Studies in  
Measurement and Assessment (CASMA)  
College of Education  
University of Iowa  
Iowa City, IA 52242  
Tel: 319-335-5439  
Fax: 319-384-0505  
Web: [www.education.uiowa.edu/casma](http://www.education.uiowa.edu/casma)

All rights reserved

## Contents

<b>1</b>	<b>Weighting Scenarios</b>	<b>1</b>
<b>2</b>	<b>Example</b>	<b>2</b>
<b>3</b>	<b>Composite for <math>[0, S_c]</math> Metric</b>	<b>3</b>
3.1	Using Total-Score Metric D Study Statistics . . . . .	3
3.2	Using Mean-Score Metric D Study Statistics . . . . .	5
<b>4</b>	<b>Composite for <math>[0, 1]</math> Metric</b>	<b>6</b>
4.1	Using Total-Score Metric D Study Statistics . . . . .	6
4.2	Using Mean-Score Metric D Study Statistics . . . . .	7
<b>5</b>	<b>Other Considerations</b>	<b>7</b>
5.1	Linear Transformations of $X_m$ and/or $X_f$ . . . . .	8
5.2	Item-score Increments Other than 1 . . . . .	9
5.3	Effective Weights . . . . .	9
5.4	Composite Generalizability Coefficient and Stratified $\alpha$ . . . . .	10
<b>6</b>	<b>Concluding Comments</b>	<b>11</b>
<b>7</b>	<b>References</b>	<b>12</b>

Suppose a test consists of  $n_m$  multiple-choice (MC) items scored  $[0, 1, \dots, s_m]$  (almost always  $s_m = 1$ ) and  $n_f$  free-response (FR) items scored  $[0, 1, \dots, s_f]$ . Let  $X_m$  represent MC scores in the total-score metric  $[0, 1, \dots, S_m = n_m s_m]$  and  $X_f$  represent FR scores in the total-score metric  $[0, 1, \dots, S_f = n_f s_f]$ . Note that each of the  $s$  and  $S$  constants has two interpretations: (1) a highest score; and (2) a range of scores.<sup>1</sup> In these notes, conceptually the range interpretation is much more sensible. (Section 5.2 considers cases in which the range and maximum scores are different.)

For this situation, these notes address multivariate generalizability (G) theory issues (see Brennan, 2001a) associated with four sets of nominal weights for forming a composite of MC and FR scores. Here, all four sets of weights are subject to the following:

*Constraint:* MC items contribute  $r$  times as much to the composite score range as FR items.

It is assumed that the G study variance and covariance components, or estimates of them, are available. Usually nominal weights are defined such that they sum to one. Here there is no such restriction.

## 1 Weighting Scenarios

The four weighting scenarios are described next.

- (a) Suppose an investigator wishes to construct a composite,  $C$ , that ranges from 0 to  $S_c$ . We will call this the  $[0, S_c]$  composite-score metric. For this metric, what are  $w_m$  and  $w_f$  such that

$$w_m X_m + w_f X_f = C, \quad (1)$$

given the constraint? Note that since  $X_m$  and  $X_f$  are random variables in the total-score metric for MC and FR items, respectively,  $w_m$  and  $w_f$  are associated with D study variance and covariance components expressed in the total-score metric.

- (b) If computations are performed using mGENOVA (Brennan, 2001b), the D study variance and covariance components will be on the mean-score metric ( $[0, 1, \dots, s_m]$  for MC items and  $[0, 1, \dots, s_f]$  for FR items). We designate the random variables for the mean-score metric as  $\bar{X}_m$  and  $\bar{X}_f$ . For this metric, what are weights  $v_m$  and  $v_f$  such that

$$v_m \bar{X}_m + v_f \bar{X}_f = C, \quad (2)$$

given the constraint? These are the weights that would need to be used in mGENOVA to obtain results for the composite score scale  $C$ .

---

<sup>1</sup>The total number of possible scores is the range plus 1.

- (c) What are the weights needed to transform the composite  $[0, S_c]$  metric to a composite  $[0, 1]$  metric? That is, what are  $w'_m$  and  $w'_f$  such that

$$w'_m X_m + w'_f X_f = 1, \quad (3)$$

given the constraint?

- (d) What are the weights  $v'_m$  and  $v'_f$  such that

$$v'_m \bar{X}_m + v'_f \bar{X}_f = 1, \quad (4)$$

given the constraint? These are the weights that would need to be used in mGENOVA to obtain results for the composite  $[0, 1]$  metric.

## 2 Example

Table 1 provides an example based on 100 MC items and four FR items each of which is scored  $[0, 1, \dots, 10]$ .<sup>2</sup> The first quarter of the table provides estimates of the G study variance and covariance components. Note that  $\widehat{\Sigma}_p$  is a  $2 \times 2$  matrix with a covariance in the off-diagonal positions. Technically, there are two other matrices,  $\widehat{\Sigma}_i$  and  $\widehat{\Sigma}_{pi}$ , but since covariances for these two matrices are all zero, only the estimated variance components are provided to save space.

The left side of the second quarter of Table 1 provides the usual D study statistics for MC and FR in the mean-score metric. The right side provides the same statistics in the total-score metric. These can be obtained easily from the mean-score metric statistics as follows:

- multiply MC variance components on the left side of Table 1 by  $n_m^2 = 100^2 = 10,000$ ;
- multiply FR variance components on the left side of Table 1 by  $n_f^2 = 4^2 = 16$ ;
- multiply the person covariance component on the left side of Table 1 by  $n_m n_f = 100 \times 4 = 400$ ; and
- obtain the total-score metric error variances and coefficients from the total score metric D study variance and covariance components using the usual rules (see, for example, Brennan, 2001a, p. 109).

The total-score D study statistics are identified in Table 1 with an appended “+”.

The bottom half of Table 1 provides composite score results associated with the four weighting schemes, (a)–(d), discussed previously. The third quarter of Table 1 provides results for the composite metric  $[0, S_c = 150]$ , and the bottom

<sup>2</sup>This example has some similarities with a real data example, but it is not real data per se.

quarter provides results for the composite metric  $[0,1]$ . The  $[0,1]$  composite statistics are identified with an appended “\*”. Results labelled (b) and (d) on the left are obtained using mean-score D study statistics; results labelled (a) and (c) on the right are obtained using total-score D study statistics. For this example,  $r = 1.5$  for the constraint on page 1.

### 3 Composite for $[0, S_c]$ Metric

The constraint states that that MC items contribute  $r$  times as many (fractional) score points as FR items. Let  $u_m = r/(r+1)$  be the proportion of the composite score range  $S_c$  associated with MC items and  $u_f = 1/(r+1)$  be the proportion associated with FR items. The constraint means that

$$u_m + u_f = 1. \quad (5)$$

This equation applies to all four cases on pages 1–2. Cases (a) and (b) for the composite  $[0, S_c]$  metric are considered first.

#### 3.1 Using Total-Score Metric D Study Statistics

For case (a) on page 1 we wish to find  $w_m$  and  $w_f$  such that

$$w_m S_m + w_f S_f = S_c, \quad (6)$$

given the constraint in Equation 5. Dividing both sides of Equation 6 by  $S_c$  gives

$$\frac{w_m S_m}{S_c} + \frac{w_f S_f}{S_c} = 1.$$

The two terms to the left of the equal sign are the definitions of  $u_m$  and  $u_f$ , respectively; i.e.,

$$\frac{w_m S_m}{S_c} = u_m$$

and

$$\frac{w_f S_f}{S_c} = u_f.$$

It follows that

$$w_m = \left( \frac{u_m}{S_m} \right) S_c, \quad (7)$$

and

$$w_f = \left( \frac{u_f}{S_f} \right) S_c. \quad (8)$$

For the example in Table 1,

$$w_m = \left( \frac{.6}{100} \right) 150 = .9 \quad \text{and} \quad w_f = \left( \frac{.4}{40} \right) 150 = 1.5.$$

Table 1: Example: Weights for  $[0, S_c = 150]$  and  $[0, 1]$  Composite Metrics with  $r = 1.5$ 

	MC	FR		MC	FR
$n$	100	4			
$s$	1	10			
$S$	100	40			
G study components					
$\widehat{\Sigma}_p$	.02782	.29081			
	.29081	3.29714			
$\hat{\sigma}^2(i)$	.03388	.70123			
$\hat{\sigma}^2(pi)$	.18434	3.44762			
Mean-score metric D study statistics			Total-score metric D study statistics		
$\widehat{\Sigma}_p$	.02782	.29081	$\widehat{\Sigma}_{p+}$	278.20000	116.32400
	.29081	3.29714		116.32400	52.75424
$\hat{\sigma}^2(I)$	.00034	.17531	$\hat{\sigma}^2(I+)$	3.38800	2.80492
$\hat{\sigma}^2(pI)$	.00184	.86191	$\hat{\sigma}^2(pI+)$	18.43400	13.79048
$\hat{\sigma}^2(\delta)$	.00184	.86191	$\hat{\sigma}^2(\delta+)$	18.43400	13.79048
$\hat{\sigma}^2(\Delta)$	.00218	1.03721	$\hat{\sigma}^2(\Delta+)$	21.82200	16.59540
$\mathbf{E}\hat{\rho}^2$	.93786	.79276	$\mathbf{E}\hat{\rho}^2+$	.93786	.79276
$\widehat{\Phi}$	.92727	.76070	$\widehat{\Phi}+$	.92727	.76070
(b) Composite for $[0, 150]$ metric using mean-score metric D study statistics			(a) Composite for $[0, 150]$ metric using total-score metric D study statistics		
$v$	90.00000	6.00000	$w$	.90000	1.50000
$\hat{\sigma}_c^2(p+)$	658.11384		$\hat{\sigma}_c^2(p+)$	658.11384	
$\hat{\sigma}_c^2(\delta+)$	45.96012		$\hat{\sigma}_c^2(\delta+)$	45.96012	
$\hat{\sigma}_c^2(\Delta+)$	55.01547		$\hat{\sigma}_c^2(\Delta+)$	55.01547	
$\mathbf{E}\hat{\rho}_c^2+$	.93472		$\mathbf{E}\hat{\rho}_c^2+$	.93472	
$\widehat{\Phi}_c+$	.92285		$\widehat{\Phi}_c+$	.92285	
(d) Composite for $[0, 1]$ metric using mean-score metric D study statistics			(c) Composite for $[0, 1]$ metric using total-score metric D study statistics		
$v'$	.60000	.04000	$w'$	.00600	.01000
$\hat{\sigma}_c^2(p*)$	.02925		$\hat{\sigma}_c^2(p*)$	.02925	
$\hat{\sigma}_c^2(\delta*)$	.00204		$\hat{\sigma}_c^2(\delta*)$	.00204	
$\hat{\sigma}_c^2(\Delta*)$	.00245		$\hat{\sigma}_c^2(\Delta*)$	.00245	
$\mathbf{E}\hat{\rho}_c^2*$	.93472		$\mathbf{E}\hat{\rho}_c^2*$	.93472	
$\widehat{\Phi}_c*$	.92285		$\widehat{\Phi}_c*$	.92285	

Using the total-score metric D study statistics, it follows that

$$\begin{aligned}\hat{\sigma}_c^2(p+) &= .9^2(278.2) + 1.5^2(52.75424) + 2(.9)(1.5)(116.324) = 658.11384, \\ \hat{\sigma}_c^2(\delta+) &= .9^2(18.434) + 1.5^2(13.79048) = 45.96012, \\ \hat{\sigma}_c^2(\Delta+) &= .9^2(21.822) + 1.5^2(16.59540) = 55.01547, \\ \mathbf{E}\hat{\rho}_c^2+ &= 658.11384/(658.11384 + 45.96012) = .93472, \text{ and} \\ \hat{\Phi}_c+ &= 658.11384/(658.11384 + 55.01547) = .92285.\end{aligned}$$

### 3.2 Using Mean-Score Metric D Study Statistics

It is important to note that in Section 3.1, the results assume that the D study statistics to be used in obtaining composite results are for the *total*-score metric, which is consistent with the use of  $S_m$  and  $S_f$  in Equation 6. However, total-score metric D study statistics are rarely employed in G theory. Rather, the usual convention is to use mean-score metric D study statistics, as in the case for mGENOVA. An obvious question to ask, then, is, “What weights should we use when composite results will be computed using mean-score metric D study statistics?”

The answer employs the same logic as in the previous section. We merely replace  $S_m$  and  $S_f$  (total-score metric ranges) with  $s_m$  and  $s_f$  (mean-score metric ranges). Specifically, to obtain the weights for case (b) on page 1 Equations 6–8 are replaced by

$$v_m s_m + v_f s_f = S_c, \quad (9)$$

$$v_m = \left( \frac{u_m}{s_m} \right) S_c, \quad (10)$$

and

$$v_f = \left( \frac{u_f}{s_f} \right) S_c. \quad (11)$$

For the example in Table 1,

$$v_m = \left( \frac{.6}{1} \right) 150 = 90 \quad \text{and} \quad v_f = \left( \frac{.4}{10} \right) 150 = 6.$$

Using the mean-score metric D study statistics, it follows that

$$\begin{aligned}\hat{\sigma}_c^2(p+) &= 90^2(.02782) + 6^2(3.29714) + 2(90)(6)(.29081) = 658.11384, \\ \hat{\sigma}_c^2(\delta+) &= 90^2(.00184) + 6^2(.86191) = 45.96012, \\ \hat{\sigma}_c^2(\Delta+) &= 90^2(.00218) + 6^2(1.03721) = 55.01547, \\ \mathbf{E}\hat{\rho}_c^2+ &= 658.11384/(658.11384 + 45.96012) = .93472, \text{ and} \\ \hat{\Phi}_c+ &= 658.11384/(658.11384 + 55.01547) = .92285.\end{aligned}$$

Computations with only five decimal digits do not quite give the final results reported above, which are the correct results. These results are identical to those in the Section 3.1.

It is evident that these numerical results are the same as those obtained in Section 3.1, as they must be. The important point is that different weights must be used with the two different types of D study variance and covariance components.

## 4 Composite for [0,1] Metric

When weighting issues of the type considered in these notes are encountered, in the author's experience the composite is usually expressed in a total score metric, as discussed in Section 3. Sometimes, however, multiple tests are under consideration, and there is some interest in comparing composite statistics for the various tests. This is easily accomplished if  $S_c$  is the same for all tests, but if the  $S_c$  values are different, comparisons are difficult because the composite scales are different. One route around this problem is to use a common value of  $S_c$  for all tests *solely* for the purpose of comparing statistics such as estimated error variances. An obvious choice is to put all composites on a [0,1] scale, which is the topic of this section. To do so, all that is required is to set  $S_c = 1$  in the equations in Section 3.

### 4.1 Using Total-Score Metric D Study Statistics

To obtain the weights for case (c) on page 2 Equations 6–8 are replaced by

$$w'_m S_m + w'_f S_f = 1, \quad (12)$$

$$w'_m = \frac{u_m}{S_m}, \quad (13)$$

and

$$w'_f = \frac{u_f}{S_f}. \quad (14)$$

For the example in Table 1,

$$w'_m = \frac{.6}{100} = .006 \quad \text{and} \quad w'_f = \frac{.4}{40} = .01.$$

Using the total-score metric D study statistics, it follows that

$$\begin{aligned} \hat{\sigma}_c^2(p^*) &= .006^2(278.2) + .01^2(52.75424) + 2(.006)(.01)(116.324) = .02925, \\ \hat{\sigma}_c^2(\delta^*) &= .006^2(18.434) + .01^2(13.79048) = .00204, \\ \hat{\sigma}_c^2(\Delta^*) &= .006^2(21.822) + .01^2(16.59540) = .00245, \\ \mathbf{E}\hat{\rho}_c^{2*} &= .02925/ (.02925 + .00204) = .93472, \quad \text{and} \\ \hat{\Phi}_c^* &= .02925/ (.02925 + .00245) = .92285. \end{aligned}$$

Computations with only five decimal digits do not quite give the final results for the coefficients reported above, which are the correct results.

## 4.2 Using Mean-Score Metric D Study Statistics

To obtain the weights for case (d) on page 2 Equations 6–8 are replaced by

$$v'_m s_m + v'_f s_f = 1, \quad (15)$$

$$v'_m = \frac{u_m}{s_m}, \quad (16)$$

and

$$v'_f = \frac{u_f}{s_f}. \quad (17)$$

For the example in Table 1,

$$v'_m = \frac{.6}{1} = .6 \quad \text{and} \quad v'_f = \frac{.4}{10} = .04.$$

Using the mean-score metric D study statistics, it follows that

$$\begin{aligned} \hat{\sigma}_c^2(p^*) &= .6^2(.02782) + .04^2(3.29714) + 2(.6)(.04)(.29081) = .02925, \\ \hat{\sigma}_c^2(\delta^*) &= .6^2(.00184) + .04^2(.86191) = .00204, \\ \hat{\sigma}_c^2(\Delta^*) &= .6^2(.00218) + .04^2(1.03721) = .00245, \\ \mathbf{E}\hat{\rho}_c^{2*} &= .02925/ (.02925 + .00204) = .93472, \quad \text{and} \\ \widehat{\Phi}_{c^*} &= .02925/ (.02925 + .00245) = .92285. \end{aligned}$$

Computations with only five decimal digits do not quite give the final results reported above, which are the correct results.

It is evident that these numerical results are the same as those obtained in Section 4.1, as they must be. The important point is that different weights must be used with the two different types of D study variance and covariance components. Note, as well that the generalizability coefficients are the same for all four cases, as are the Phi coefficients, which necessarily follows from the fact that all four transformations are linear.

## 5 Other Considerations

For simplicity, to this point, all score ranges have been defined such that the lowest score is 0. Obviously, any score range is unchanged if the same constant is added to all scores in the range.<sup>3</sup> For example, if MC items were scored [1,2], the results in Table 1 would be unchanged. As discussed next, a somewhat more challenging issue arises if the variables  $X_m$  and/or  $X_f$  are linearly transformed.

<sup>3</sup>Note also that the focus of these notes is composite score *variances* or functions of them. Clearly results are unchanged if the same constant is added to all scores.

### 5.1 Linear Transformations of $X_m$ and/or $X_f$

*Case (a).* Suppose  $\mathcal{X}_m = a_m + b_m X_m$  and  $\mathcal{X}_f = a_f + b_f X_f$ , and assume that the nominal weights (which we will designate  $\ddot{w}_m$  and  $\ddot{w}_f$ ) apply to these transformed variables. That is,

$$\ddot{w}_m \mathcal{X}_m + \ddot{w}_f \mathcal{X}_f = C.$$

To keep matters simple, we will also assume that  $u_m$ ,  $u_f$ , and  $S_c$  are unchanged. Since the range is unaffected by the addition of a constant to all scores within the range, it is sufficient to consider  $\mathcal{X}_m = b_m X_m$  and  $\mathcal{X}_f = b_f X_f$ .

Let  $\ddot{S}_m = b_m S_m$  and  $\ddot{S}_f = b_f S_f$ . We wish to find  $\ddot{w}_m$  and  $\ddot{w}_f$  such that

$$\ddot{w}_m \ddot{S}_m + \ddot{w}_f \ddot{S}_f = C,$$

subject to the constraint in Equation 5.

Following the same logic as in Section 3.1, we obtain

$$\ddot{w}_m = \left( \frac{u_m}{\ddot{S}_m} \right) S_c = \left( \frac{u_m}{b_m S_m} \right) S_c = \frac{w_m}{b_m}, \quad (18)$$

where the last result follows from Equation 7. Similarly,

$$\ddot{w}_f = \left( \frac{u_f}{\ddot{S}_f} \right) S_c = \left( \frac{u_f}{b_f S_f} \right) S_c = \frac{w_f}{b_f}. \quad (19)$$

In this sense,  $\ddot{w}_m$  and  $\ddot{w}_f$  absorb the slopes of the transformations.

Returning to the example in Table 1 and Section 3.1, the computations proceed in the same manner except that  $w_m$  and  $w_f$  are replaced by  $\ddot{w}_m$  and  $\ddot{w}_f$ , respectively. In particular, the *same* total-score metric D study statistics for MC and FR items are used; these statistics are *not* multiplied by the slopes.

*Other cases.* The same type of logic applies to the other cases. For example, for case (b) the analogue of Equation 2 is

$$\ddot{v}_m \overline{\mathcal{X}}_m + \ddot{v}_f \overline{\mathcal{X}}_f = C;$$

and the weights are

$$\ddot{v}_m = \frac{v_m}{b_m} \quad (20)$$

and

$$\ddot{v}_f = \frac{v_f}{b_f}, \quad (21)$$

where  $b_m$  and  $b_f$  are the slopes of the transformations for converting  $\overline{X}_m$  to  $\overline{\mathcal{X}}_m$  and  $\overline{X}_f$  to  $\overline{\mathcal{X}}_f$ . In this case, computations use the mean-score metric D study statistics for MC and FR items, and these statistics are *not* multiplied by the slopes.

## 5.2 Item-score Increments Other than 1

Since  $s_m$ ,  $s_f$ ,  $S_m$ ,  $S_f$ , and  $S_c$  are all ranges, the results provided in these notes do not require that contiguous scores all differ by 1, although the early part of this paper used 1 as the increment. For example, a three-point rubric might have scores of [.1, .2, .3] or [10, 20, 30]. As illustrated below, however, when the increments for MC and FR item scores differ, the composite score scale may have characteristics judged undesirable.

Suppose, there are three MC items scored [0, 1] and two FR items scored [10, 20, 30], with  $u_m = .4$ ,  $u_f = .6$ , and  $S_c = 90$ . Using the results in case (a) it is easy to determine that  $w_m = 12$ ,  $w_f = 1.35$ , and the possible composite scores are:

27.0	39.0	40.5	51.0	52.5	54.0	63.0	64.5	66.0	67.5
76.5	78.0	79.5	81.0	90.0	91.5	93.0	103.5	105.0	117.0

Note that the distances between contiguous scores differ dramatically, varying between 1.5 and 12.0. This characteristic of the composite score scale may render it less than ideal in some circumstances.

Also, the results derived in these notes do not require that there be a constant difference between contiguous item scores. So, for example, these results apply to formula-scored MC items<sup>4</sup> where the possible scores are  $[-1/(k-1), 0, 1]$  with  $k$  being the number of alternatives. In such cases, however, almost certainly the composite score scale will have characteristics similar to (or even more extreme than) the example in the preceding paragraph.

## 5.3 Effective Weights

Effective weights in multivariate G theory are discussed by Brennan (2001a, pp. 306–307). In general, an effective weight provides the proportion of a particular composite variance attributable to a level of a fixed facet. The effective weights considered here are the same for cases (a)–(d).

Relative to composite universe score variance, the effective weights for MC and FR items, expressed in terms of case (a), are:

$$ew_m(\tau) = \frac{w_m^2 \sigma_m^2(p+) + w_m w_f \sigma_{mf}(p+)}{\sigma_c^2(p+)} \quad (22)$$

and

$$ew_f(\tau) = \frac{w_m w_f \sigma_{mf}(p+) + w_f^2 \sigma_f^2(p+)}{\sigma_c^2(p+)}, \quad (23)$$

respectively. For the example,

$$\widehat{ew}_m(\tau) = \frac{.9^2(278.2) + .9(1.50)(116.324)}{658.11384} = .58102$$

<sup>4</sup>The author is not recommending use of formula scoring.

and

$$\widehat{ew}_f(\tau) = \frac{.9(1.5)(116.324) + 1.5^2(52.75424)}{658.11384} = .41898.$$

Relative to composite relative error variance, the effective weights for MC and FR items, expressed in terms of case (a), are:

$$ew_m(\delta) = \frac{w_m^2 \sigma_m^2(\delta+)}{\sigma_c^2(\delta+)} \quad \text{and} \quad ew_f(\delta) = \frac{w_f^2 \sigma_f^2(\delta+)}{\sigma_c^2(\delta+)}.$$

Relative to composite absolute error variance, the effective weights for MC and FR items, are:

$$ew_m(\Delta) = \frac{w_m^2 \sigma_m^2(\Delta+)}{\sigma_c^2(\Delta+)} \quad \text{and} \quad ew_f(\Delta) = \frac{w_f^2 \sigma_f^2(\Delta+)}{\sigma_c^2(\Delta+)}.$$

The following table provides estimated effective weights for the example along with  $u_m$  and  $u_f$  in Equation 5.

	MC	FR
$u$	.6	.4
$\widehat{ew}(\tau)$	.58102	.41898
$\widehat{ew}(\delta)$	.32488	.67512
$\widehat{ew}(\Delta)$	.32129	.67871

It is evident that the effective weights relative to composite universe score variance are quite close to the proportional weights for the composite score range. However these are quite different from the effective weights relative to composite error variance (both the  $\delta$  and  $\Delta$  types), which clearly indicate that FR items contribute more than twice as much as MC items to error variance.

#### 5.4 Composite Generalizability Coefficient and Stratified $\alpha$

As noted previously and illustrated in Table 1,  $(\mathbf{E}\hat{\rho}_c^2+) = (\mathbf{E}\hat{\rho}_c^2*)$ , since all the transformations considered here are linear. Furthermore, as discussed below, for the design considered in these notes, these coefficients are identical to stratified  $\alpha$  (Rajaratnam, Cronbach, & Gleser, 1965).

In the terminology and notation of generalizability theory, the design discussed in this paper is the multivariate  $p^\bullet \times I^\circ$  design (see Brennan, 2001a, sect. 9.1). This notation means that: (i) for both MC and FR items, the design is  $p \times i$ ; (ii) persons respond to both types of items (signified by the closed circle following  $p$ ); (iii) any given item is either MC or FR (signified by the open circle following  $i$ )—i.e., items are nested within levels of the fixed item-type facet). For this design, Jarjoura and Brennan (1982, 1983) demonstrated that the generalizability coefficient is identical stratified  $\alpha$ .

In terms of the notation for the  $[0, S_c]$  metric used in these notes,

$$\hat{\alpha}_{strat} = 1 - \frac{w_m^2 \hat{\sigma}_m^2(\delta+) + w_f^2 \hat{\sigma}_f^2(\delta+)}{\hat{\sigma}_c^2(p+) + \hat{\sigma}_c^2(\delta+)}, \quad (24)$$

and for the example

$$\alpha_{strat} = 1 - \frac{.9^2(18.43400) + 1.5^2(13.79048)}{658.11384 + 45.96012} = .93472,$$

which is identical to  $\mathbf{E}\hat{\rho}_c^2+$ . Changing “+” to “\*” in Equation 24 gives the corresponding equation for the  $[0,1]$  metric.

Whether  $\mathbf{E}\hat{\rho}_c^2+$ ,  $\mathbf{E}\hat{\rho}_c^2*$ , or  $\hat{\alpha}_{strat}$  is used, of course, the nominal weights must be determined, and that is the principal purpose of these notes. Beyond that, however, since all three coefficients are equal, one might ask, “Why even consider the more complicated multivariate G theory approach in this paper?” There are several reasons. First, the multivariate G theory approach provides a much richer set of results than a single coefficient. Second, the multivariate G theory approach for the  $p^\bullet \times i^\circ$  design considered here can be used directly with mGENOVA. Third, the approach is easily generalized to many other multivariate designs, the most common of which are accommodated by mGENOVA.

## 6 Concluding Comments

These notes have considered four types of nominal weights in multivariate G theory that might be used to form a composite of MC and FR scores, where the composite has a specified range, and weights are subject to the constraint that MC items contribute  $r$  times as much to the composite score range as FR items. Usually nominal weights are defined such that they sum to one. Here there is no such restriction.

Two sets of weights [labelled cases (b) and (d) on pages 1 and 2, respectively] use mean-score metric D study statistics, which makes them directly applicable using mGENOVA. The other two sets of weights, [labelled cases (a) and (c) on pages 1 and 2, respectively] use total-score metric D study statistics. Composite results for cases (a) and (b) are identical; a similar statement holds for cases (c) and (d). Composite generalizability coefficients and Phi coefficients are the same for all four cases.

The types of weights considered in these notes are used in some operational testing programs and appear reasonable in certain circumstances. However, the author offers no general endorsement of such weights.

To keep the discussion relatively simple, these notes have assumed that a composite consists of only two sections or types of items, MC and FR. Results are easily generalized to more than two sections or item types.

Also, to keep the discussion relatively simple, it has been assumed that  $s_f$  is a constant for all  $n_f$  FR items. If this is not true, one alternative is to treat all FR items with the same  $s_f$  as belonging to the same section or stratum. Then

there will be as many FR strata as there are values of  $s_f$ . Obviously, if this approach is adopted, then there will have to be weights associated with each of these strata, and choosing those weights will involve the types of considerations addressed in these notes.

## 7 References

- Brennan, R. L. (2001a). *Generalizability theory*. New York: Springer-Verlag.
- Brennan, R. L. (2001b). *mGENOVA* (Version 2.1) [Computer software and manual]. Iowa City, IA: University of Iowa. (Available on <http://www.education.uiowa.edu/casma/>)
- Jarjoura, D. & Brennan, R. L. (1982). A variance components model for measurement procedures associated with a table of specifications. *Applied Psychological Measurement*, *6*, 161–171.
- Jarjoura, D. & Brennan, R. L. (1983). Multivariate generalizability models for tests developed according to a table of specifications. In L. J. Fyans (Ed.), *New directions for testing and measurement: Generalizability theory* (No.18) (pp. 83–101). San Francisco: Jossey-Bass.
- Rajaratnam, N., Cronbach, L. J., & Gleser, G. C. (1965). Generalizability of stratified-parallel tests. *Psychometrika*, *30*, 39–56.