*Center for Advanced Studies in Measurement and Assessment*

*CASMA Research Report*

*Number 44*

# Utility Indexes for Composite Scores: Matrix Formulation Including R Code and Examples

*Hyung Jin Kim*
*Robert Brennan*†

July 2015
Revised August 2016

†Hyung Jin Kim is Research Scientist, Center for Advanced Studies in Measurement and Assessment (CASMA), College of Education, University of Iowa (email: hyungjin-kim@uiowa.edu). Robert Brennan is E. F. Lindquist Chair in Measurement and Testing and Founding Director, Center for Advanced Studies in Measurement and Assessment (CASMA), College of Education, University of Iowa (email: robert-brennan@uiowa.edu).

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

# Contents

**Abstract** **iv**

**1 Variables in Matrix Form** **1**

**2 Relative Utility Index $U_r$ for Composite Scores** **3**

**3 Relative Utility $\tilde{U}_r$ for Scale Scores** **5**
3.1 Matrix Notation . . . . . 5
3.2 Covariance of $\tilde{Z}$ and $\tilde{X}$ . . . . . 5
3.3 Estimating $\sigma^2(\tilde{E}_i)$ and $\sigma(\tilde{T}, \tilde{Z})$ . . . . . 6

**4 Example** **7**
4.1 Number-correct Scores on Items . . . . . 7
4.2 Scale Scores . . . . . 9
4.3 Results . . . . . 9

**5 Comments** **10**

**6 References** **10**

**7 APPENDIX A: R Code** **11**

**8 APPENDIX B: Example Using R Code** **17**

## Abstract

Suppose there exist $k$ tests in a battery, any weighted combination of which generates a composite score denoted $Z$. Suppose there also exists an additional composite test score, denoted $X$, that is created based on a subset of items from at least one, and possibly all, of the $k$ tests. Brennan (2015) addresses the question of whether or not it is preferable to estimate true scores on $X$ using $X$ itself or using $Z$. Brennan (2015) presents the theory to obtain utility indexes for composite scores that can apply to both raw scores and scale scores. In this paper, equations in Brennan (2015) are converted into matrix form so that computation is straightforward and facilitated. This paper includes the R code for computing utility indexes for composite scores. As an example, a data set was simulated, scoring was performed both for raw scores and scale scores, and utility indexes for composite scores were computed using the R code.

Suppose there exist $k$ tests, each of which generates a score, and an additional test score based on a subset of items from at least one, and possibly all, of the $k$ tests. Let $Z$ denote a weighted composite of the full-length $k$ tests, let $X$ be the composite of subsets of items from the $k$ tests, and let $T_X$ be true score associated with $X$. Sometimes, we refer to $X$ as a pseudo test since it is not an actually administered intact test. Such subsets of items forming the composite score $X$ are typically similar in some content or construct-relevant sense, and different from the items not used for $X$. Brennan (2015) considers a procedure for determining whether or not the composite of subsets of items (i.e., $X$) provides "more useful" information about $T_X$ than that provided by the composite of full-length tests (i.e., $Z$). It is assumed here that the reader is familiar with Brennan (2015).

In Brennan's (2015) procedure, the relative utility index $U_r$ for composite scores is defined by

$$\begin{aligned} U_r &= \frac{U}{\rho_X^2} \\ &= \left[\frac{\sigma^2(X)}{\sigma^2(Z)}\right]\left[\frac{\sigma(T_X, Z)}{\sigma^2(T_X)}\right]^2, \end{aligned} \tag{1}$$

where the utility index $U$ is

$$U = \left[\frac{\sigma(T_X, Z)}{\sigma(T_X)\sigma(Z)}\right]^2. \tag{2}$$

The basic notion is that, if $U_r > 1$, then $Z$ is preferable to $X$ as an estimate of $T_X$; if $U_r \leq 1$, $X$ is preferable to $Z$ as an estimate of $T_X$. It is important to note that, since we will never need trues scores on $Z$, we will often simply use $T$ to represent $T_X$, and $E$ to represent $E_X$.

This paper considers the equations in Brennan (2015) and converts them into matrix functions so that computation is more straightforward and facilitated. A boldfaced variable (e.g., $\mathbf{Z}$) implies that the variable is in matrix form, and $\sim$ above a variable (e.g., $\tilde{Z}$) implies that the variable is associated with scale scores. A bold-faced variable with $\sim$ above implies that the variable is associated with scale scores and represented in matrix form. A boldfaced variable with $'$ in the upper right corner represents the transpose of the matrix that the boldfaced variable refers to.

# 1 Variables in Matrix Form

As considered in Brennan (2015), suppose there exist $k$ tests, each of which generates a score, and an additional test score consisting of a subset of items from the $k$ tests. For the rest of this paper, we let $Z_i$ and $X_i$ denote scores for test $i$ and scores for a subset of items on test $i$.

Let $\mathbf{Z}$ and $\mathbf{w}$ be column vectors as follows:

$$\underset{(k\times 1)}{\mathbf{Z}} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_i \\ \vdots \\ Z_k \end{bmatrix} \quad \text{and} \quad \underset{(k\times 1)}{\mathbf{w}} = \begin{bmatrix} w_1 \\ \vdots \\ w_i \\ \vdots \\ w_k \end{bmatrix}$$

where, for $1 \leq i \leq k$, $Z_i$ denotes scores for test $i$ and the $w_i$ are nominal weights ($w_i \geq 0$ for all $i$) specified by the user. The composite of full-length tests, then, can be represented as:

$$Z = w_1 Z_1 + w_2 Z_2 + \cdots + w_k Z_k = \mathbf{w}'\mathbf{Z}\,. \tag{3}$$

Similarly, let $T_i$ and $E_i$ denote true and error scores associated with $X_i$ for a subset of items on test $i$. $\mathbf{T}$, $\mathbf{E}$, and $\mathbf{X}$ can be denoted

$$\underset{(k\times 1)}{\mathbf{T}} = \begin{bmatrix} T_1 \\ \vdots \\ T_i \\ \vdots \\ T_k \end{bmatrix},\ \underset{(k\times 1)}{\mathbf{E}} = \begin{bmatrix} E_1 \\ \vdots \\ E_i \\ \vdots \\ E_k \end{bmatrix},\ \text{and}\ \underset{(k\times 1)}{\mathbf{X}} = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_k \end{bmatrix}$$

such that $\mathbf{X} = \mathbf{T} + \mathbf{E}$ (i.e., $X_i = T_i + E_i$ for $1 \leq i \leq k$). The nominal weights for $\mathbf{X}$ can also be denoted

$$\underset{(k\times 1)}{\mathbf{v}} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix},$$

and defined by the user. Similar to Equation 3, the composite of subsets of items from the $Z_i$ can be represented as:

$$X = v_1 X_1 + v_2 X_2 + \cdots + v_k X_k = \mathbf{v}'\mathbf{X}\,,$$

where the true and error scores associated with $X$ are

$$T = v_1 T_1 + v_2 T_2 + \cdots + v_k T_k = \mathbf{v}'\mathbf{T}\,,$$

and

$$E = v_1 E_1 + v_2 E_2 + \cdots + v_k E_k = \mathbf{v}'\mathbf{E}\,.$$

Let the error variance matrix for $\mathbf{X}$ be denoted

$$\underset{(k\times k)}{\mathbf{\Sigma_E}} = \begin{bmatrix} \sigma^2(E_1) & & & & 0 \\ & \ddots & & & \\ & & \sigma^2(E_i) & & \\ & & & \ddots & \\ 0 & & & & \sigma^2(E_k) \end{bmatrix},$$

where, for $1 \le i \le k$, $\sigma^2(E_i)$ stands for error variance for $X_i$. While the diagonal elements are non-zero, the off-diagonal elements are zero because the errors in $X_i$ is uncorrelated with the errors in $X_j$ for for $1 \le i, j \le k$ and $j \ne i$. The variance-covariance matrix for $\mathbf{X}$ can be denoted

$$\underset{(k\times k)}{\boldsymbol{\Sigma}_{\mathbf{X}}} = \begin{bmatrix} \sigma^2(X_1) & \sigma(X_1,X_2) & \cdots & \sigma(X_1,X_k) \\ \sigma(X_2,X_1) & \sigma^2(X_2) & \cdots & \sigma(X_2,X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(X_k,X_1) & \sigma(X_k,X_2) & \cdots & \sigma^2(X_k) \end{bmatrix},$$

where $\sigma^2(X_i)$ and $\sigma(X_i, X_j)$ for $i \ne j$ are the variance for scores on a subset of items on test $i$ and the covariance between scores on subsets of items on test $i$ and test $j$, respectively. In a similar manner, the variance-covariance matrix for $\mathbf{Z}$ and the covariance matrix between $\mathbf{Z}$ and $\mathbf{X}$ can be denoted

$$\underset{(k\times k)}{\boldsymbol{\Sigma}_{\mathbf{Z}}} = \begin{bmatrix} \sigma^2(Z_1) & \sigma(Z_1,Z_2) & \cdots & \sigma(Z_1,Z_k) \\ \sigma(Z_2,Z_1) & \sigma^2(Z_2) & \cdots & \sigma(Z_2,Z_k) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(Z_k,Z_1) & \sigma(Z_k,Z_2) & \cdots & \sigma^2(Z_k) \end{bmatrix}$$

and

$$\underset{(k\times k)}{\boldsymbol{\Sigma}_{\mathbf{ZX}}} = \begin{bmatrix} \sigma(Z_1,X_1) & \sigma(Z_1,X_2) & \cdots & \sigma(Z_1,X_k) \\ \sigma(Z_2,X_1) & \sigma(Z_2,X_2) & \cdots & \sigma(Z_2,X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(Z_k,X_1) & \sigma(Z_k,X_2) & \cdots & \sigma(Z_k,X_k) \end{bmatrix},$$

respectively.

## 2 Relative Utility Index $U_r$ for Composite Scores

The variables expressed in matrix form from Section 1 can be used to compute variances of $X$, $E$, and $T$ as follows:

$$\begin{aligned} \sigma^2(X) &= \sigma^2(v_1X_1 + v_2X_2 + \cdots + v_kX_k) \\ &= \sum_i \sum_j v_i \sigma(X_i, X_j) v_j \\ &= \mathbf{v}' \, \boldsymbol{\Sigma}_{\mathbf{X}} \, \mathbf{v}\,, \end{aligned} \tag{4}$$

$$\begin{aligned} \sigma^2(E) &= \sigma^2(v_1E_1 + v_2E_2 + \cdots + v_kE_k) \\ &= \sum_{i=1}^{k} v_i^2 \sigma^2(E_i) \\ &= \mathbf{v}' \, \boldsymbol{\Sigma}_{\mathbf{E}} \, \mathbf{v}\,, \end{aligned} \tag{5}$$

and

$$\begin{aligned} \sigma^2(T) &= \sigma^2(X) - \sigma^2(E) \\ &= \mathbf{v}'\,\mathbf{\Sigma_X}\,\mathbf{v} - \mathbf{v}'\,\mathbf{\Sigma_E}\,\mathbf{v}\,. \end{aligned} \tag{6}$$

Also,

$$\begin{aligned} \sigma^2(Z) &= \sigma^2(w_1 Z_1 + w_2 Z_2 + \cdots + w_k Z_k) \\ &= \sum_{i=1}^{k} w_i^2 \sigma^2(Z_i) + \sum\sum_{i \neq j} w_i w_j \sigma(Z_i, Z_j) \\ &= \sum_i \sum_j w_i \sigma(Z_i, Z_j) w_j \\ &= \mathbf{w}'\,\mathbf{\Sigma_Z}\,\mathbf{w}\,. \end{aligned} \tag{7}$$

A covariance between $T$ and $Z$ can also be obtained as follows:

$$\begin{aligned} \sigma(T, Z) &= \sum_{i=1}^{k}\sum_{j=1}^{k} \sigma(v_i T_i, w_j Z_j) \\ &= \sum_{i=1}^{k} \sigma(v_i T_i, w_i Z_i) + \sum\sum_{i \neq j} \sigma(v_i T_i, w_j Z_j) \\ &= \sum_{i=1}^{k} v_i w_i [\sigma(Z_i, T_i)] + \sum\sum_{i \neq j} v_i w_j \sigma(Z_j, T_i) \\ &= \sum_{i=1}^{k} v_i w_i [\sigma(Z_i, X_i) - \sigma^2(E_i)] + \sum\sum_{i \neq j} v_i w_j \sigma(Z_j, X_i) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} v_i w_j \sigma(Z_j, X_i) - \sum_{i=1}^{k} v_i w_i \sigma(E_i)^2 \\ &= \mathbf{w}'\,\mathbf{\Sigma_{ZX}}\,\mathbf{v} - \mathbf{w}'\,\mathbf{\Sigma_E}\,\mathbf{v}\,. \end{aligned} \tag{8}$$

Therefore, the product of Equations 6 and 7 gives the denominator of $U$ in Equation 2, and the square of Equation 8 gives the numerator of $U$ in Equation 2.

As a result, the relative utility index $U_r$ for composite scores becomes

$$\begin{aligned} U_r &= \left[\frac{\sigma^2(X)}{\sigma^2(Z)}\right]\left[\frac{\sigma(T, Z)}{\sigma^2(T)}\right]^2 \\ &= \left[\frac{\mathbf{v}'\,\mathbf{\Sigma_X}\,\mathbf{v}}{\mathbf{w}'\,\mathbf{\Sigma_Z}\,\mathbf{w}}\right]\left[\frac{\mathbf{w}'\,\mathbf{\Sigma_{ZX}}\,\mathbf{v} - \mathbf{w}'\,\mathbf{\Sigma_E}\,\mathbf{v}}{\mathbf{v}'\,\mathbf{\Sigma_X}\,\mathbf{v} - \mathbf{v}'\,\mathbf{\Sigma_E}\,\mathbf{v}}\right]^2, \end{aligned} \tag{9}$$

using Equations 4 through 8. If $U_r > 1$, then $Z$ is preferable to $X$ as an estimate of $T$ (or, more specifically $T_X$); if $U_r \leq 1$, then $X$ is preferable to $Z$ as an estimate of $T$.

# 3 Relative Utility $\tilde{U}_r$ for Scale Scores

Note that a $\sim$ above a variable implies that the variable is associated with scale scores. The scale-score relative utility index, $\tilde{U}_r$, has the same form as $U_r$, namely,

$$\tilde{U}_r = \frac{\tilde{U}}{\rho^2_{\tilde{X}}} = \left[\frac{\sigma^2(\tilde{X})}{\sigma^2(\tilde{Z})}\right]\left[\frac{\sigma(\tilde{T},\tilde{Z})}{\sigma^2(\tilde{T})}\right]^2.$$

## 3.1 Matrix Notation

Let

$$\underset{(k\times 1)}{\tilde{\mathbf{Z}}} = \begin{bmatrix} \tilde{Z}_1 \\ \tilde{Z}_2 \\ \vdots \\ \tilde{Z}_k \end{bmatrix}$$

where $\tilde{Z}_j$ designates the scale-score version of $Z_j$. The composite scale score of full-length tests can be represented as:

$$\tilde{Z} = w_1\tilde{Z}_1 + w_2\tilde{Z}_2 + \cdots + w_k\tilde{Z}_k = \mathbf{w}'\tilde{\mathbf{Z}}\,.$$

Similarly, let $\tilde{X}_i$ denote the scale-score version of $X_i$ and let $\tilde{T}_i$ and $\tilde{E}_i$ be true and error scores associated with $\tilde{X}_i$. Then, $\tilde{\mathbf{T}}$, $\tilde{\mathbf{E}}$, and $\tilde{\mathbf{X}}$ can be denoted

$$\underset{(k\times 1)}{\tilde{\mathbf{T}}} = \begin{bmatrix} \tilde{T}_1 \\ \tilde{T}_2 \\ \vdots \\ \tilde{T}_k \end{bmatrix}, \quad \underset{(k\times 1)}{\tilde{\mathbf{E}}} = \begin{bmatrix} \tilde{E}_1 \\ \tilde{E}_2 \\ \vdots \\ \tilde{E}_k \end{bmatrix}, \text{ and } \underset{(k\times 1)}{\tilde{\mathbf{X}}} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_k \end{bmatrix}$$

such that $\tilde{\mathbf{X}} = \tilde{\mathbf{T}} + \tilde{\mathbf{E}}$. The composite scale score for subsets of items from the $\tilde{Z}_i$ is

$$\tilde{X} = \mathbf{v}'\tilde{\mathbf{X}}\,,$$

where the true and error scores associated with $\tilde{X}$ are

$$\tilde{T} = \mathbf{v}'\tilde{\mathbf{T}} \quad \text{and} \quad \tilde{E} = \mathbf{v}'\tilde{\mathbf{E}}\,.$$

## 3.2 Covariance of $\tilde{Z}$ and $\tilde{X}$

According to Brennan (2015), it is assumed that we have raw-to-scale-score transformations for each of the $Z_i$ and for $X$, but we do not have raw-to-scale-score transformations for the $X_i$. In other words, $\tilde{X}$ and $\tilde{Z}_i$ for $1 \le i \le k$ are observable, whereas the $\tilde{X}_i$ are not. Since $\tilde{X}$ and $\tilde{Z}_i$ are available, the error variance for $\tilde{X}$ as well as covariances between

$\tilde{X}$ and $\tilde{Z}_i$ can be computed directly from observable data. Let the covariance between $\tilde{X}$ and $\tilde{\mathbf{Z}}$ be defined as

$$\underset{(k\times 1)}{\boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}\tilde{X}}} = \begin{bmatrix} \sigma(\tilde{Z}_1, \tilde{X}) \\ \vdots \\ \sigma(\tilde{Z}_k, \tilde{X}) \end{bmatrix}.$$

The covariance between $\tilde{Z}$ and $\tilde{X}$ is:

$$\begin{aligned} \sigma(\tilde{Z}, \tilde{X}) &= \sum_{i=1}^{k} w_i \sigma(\tilde{Z}_i, \tilde{X}) \\ &= \mathbf{w}'\boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}\tilde{X}} \,. \end{aligned}$$

Similar to Equation 7 for $\sigma^2(Z)$, the variance of $\tilde{Z}$ can be computed as follows:

$$\begin{aligned} \sigma^2(\tilde{Z}) &= \sigma^2(w_1\tilde{Z}_1 + w_2\tilde{Z}_2 + \cdots + w_k\tilde{Z}_k) \\ &= \sum_i \sum_j w_i \sigma(\tilde{Z}_i, \tilde{Z}_j) w_j \\ &= \mathbf{w}' \, \boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}} \, \mathbf{w} \end{aligned}$$

where $\boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}}$ is defined as

$$\underset{(k\times k)}{\boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}}} = \begin{bmatrix} \sigma^2(\tilde{Z}_1) & \sigma(\tilde{Z}_1, \tilde{Z}_2) & \cdots & \sigma(\tilde{Z}_1, \tilde{Z}_k) \\ \sigma(\tilde{Z}_2, \tilde{Z}_1) & \sigma^2(\tilde{Z}_2) & \cdots & \sigma(\tilde{Z}_2, \tilde{Z}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(\tilde{Z}_k, \tilde{Z}_1) & \sigma(\tilde{Z}_k, \tilde{Z}_2) & \cdots & \sigma^2(\tilde{Z}_k) \end{bmatrix},$$

and $\sigma^2(\tilde{Z}_i)$ is the variance for scale scores for test $i$ and $\sigma(\tilde{Z}_i, \tilde{Z}_j)$ for $i \neq j$ is the covariance between scale scores for test $i$ and test $j$.

### 3.3 Estimating $\sigma^2(\tilde{E}_i)$ and $\sigma(\tilde{T}, \tilde{Z})$

Let the error variance for $\tilde{\mathbf{X}}$ be notated as

$$\underset{(k\times k)}{\boldsymbol{\Sigma}_{\tilde{\mathbf{E}}}} = \begin{bmatrix} \sigma^2(\tilde{E}_1) & & 0 \\ & \ddots & \\ 0 & & \sigma^2(\tilde{E}_k) \end{bmatrix}.$$

However, since $\tilde{X}_i$, $\tilde{T}_i$, and $\tilde{E}_i$ for all $1 \leq i \leq k$ are not available for a given set of data, error variances for $\tilde{X}_i$ cannot be directly computed. In order to estimate $\boldsymbol{\Sigma}_{\tilde{\mathbf{E}}}$, Brennan (2015) suggests assuming that error variances for raw and scale scores are proportional in the sense that

$$\frac{\sigma^2(\tilde{E}_i)}{\sigma^2(\tilde{E})} = \frac{\sigma^2(E_i)}{\sigma^2(E)}\,,$$

which implies that

$$\sigma^2(\tilde{E}_i) = \frac{\sigma^2(E_i)}{\sigma^2(E)}\sigma^2(\tilde{E})\,. \tag{10}$$

As a result, using Equation 25 in Brennan (2015), the covariance between $\tilde{T}$ and $\tilde{Z}$ can be computed as:

$$\begin{aligned}\sigma(\tilde{T},\tilde{Z}) &= \sum_{i=1}^{k} w_i\sigma(\tilde{Z}_i,\tilde{X}) - \frac{\sigma^2(\tilde{E})}{\sigma^2(E)}\sum_{i=1}^{k} w_i v_i \sigma^2(E_i)\\ &= \mathbf{w}'\mathbf{\Sigma}_{\tilde{\mathbf{Z}}\tilde{X}} - \frac{\sigma^2(\tilde{E})}{\sigma^2(E)}\mathbf{w}'\mathbf{\Sigma}_{\mathbf{E}}\mathbf{v}\,. \end{aligned}\tag{11}$$

Therefore, the relative utility index $\tilde{U}_r$ for composite scale scores becomes

$$\begin{aligned}\tilde{U}_r &= \left[\frac{\sigma^2(\tilde{X})}{\sigma^2(\tilde{Z})}\right]\left[\frac{\sigma(\tilde{T},\tilde{Z})}{\sigma^2(\tilde{T})}\right]^2\\ &= \left[\frac{\sigma^2(\tilde{X})}{\sigma^2(\tilde{Z})}\right]\left[\frac{\mathbf{w}'\mathbf{\Sigma}_{\tilde{\mathbf{Z}}\tilde{X}} - \frac{\sigma^2(\tilde{E})}{\sigma^2(E)}\mathbf{w}'\mathbf{\Sigma}_{\mathbf{E}}\mathbf{v}}{\sigma^2(\tilde{X}) - \sigma^2(\tilde{E})}\right]^2, \end{aligned}\tag{12}$$

for which $\sigma^2(\tilde{X})$ and $\sigma^2(\tilde{Z})$ can be computed directly from a given data. A detailed description about estimating $\sigma^2(\tilde{E})$ is provided in Section 3.3 in Brennan (2015).

# 4 Example

This paper includes the R code for computing the relative utility index for composite scores using the matrix formulations provided defined above. The R code and examples are freely available on `www.education.uiowa.edu/casma/publications-data-file`. In order to demonstrate how the R code works, data was simulated. As in Brennan (2015), suppose there were three full-length tests, $Z_1$, $Z_2$, and $Z_3$, with 30, 40, and 50 dichotomous items, respectively. Responses were generated for 5000 examinees using the 3PL IRT Model. Scoring was conducted based on observed number-correct scores on items.

For the example in Sections 4.1 and 4.2, one composite score $Z$ was formed based on the three full-length test scores such that $Z = Z_1 + Z_2 + Z_3$. Additionally, two composite scores (i.e., $X$ and $Y$) were formed based on subsets of items from the three tests such that $X = X_1 + X_2 + X_3$ and $Y = Y_1 + Y_2 + Y_3$. There were a total of 35 items contributing to $X$ and $Y$. Note that the item sets for $Y$ are different from the item sets for $X$.

## 4.1 Number-correct Scores on Items

Using the simulated data, the variance-covariance matrix for $\mathbf{Z}$ (i.e., $\mathbf{\Sigma}_{\mathbf{Z}}$) was computed. For composite score $X$, the error variance and variance-covariance matrices for scores

Table 4.1: Variance-covariance Matrices for Number-correct Scores on Items

| | | |
|---|---|---|
| Test Scores $\mathbf{Z}$ | $\boldsymbol{\Sigma}_{\mathbf{Z}} = \begin{bmatrix} 14.5404 & 13.6433 & 20.0893 \\ 13.6433 & 24.8964 & 29.4443 \\ 20.0893 & 29.4443 & 51.8031 \end{bmatrix}$ | |
| Scores on Subsets $\mathbf{X}$ | $\boldsymbol{\Sigma}_{\mathbf{X}} = \begin{bmatrix} 2.6658 & 0.8715 & 1.4879 \\ 0.8715 & 2.1428 & 2.6820 \\ 1.4879 & 2.6820 & 4.4048 \end{bmatrix}$ | $\boldsymbol{\Sigma}_{\mathbf{ZX}} = \begin{bmatrix} 4.8490 & 4.0370 & 5.8870 \\ 4.5475 & 3.7230 & 7.1699 \\ 6.5921 & 5.6345 & 9.5843 \end{bmatrix}$ |
| Scores on Subsets $\mathbf{Y}$ | $\boldsymbol{\Sigma}_{\mathbf{Y}} = \begin{bmatrix} 3.4031 & 1.3182 & 1.8292 \\ 1.3182 & 2.7560 & 3.1278 \\ 1.8292 & 3.1278 & 4.1124 \end{bmatrix}$ | $\boldsymbol{\Sigma}_{\mathbf{ZY}} = \begin{bmatrix} 5.7311 & 4.2141 & 5.3526 \\ 5.6931 & 5.2791 & 7.4467 \\ 8.5346 & 6.7116 & 9.4478 \end{bmatrix}$ |

Table 4.2: KR-20 Reliability and Error Variance for Scores on Subsets of Items

| | Composite $X$ | | | Composite $Y$ | | |
|---|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | $X_3$ | $Y_1$ | $Y_2$ | $Y_3$ |
| KR-20 | 0.3862 | 0.2940 | 0.4562 | 0.4668 | 0.4231 | 0.4921 |
| Error Variance Matrix | $\boldsymbol{\Sigma}_{\mathbf{E_X}} = \begin{bmatrix} 1.6361 & 0 & 0 \\ 0 & 1.5128 & 0 \\ 0 & 0 & 2.3952 \end{bmatrix}$ | | | $\boldsymbol{\Sigma}_{\mathbf{E_Y}} = \begin{bmatrix} 1.8145 & 0 & 0 \\ 0 & 1.5898 & 0 \\ 0 & 0 & 2.0888 \end{bmatrix}$ | | |

on the three subsets of items were obtained and denoted $\boldsymbol{\Sigma}_{\mathbf{E_X}}$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$, respectively. In order to obtain error variances, KR-20 was computed as a measure of reliability for each score for the subsets of items. Error variances were obtained as follows:

$$\sigma^2_{E_i} = \sigma^2_{X_i}(1 - \rho^2_i)$$

where $\sigma^2_{E_i}$, $\sigma^2_{X_i}$, and $\rho^2_i$ represent error variance, observed score variance, and reliability, respectively. Similarly, for composite score $Y$, the error variance and variance-covariance matrices for scores on the three subsets of items were also obtained and denoted $\boldsymbol{\Sigma}_{\mathbf{E_Y}}$ and $\boldsymbol{\Sigma}_{\mathbf{Y}}$, respectively. Covariance matrices between the $Z_i$ test scores and scores on subsets of items for composite scores (i.e., $\boldsymbol{\Sigma}_{\mathbf{ZX}}$ and $\boldsymbol{\Sigma}_{\mathbf{ZY}}$) were also computed. Table 4.1 shows the variance and variance-covariance matrices for the given simulated data. For raw scores, the R code requires the matrices in Table 4.1 and the reliability estimates in Table 4.2. Table 4.2 shows the KR-20 reliability and error variances for scores on subsets of items for both composite scores $X$ and $Y$. Note that the error variances are automatically computed within the R code, but are not produced as output.

Table 4.3: Variance-covariance Matrices for Scale Scores

| | Scale Scores for Three Full-length Tests | |
|---|---|---|
| Variance-covariance Matrix | $\boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}} = \begin{bmatrix} 14.5404 & 14.3047 & 13.6433 \\ 14.3047 & 14.0967 & 13.4107 \\ 13.6433 & 13.4107 & 24.8964 \end{bmatrix}$ | |
| | Scale Scores for Two Composite Tests | |
| | Scale Scores $\tilde{X}$ | Scale Scores $\tilde{Y}$ |
| Error Variance | 3.61 | 4.41 |
| Covariances of $\tilde{Z}_i$ with $\tilde{X}$ and $\tilde{Y}$ | $\boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}\tilde{X}} = \begin{bmatrix} 10.0818 \\ 9.9234 \\ 18.2451 \end{bmatrix}$ | $\boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}\tilde{Y}} = \begin{bmatrix} 20.0893 \\ 19.6907 \\ 29.4443 \end{bmatrix}$ |

## 4.2 Scale Scores

In order to demonstrate how the R code can be used for scale scores, artificial raw-to-scale conversion tables were created for raw scores $Z_1$, $Z_2$, $Z_3$, $X$, and $Y$. The arcsine transformation was used to stabilize the magnitude of the conditional standard error of measurement for raw score points. A detailed description can be found in Kolen and Brennan (2014, chap. 9). For five test scores, raw-to-scale conversion tables were created so that rounded scale scores range from 10 to 50 in increments of one. These computations resulted in $\tilde{Z}_1$, $\tilde{Z}_2$, $\tilde{Z}_3$, $\tilde{X}$, and $\tilde{Y}$ have conditional standard errors of measurement of 2.2. 1.8, 1.7, 1.9, and 2.1, respectively.

Table 4.3 shows the variance-covariance matrix for scale scores for the three full-length tests as well as the covariances of $\tilde{Z}_i$ with $\tilde{X}$ and $\tilde{Y}$. Since the conditional standard error of measurement has been standardized, the error variances for $\tilde{X}$ and $\tilde{Y}$ are simply the square of 1.9 and 2.1, respectively.

## 4.3 Results

For the three full-length tests, the simulation study considers six sets of weights for the $w_i$ as illustrated in Brennan (2015). The first four columns in Table 4.4 present these weights. For $X$ and $Y$, the $v_i$ weights were set to one for all scores on subsets of items from the three tests. Table 4.4 presents the relative utility indexes for composite scores using number-correct scores on items. All computations were performed using the R code provided in Appendix A. Additionally, Appendix B provides the variance and variance-covariance matrices as well as the KR-20 reliabilities for the given example. The reader can run the R code in Appendix B and confirm the results in Table 4.4. Note that the functions in Appendix A should be defined prior to running the R code in Appendix B.

Table 4.4 presents the relative utility indexes for both number-correct scores on items and scale scores for $X$ and $Y$ for all sets of weights. When number-correct scores on

Table 4.4: Results for Composite Score Utility Index

| | Full-length Tests | | | Composite $X$ | | Composite $Y$ | |
|---|---|---|---|---|---|---|---|
| | $Z_1$ | $Z_2$ | $Z_3$ | $U_r$ | $\tilde{U}_r$ | $U_r$ | $\tilde{U}_r$ |
| weight $\mathbf{w}$ | 1/3 | 1/3 | 1/3 | 1.0130 | 1.2303 | 0.9781 | 0.7111 |
| | 1/4 | 1/4 | 1/2 | 0.9385 | 1.3884 | 0.9183 | 0.7459 |
| | 1/2 | 1/2 | 0 | 1.1201 | 0.7867 | 1.0466 | 0.5529 |
| | 1 | 0 | 0 | 1.2110 | 0.7810 | 0.9502 | 0.5507 |
| | 0 | 1 | 0 | 0.7950 | 0.7917 | 0.8646 | 0.5548 |
| | 0 | 0 | 1 | 0.7425 | 1.5621 | 0.7497 | 0.7143 |

items are considered for the composite $X$, $X$ is preferable to $Z$ as an estimate of $T_X$ for the second, fifth, and sixth sets of weights. More often than not, $X$ provides more information than that provided by $Z$ when $Z_1$ is not included in forming the composite $Z$ (i.e., $w_1 = 0$). When scale scores are considered for the composite $X$, different results were obtained: $X$ provides more information than $Z$ does when $Z_3$ is not considered in forming the composite $Z$ (i.e., $w_3 = 0$).

For the composite $Y$ number-correct scores, there is only one set of weights for which $Z$ is preferable to $Y$ as an estimate of $T_Y$: that is $(1/2, 1/2, 0)$. For other sets of weights, $Y$ provides more information than that provided by $Z$ no matter which score type (i.e., number-correct scores, or scale score) was considered.

## 5 Comments

The discussion in this paper has been intensive in terms of matrix notation and operations. Most importantly, since dimensions of matrices are crucial in matrix multiplication, the user should be very careful when setting up variance and covariances matrices for computation.

## 6 References

Brennan, R. L. (2015, May). *Utility Indexes for Composite Scores.* (CASMA Research Report No. 42). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa. (Available on http://www.education.uiowa.edu/casma)

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York: Springer.

# 7 APPENDIX A: R Code

```
##  According to Brennan(2015),
##        Z = w_1*Z_1 + .... + w_k*Z_k       (eq. 3)
##        X = v_1*X_1 + .... + v_k*X_k       (eq. 4)
##        SZ = w_1*SZ_1 + .... + w_k*SZ_k
##        SX = v*_1*SX_1 + .... + v*_k*SX_k
##
##  where Z_i   denotes raw scores for test i,
##
##        X_i   stands for raw scores for the subset of items
##              for test i,
##
##        SZ_i  stands for scales scores for test i,
##
##        SX_i  stands for scale scores for the subset of
##              items for test i,
##
##        Z     denotes composite scores for all k tests,
##
##        X     denotes composite scores for scores for the
##              subsets of items for all k tests,
##
##        SZ    stands for composite scale scores for all k
##              tests; and,
##
##        SX    stands for composite scale scores for scores
##              for the subset of items for all k tests.
##
##  Unless stated otherwise, all equation numbers in for Brennan (2015).
```

```
## Composite.Utility function computes a composite utility and relative utility
#  indexes (U and Ur, respectively) reflecting the merits of using Z rather
#  than X.
#
## Input:
#
## 1. w                a list of vectors of weights for Z_1, ..., Z_k
#                      (or pseudo Y_1, ..., Y_k)
#                              w = (w_1, ..., w_k)
#
## 2. v                a vector of weights for X_1, ..., X_k
#                              v = (v_1, ..., v_k)
#
## 3. Z.cov.mat        a (k x k) covariance matrix for raw scores
#                      on k tests
#                                    Z_1 Z_2 ... Z_k
#                              Z_1 |                 |
#                              Z_2 |                 |
#                               :  |                 |
#                              Z_k |                 |
#
## 4. X.cov.mat        a (k x k) variance-covariance matrix for raw scores
#                      on the k subsets of items, each subset for each
#                      test
#                                    X_1 X_2 ... X_k
#                              X_1 |                 |
#                              X_2 |                 |
#                               :  |                 |
#                              X_k |                 |
#
## 5. Z_X.cov.mat      a (k x k) covariance matrix between raw scores
#                      for k tests and raw scores for subset of items for
#                      all k tests
#                                    X_1 X_2 ... X_k
#                              Z_1 |                 |
#                              Z_2 |                 |
#                               :  |                 |
#                              Z_k |                 |
#
## 6. XZ.rel           a vector of reliability coefficients for X_1, ..., X_k
#                              X.rel = (r_1, ..., r_k)
#
## 7. SZ.cov.mat       a (k x k) covariance matrix for scales scores
#                      on k tests
#                                    SZ_1 SZ_2 ... SZ_k
#                             SZ_1 |                    |
#                             SZ_2 |                    |
#                               :  |                    |
#                             SZ_k |                    |
```

```
#
## 8. SX.var          a variance of composite scale scores for the subset
#                     of items for all k tests
#
## 9. SZ_SX.cov.mat   a (k x 1) covariance matrix between scale scores
#                     for k tests and a composite scale score for
#                     subset of items for all k tests
#                                  SX
#                           SZ_1 |    |
#                           SZ_2 |    |
#                             :  |    |
#                           SZ_k |    |
#
## 10. SX.error.var   error variance for scale score X
#
## 11. Scale          "Scale" to compute utilities for scale scores and
#                     "Raw" to compute utilities for raw scores
#
## Output:
#
#  Utility Matrix     a (n x 2) matrix where n refers to the number of w
#                     weights. The first and second columns give utilities
#                     and relative utilities for corresponding w weights,
#                     respectively.
#                                   U   Ur
#                            w_1 |          |
#                            w_2 |          |
#                             :  |          |
#                            w_n |          |
#
#                     i.  Ur > 1: Z (or pseudo Y) is preferable to X as
#                         an estimate of Tx
#                     ii. Ur < 1: X is preferable to Z (or pseudo Y) as
#                         an estimate of Tx
```

```
Composite.Utility <- function(w, v, Z.cov.mat, X.cov.mat, Z_X.cov.mat, X.rel,
                             SZ.cov.mat, SX.var, SZ_SX.cov.mat,
                             SX.error.var, Scale){
    if (Scale == "Scale") {
        # For scale scores, Scale.Utility function is called
        Scale.Utility(w, v, SZ.cov.mat, SX.var, SZ_SX.cov.mat,
                         SX.error.var, X.cov.mat, X.rel)
    } else if (Scale == "Raw") {
        # For raw scores, Raw.Utility function is called
        Raw.Utility(w, v, Z.cov.mat, X.cov.mat, Z_X.cov.mat, X.rel)
    }
}
```

```
Raw.Utility <- function(w, v, Z.cov.mat, X.cov.mat, Z_X.cov.mat, X.rel) {

    nX <- ncol(X.cov.mat)  # number of X_i's
    nZ <- ncol(Z.cov.mat)  # number of Z_j's
    nw <- length(w)        # number of w weights considered for Z

    X_E2.mat <- matrix(0, nrow = nX, ncol = nX)
        # a matrix storing error variances for X_i's
    utility.mat <- matrix(nrow = nw, ncol = 2)
        # a matrix storing utility and relative utility for each w weight
        # 1st column: utility
        # 2nd column: relative utility

    j <- 1
    for (j in 1:nX) {
        X_E2.mat[j,j] <- X.cov.mat[j,j]*(1-X.rel[j])
            # calculates error variances
            # for X_i's and stores them in X_E2 matrix
            # off-diagonal entries are zero
            # (i.e., cov(E_i, E_j) = 0 for i!= j)
        j <- j + 1
    }

    X.var <- t(v) %*% X.cov.mat %*% t(t(v))
        # calculates a variance of X considering v weight (eq.14)
    X_T.var <- X.var - t(v) %*% X_E2.mat %*% t(t(v))
        # calculates a variance of true scores for X considering v weight (eq. 16)

    i <- 1
    for (i in 1:length(w)) {
        Z.var <- t(w[[i]]) %*% Z.cov.mat %*% t(t(w[[i]]))
            # calculates a variance of Z considering w weight (eq. 13)

        Z.X_T.cov <- t(w[[i]]) %*% Z_X.cov.mat %*% t(t(v)) -
                     t(w[[i]]) %*% X_E2.mat %*% t(t(v))
            # calculates a covariance between Z
            # and true scores for X considering v and w weights  (eq. 11)

        utility.mat[i,1] <- (Z.X_T.cov/sqrt(Z.var)/sqrt(X_T.var))^2
            # Utility U (eq. 2)
        utility.mat[i,2] <- (X.var/Z.var)*(Z.X_T.cov/X_T.var)^2
            # Relative Utility Ur (eq. 18)

        i = i + 1
    }
    utility.mat
}
```

```
Scale.Utility = function(w, v, SZ.cov.mat, SX.var, SZ_SX.cov.mat,
                         SX.E2, X.cov.mat, X.rel) {

    nX = length(X.rel)
        # number of X_i's
    X_E2.mat = matrix(0, nrow = nX, ncol = nX)
        # a matrix storing error variances for X_i's
    utility.mat = matrix(nrow = length(w), ncol = 2)
        # a matrix storing utility and relative utility for each w weight
        # 1st column: utility
        # 2nd column: relative utility

    j = 1
    for (j in 1:nX) {
        X_E2.mat[j,j] = X.cov.mat[j,j]*(1-X.rel[j])
            # calculates error variances for X_i's
            # and stores them in X_E2 matrix
            # off-diagonal entries are zero
            # (i.e., cov(E_i, E_j) = 0 for i!= j)
        j = j + 1
    }
    X_E2 = t(v) %*% X_E2.mat %*% t(t(v)) # error variance of X

    SX_T.var = SX.var - SX.E2   # variance of true scores of SX (eq. 21)

    i = 1
    for (i in 1:length(w)) {
        SZ.var = t(w[[i]]) %*% SZ.cov.mat %*% t(t(w[[i]]))
            # calculates a variance of SZ considering w weight
            # (modified eq. 13)

        SZ_SX.cov = t(w[[i]]) %*% SZ_SX.cov.mat
            # covariance between Total SZ and SX

        SZ.SX_T.cov = SZ_SX.cov -
            SX.E2/X_E2 * (t(w[[i]]) %*% X_E2.mat %*% t(t(v)))
            # covariance between SZ and true SX (eq. 26)

        utility.mat[i,1] = SZ.SX_T.cov^2/(SZ.var*SX_T.var)
            # Utility U (eq. 2)
        utility.mat[i,2] = (SX.var/SZ.var)*(SZ.SX_T.cov/SX_T.var)^2
            # Relative Utility Ur (eq. 20)

        i = i + 1
    }
    utility.mat
}
```

# 8 APPENDIX B: Example Using R Code

```
##### Number-correct Scores on Items #####

# Variance-covariance Matrix of (Z1, Z2, Z3)
Z.cov.mat = matrix(c(14.54041464, 13.64333187, 20.08933851,
                     13.64333187, 24.89638328, 29.44426965,
                     20.08933851, 29.44426965, 51.80307678),
                   nrow = 3, ncol = 3, byrow=TRUE)

# Variance-covariance Matrix of (X1, X2, X3)
X.cov.mat = matrix(c(2.6657795159, 0.8715068214, 1.487939228,
                     0.8715068214, 2.1428122024, 2.682019884,
                     1.4879392278, 2.6820198840, 4.404751310),
                   nrow = 3, ncol = 3, byrow=TRUE)

# Variance-covariance Matrix between (Z1, Z2, Z3) and (X1, X2, X3)
Z_X.cov.mat = matrix(c(4.849008842, 4.036992679, 5.886970354,
                       4.547488298, 3.723036207, 7.169895179,
                       6.592124185, 5.634495219, 9.584319104),
                     nrow = 3, ncol = 3, byrow=TRUE)

# Variance-covariance Matrix of (Y1, Y2, Y3)
Y.cov.mat = matrix(c(3.403096619, 1.318182036, 1.829201040,
                     1.318182036, 2.755994359, 3.127752030,
                     1.829201040, 3.127752030, 4.112367914),
                   nrow = 3, ncol = 3, byrow=TRUE)

# Variance-covariance Matrix between (Z1, Z2, Z3) and (Y1, Y2, Y3)
Z_Y.cov.mat = matrix(c(5.731099820, 4.214067453, 5.352632446,
                       5.693130626, 5.279066613, 7.446731746,
                       8.534585317, 6.711634487, 9.447838048),
                     nrow = 3, ncol = 3, byrow=TRUE)

X.var = 19.29627489    # Variance of X
Y.var = 22.82172911    # Variance of Y
X.rel = 0.7176411267   # Reliability of X
Y.rel = 0.763122936    # Reliability of Y

# Reliability for (X1, X2, X3)
XZ.rel = c(0.3862486461, 0.2940013553, 0.4562130603)
# Reliability for (Y1, Y2, Y3)
YZ.rel = c(0.4667870404, 0.4231401335, 0.4920775381)
```

```
##### Scale Scores #####

# Variance-covariance Matrix for Scale Scores of (Z1, Z2, Z3)
SZ.cov.mat = matrix(c(14.54041464, 14.30470470, 13.64333187,
                      14.30470470, 14.09673251, 13.41066472,
                      13.64333187, 13.41066472, 24.89638328),
                   nrow = 3, ncol = 3, byrow=TRUE)

# Variance-covariance Matrix between
# Scale Scores of (Z1, Z2, Z3) and Scale Scores of X
SZ_SX.cov.mat = matrix(c(10.081769759, 9.923449195, 18.245124597),
                       nrow = 3, ncol = 1, byrow=TRUE)

# Variance-covariance Matrix between
# Scale Scores of (Z1, Z2, Z3) and Scale Scores of Y
SZ_SY.cov.mat = matrix(c(20.08933851, 19.69073338, 29.44426965),
                      nrow = 3, ncol = 1, byrow=TRUE)

# Variance of Scale Scores of X
SX.var = 13.40639402

# Variance of Scale Scores of Y
SY.var = 51.80307678

# Standard Error of Measurement for Scales Scores of X and Y.
SX.sem = 1.9
SY.sem = 2.1

w1 = c(1/3, 1/3, 1/3)
w2 = c(1/4, 1/4, 1/2)
w3 = c(1/2, 1/2, 0)
w4 = c(1, 0, 0)
w5 = c(0, 1, 0)
w6 = c(0, 0, 1)

w = list(w1, w2, w3, w4, w5, w6)
v = c(1, 1, 1)
```

```
##### Number-correct Scores on Items #####

Composite.Utility(w, v, Z.cov.mat, X.cov.mat, Z_X.cov.mat, XZ.rel,
                  NULL, NULL, NULL, NULL, "Raw")

Composite.Utility(w, v, Z.cov.mat, Y.cov.mat, Z_Y.cov.mat, YZ.rel,
                  NULL, NULL, NULL, NULL, "Raw")

##### Scale Scores #####

Composite.Utility(w, v, Z.cov.mat, X.cov.mat, Z_X.cov.mat, XZ.rel,
                  SZ.cov.mat, SX.var, SZ_SX.cov.mat, SX.sem^2, "Scale")

Composite.Utility(w, v, Z.cov.mat, Y.cov.mat, Z_Y.cov.mat, YZ.rel,
                  SZ.cov.mat, SY.var, SZ_SY.cov.mat, SY.sem^2, "Scale")
```