If you have issues viewing or accessing this file contact us at NCJRS.gov.

National Criminal Justice Reference Service

ncjrs

This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504.

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U. S. Department of Justice.

National Institute of Justice
United States Department of Justice
Washington, D.C. 20531

5/16/84

A MEASUREMENT MODEL APPROACH TO THE CONSTRUCTION OF SIMPLE LINEAR COMPOSITE MEASURES

Terry Connolly
Georgia Institute of Technology
August, 1979

92505

NCJRS

ACQUISITIONS

Work in this area was supported, in part, by Grant 78-NI-AX-0003 from the National Institute of Law Enforcement and Criminal Justice, Principal Investigator: S. J. Deutsch. The opinions expressed are those of the author and do not necessarily represent the views or policies of the U.S. Department of Justice.

U.S. Department of Justice
National Institute of Justice

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this copyrighted material has been granted by
Public Domain/LEAA
U.S. Department of Justice
to the National Criminal Justice Reference Service (NCJRS)

Further reproduction outside of the NCJRS system requires permission of the copyright owner.

92505

Interest in the construction of composite measures arises in the following familiar situation: The investigator is interested in some variable or construct that is not directly measurable without error. A number of indicators or measures are available, each tapping the variable of interest in a somewhat imperfect way. The investigator wishes to combine these several indicators into a single 'composite measure' that will reflect the underlying variable better than does any one of the indicators taken alone. The paper considers several alternative approaches to the construction of such composites, and assesses the adequacy of the resulting measures.

We shall not be concerned here with approaches to the development of such multiple indicators. (A general rationale, and many ingenious examples, of such development are given in Webb, Campbell, Schwartz and Sechrest, 1966). Nor shall we consider the practical problems involved in sampling, estimation, and testing of particular models in particular measurement settings. Rather, we shall assume complete knowledge of the relevant structure of a measurement process, and examine the performance of several alternative strategies of composite measure construction as a function of the parameters of this structure. The aim is to suggest the circumstances under which the effort involved in devising optimal composite measures, rather than using standard rules of thumb, might be justified by the resulting improvement in measurement quality. That is, the emphasis of the paper is less on the practical steps involved in devising an optimal measurement procedure in any particular case than on the conditions under which such an effort might be justified.

In the psychological literature, interest in the relative efficacy of alternative linear composites as either estimators or predictors has been recently revived by a paper by Wainer (1976). Wainer proposes an 'equal weights theorem' that sets a surprisingly small upper bound on the expected loss of explained variance when optimal weights are replaced by equal weights in linear prediction equations. Though the theorem has since been shown to underestimate the loss of explained variance by a factor of two (Laughlin, 1978), and to be of considerably less generality than originally claimed (Pruzek and Frederick, 1978), Wainer (1978) remains essentially unmoved from his original contention that the coefficients in linear models generally "don't make no nevermind," and that the simplest (i.e. equal) weights are almost always adequate.

The debate has a lengthy history in both methodological and substantive literatures. Wilks (1938) showed that, for sufficiently large numbers of predictors, and under reasonably general conditions, predictions from a linear composite are insensitive to changes in weighting schemes. Similar conclusions are reached by Burt (1950) and Gulliksen (1950), the latter suggesting that differential weighting schemes be considered only when the number of measures to be combined is less than ten. Even when the number of measures or predictors available is small, equal weights may produce composites as good as or better than more complex approaches in a variety of practical situations (Wesman and Bennett, 1959; Lawshe and Schucker, 1959; Schmidt, 1959; Dawes and Corrigan, 1974). Given the simplicity of 'unit weights', their independence of data and estimation problems, and the apparent robustness of the resulting composites, their use across a wide range of decision making settings has been advocated (Einhorn and Hogarth, 1975; Dawes, 1979).

Against this impressive body of data and argument, most investigators

faced with a multiple-indicator problem will presumably feel justified in simply adding together their several measures, perhaps after standardization, into a simple composite. As the present paper attempts to show, however, the procedure remains a hazardous one. The resulting composites may well be excellent measures of the variable of interest. On the other hand, they may be completely misleading. Crucially, no evaluation of measure quality, and no procedure for refinement, is possible if equal weighting is routinely used.

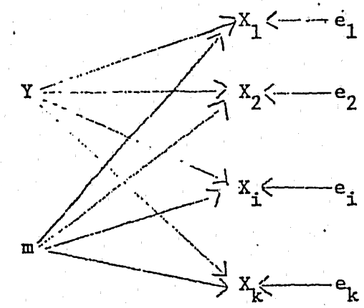
Both evaluation and refinement require the specification of an explicit measurement model. As Blalock has argued: ". . . imperfections in our measuring instruments and the necessity of indirect measurement will require that rather complex auxiliary theories be formulated so that we may pass back and forth between our conceptual variables and operational measures with minimal confusion" (Blalock, 1974: 6). In the present case, the auxiliary theory is, fortunately, not particularly complex. Nor is any great mathematical sophistication required for its analysis, beyond simple algebra. The present paper proposes a simple model of a reasonably general multiple-indicator situation, and explores the consequences of forming composites with different weighting schemes. A number of currently widespread practices and beliefs are examined, and some tentative generalizations offered. The main hope, however, is that investigators will find the approach useful in specifying and testing measurement models for particular measurement situations of substantive interest; and that the analytic approach developed here will be of value to them in assessing their present measures, and in developing better ones.

I. A GENERAL MEASUREMENT MODEL

A measure or indicator is, in general, an accessible covariate of the variable of interest. Scores obtained on a particular I.Q. test are correlated with the true intelligence of the individuals taking the test. Length measurements obtained from a given instrument are correlated with the actual length of the objects they measure. Responses to a questionnaire assessing job satiffaction are correlated with the actual satisfaction of the workers surveyed. In general, then, measurements obtained from alternative measurement procedures will also be intercorrelated, since each taps, to some extent, the same underlying variable. (Such intercorrelation of measures is noted as necessary though not sufficient for 'convergent validity' of the measurement procedure: Campbell and Fiske, 1959).

Covariation between measures is, however, not sufficient evidence to establish whether, or to what extent, the several measures are associated with the underlying variable of interest. Covariance with some other, irrelevant variable is clearly an alternative possibility. A particularly troublesome source of such irrelevant covariation is what might be termed a 'method' variable. For example, two different I.Q. tests might both involve the reading of instructions; differences between subjects in reading skills would thus contribute to the covariation in obtained scores between the two tests, over and above that resulting from both tapping the underlying variable 'true intelligence'. Similarly, different experts might agree for reasons unrelated to the extent to which each independently taps the underlying variable (s)he is asked to assess.

In this paper, we shall consider the following simple linear measurement model:



where:

Y is the underlying variable of interest;

m is a source of 'irrelevant shared variance' (e.g. due to shared method);

$X_1, X_2, \ldots X_i, \ldots X_k$ are the multiple measures or indicators;

$e_1, e_2, \ldots e_i, \ldots e_k$ are independent error terms associated with each of the k indicators.

Each measure X_i will be considered as comprising three components: a 'true score' component, a 'method' component, and an 'independent error' component:

$$X_i = a_i Y + c_i m + e_i, \quad i = 1, 2, \ldots k \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

For convenience, we shall treat Y, m, and each e_i as independent, normally distributed random variables with zero mean and variances σ_y^2, σ_m^2 and σ_i^2 respectively. We shall define the VALIDITY of a measure as the correlation of the measure with the underlying variable of interest; and the MEAN SQUARED ERROR (MSE) of a measure as the expected value of the squared difference between the value of a measure and the value of the underlying variable. In the above model:

a. Validity of measure i, $\rho_{x_i y} = \dfrac{\mathrm{Cov}\,(X_i, Y)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(Y)}}$ (by def'n)

or: $$\rho_{x_i y} = \frac{a_i \sigma_y}{\sqrt{a_i^2\sigma_y^2 + c_i^2\sigma_m^2 + \sigma_i^2}} \quad \text{.....} \quad (2)$$

b. Mean squared error of measure i, $MSE_i = E(Y - X_i)^2$ (by def'n)

or: $$MSE_i = \sigma_y^2(1-a_i)^2 + \sigma_m^2 c_i^2 + \sigma_i^2 \quad \text{.....} \quad (3)$$

c. Intercorrelation of measure i and measure j, $\rho_{x_i x_j}$

$$\rho_{x_i x_j} = \frac{a_i a_j \sigma_y^2 + c_i c_j \sigma_m^2}{\sqrt{(a_i^2\sigma_y^2 + c_i^2\sigma_m^2 + \sigma_i^2)\,(a_j^2\sigma_y^2 + c_j^2\sigma_m^2 + \sigma_j^2)}} \quad \text{.....} \quad (4)$$

Composite Measures

We wish to examine the properties of simple linear composites, $\hat{Y}$, as measures of Y. Such measures are formed by the summation of the measures, X_i, each weighted by some coefficient b_i:

$$\hat{Y} = \sum_{i=1}^{k} b_i X_i$$

For composites of this form, it can be shown (Appendix 1) that:

d. Validity of composite, $\rho_{y\hat{y}}$, is given by:

$$\rho_{y\hat{y}} = \frac{\sigma_y \sum(a_i b_i)}{\sqrt{\sigma_y^2 (\sum a_i b_i)^2 + \sigma_m^2(\sum b_i c_i)^2 + \sum(b_i^2\sigma_i^2)}} \quad \text{.....} \quad (5)$$

e. Mean squared error of composite, $MSE_{\hat{y}}$, is given by:

$$MSE_{\hat{y}} = \sigma_y^2(1-\sum a_i b_i)^2 + \sigma_m^2 (\sum b_i c_i)^2 + \sum(b_i^2\sigma_i^2) \quad \text{.....} \quad (6)$$

Thus, for a given measurement model, Equations (5) and (6) allow the validity and mean squared error of a composite to be expressed as an explicit function of the weights, b_i, associated with each measure, X_i, in the composite. For example, if the composite is constructed by merely averaging the raw scores on each measure, this amounts to setting each $b_i = 1/k$ for all k measures. Substitution in (5) and (6) allows direct calculation of the validity and MSE of a measure so constructed.

Equations (5) and (6) also open the possibility of deriving optimal weightings for a given measurement setting. For example, setting the partial derivatives of (6) with respect to each b_i equal to zero yields a set of k equations which, when solved simultaneously, give values for b_i that minimize $MSE_{\hat{y}}$. (The general procedure is shown in Appendix 1. It should be noted that the algebra involved is tedious for the general case for values of k greater than 2. In the examples that follow, we shall generally simplify the model to specific cases of interest.)

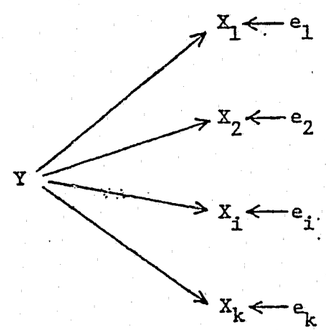
Equations (2) through (6) express the key properties of individual measures (validity and mean squared error), the intercorrelation between pairs of measures, and the properties of simple linear composites of several measures (validity and mean squared error) as explicit functions of the structure and parameters of the measurement model plus, for the composites, the weighting scheme chosen. The implications of these equations for several familiar measurement situations will be explored in the remainder of this paper.

In the following Section, we shall examine the behavior of composite measures formed from several equally-valid measures, both for the case in which the measures share no method variance, and for the case in which this shared variance is significant. In the Third Section of the paper, we shall explore the consequences of relaxing this assumption of equal validity.

II. COMPOSITES OF EQUALLY-VALID MEASURES

a. Equally-Valid Measures, Zero Method Variance

The simplest special case of the general model presented earlier is that in which each measure comprises only 'true score' and 'random error' terms, with no 'shared method' variance; and in which all measures are of equal variance, and equally valid. This simplification of the general model amounts to setting $\sigma_m^2 = 0$ (or, equivalently, $c_i = 0$); setting $a_i = 1$, $i=1,\ldots k$; and setting all σ_i^2 equal. That is:



$$X_i = Y + e_i, \quad i=1,\ldots k,$$

$$\sigma_i^2 = \sigma_j^2 = \sigma_e^2 \text{ for all } i, j.$$

From Eqn (2):

Validity of each measure, $$\rho_{x_i y} = \frac{\sigma_y}{\sqrt{\sigma_y^2 + \sigma_e^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

From Eqn (4):

Intercorrelation between any pair of measures,

$$\rho_{x_i x_j} = \frac{\sigma_y^2}{\sigma_y^2 + \sigma_e^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

$$= \rho^2_{x_i y}$$

From Eqn (5):

Validity of composite formed by $b_i = 1/k$,

$$\rho_{y\hat{y}} = \frac{\sigma_y}{\sqrt{\sigma_y^2 + \sigma_e^2/k}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

Note that, by Eqn (9), as k becomes very large, the validity of the composite approaches 1.0. The form of this function for various values of individual-measure validity, $\rho_{x_i y}$, is shown in Figure 1.

As expected, the validity of the composite increases as k increases. However, the rate at which this improvement occurs falls quickly as k increases, particularly when the individual measures are themselves reasonably valid. For example, if the individual measures have validities over about .70, little is gained by aggregating more than half a dozen of them. For less valid individual measures, composites of larger numbers may be justified by the improvement of composite validity.

Ghiselli (1964) provides an alternative formulation for the validity of equally-weighted composites. Using individual measures that have been expressed in 'standardized' form (i.e. transformed to zero mean and unit variance), Ghiselli shows that:

$$\rho_{y\hat{y}} = k^{1/2}\bar{\rho}_{yx} / [1 + (k-1)\,\bar{\rho}_{x_i x_j}]^{1/2} \quad \ldots\ldots\ldots\ldots\ldots \quad (10)$$

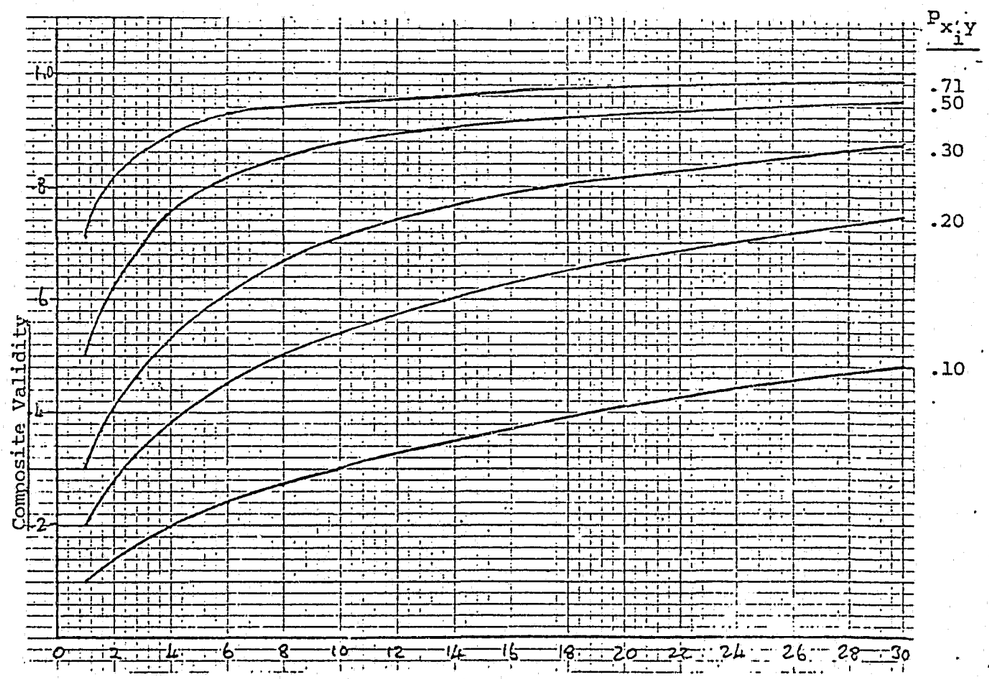
where:

$\rho_{y\hat{y}}$ = correlation of composite with y, i.e. the validity of the composite;

k = number of measures included in the composite;

$\bar{\rho}_{yx}$ = mean validity of the k measures included;

FIGURE 1: Composite Validity as a Function of k (Number of Measures Included) and Validity of Individual Measures, $\rho_{x_i y}$.



and

$\bar{\rho}_{x_i x_j}$ = mean intercorrelation of the k measures.

For k equally-valid measures, Eqn (10) produces curves identical with those generated from Eqn (9).

Ghiselli's formulation needs to be treated with some caution. From Eqn (10), it can be readily shown that, as k becomes large:

$$\lim_{k \to \infty} \rho_{y\hat{y}} = \bar{\rho}_{yx} / (\bar{\rho}_{x_i x_j})^{1/2} \quad \text{.....} \quad (11)$$

That is, the upper limit on the validity of the composite is the mean validity of the k measures divided by the square root of the mean intercorrelation. Given a choice of several indicators or measures, then, we should choose those that are (a) highly valid in themselves; and (b) minimally correlated with one another. However, a puzzling implication Eqn (11) is that a pair of orthogonal measures (i.e. measures that show zero intercorrelation) produces a composite of infinite validity if either one of them has positive validity. Given an explicit measurement model, such as that considered here, no such violation can arise. By Eqn (8), the correlation between any two measures generated by this measurement model is equal to the square of the validity of each, so that the upper limit expressed by Eqn (11) on validity of a composite is, as it must be, equal to 1.0.

Summary

For this simplest multi-measure case, in which each measure is of equal validity and contains only 'true score' and 'random error' components, an optimally-valid composite is formed by any equal-weighting scheme. The validity of the composite measure thus formed is a positive function of k,

the number of measures included, and is upper-bounded at 1.0. The rate at which this upper bound is approached as k increases is a function of the validity of the individual measures.

b. Equally-Valid Measures, Shared Method Variance

We noted earlier that, in practical measurement situations, there will often be variance shared between measures over and above their common 'true score' component. We labelled this additional shared variance as due to 'shared method', though it can, of course, arise from other causes. In general, the effects of this additional shared variance will be to raise the intercorrelation between measures, while depressing the validity both of the individual measures, and of their composites.

For example, consider the measurement model that results if we set all a_i and all c_i equal to unity and all error terms e_i of equal variance σ_e^2. Then, from Eqn (2):

$$\text{Validity of measure i, } \rho_{x_i y} = \frac{\sigma_y}{\sqrt{\sigma_y^2 + \sigma_m^2 + \sigma_e^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (12)$$

From Eqn (4), the intercorrelation between two measures of this form is

$$\rho_{x_i x_j} = \frac{\sigma_y^2 + \sigma_m^2}{\sigma_y^2 + \sigma_m^2 + \sigma_e^2} = \rho_{x_i y}^2 + \frac{\sigma_m^2}{\sigma_y^2 + \sigma_m^2 + \sigma_e^2} \quad \ldots\ldots\ldots\ldots \quad (13)$$

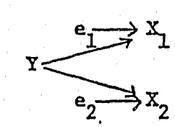
Forming a composite by setting $b_i = 1/k$ yields, by Eqn (5):

$$\text{Validity of composite, } \rho_{y\hat{y}} = \frac{\sigma_y}{\sqrt{\sigma_y^2 + \sigma_m^2 + \sigma_e^2/k}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (14)$$

Comparison of Eqn (12) with Eqn (7) shows that the validity of each measure is

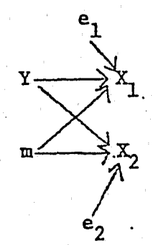
reduced; of (13) with (8) that intercorrelation between measures is increased; and of (14) with (9) that composite validity is reduced. Importantly, the upper limit on composite validity, which was 1.0 for the 'random errors' model, is reduced to $\sigma_y/(\sigma_y^2 + \sigma_m^2)^{1/2}$ for the 'shared method variance' model. For example, if 'shared method variance' is equal to 'true score' variance in each measure, the upper limit on validity for a composite of infinite number of such measures is $1/\sqrt{2}$ or .707, rather than unity, the upper limit with no shared method term. As k, the number of measures in the composite, increases, this limit is approached by a family of curves similar to those shown in Figure 1.

Shared method (or other irrelevant) variance between measures presents a trap for the unwary empiricist. Suppose, for example, a particular study includes two measures intended to tap a single underlying variable or construct. The data analysis shows the two measures to be correlated .707. Clearly, such a result is consistent with a measurement model of the following form:



where $\sigma_1^2 = \sigma_2^2 = .414\ \sigma_y^2$

That is, if the investigator is prepared to assume a 'true score plus independent random error' model, the observed intercorrelation is consistent with two measures, each valid .841; an equally-weighted composite of the two is valid .908 — a highly satisfactory measure. However, the observed intercorrelation is equally consistent with an alternative model including shared method variance:



for which a range of parameter values might generate the observed value of $\rho_{x_1x_2}$. In the worst possible case, neither measure need be at all valid ($\rho_{x_1y} = \rho_{x_2y} = 0.0$), in which case any composite of the two will also have zero validity. That is, merely observing positive intercorrelation between measures is not sufficient to establish the validity of either, or of any composite of them. As the example suggests, a modestly-high correlation between a pair of measures can imply anything from excellent to impossible composite measures. The critical difference between the two situation is not in the value of the intercorrelation found between the measures, but in the structure and parameters of the relevant measurement model.

Summary

For a given measurement model, the introduction of shared method variance will increase the intercorrelation between measures, while depressing the validity both of individual measures and of composites formed from them. These effects constitute a potential trap for empirical studies, in that high intercorrelation between measures can be mistakenly treated as evidence for high validity of individual measures, and thus for supposing composites of the measures to be highly valid. In fact, as we have shown here, high correlation between obtained measures is equally consistent with measures of zero validity, composites of which will have zero validity also.

c. Mean Squared Error Considerations

For k measures of equal validity and equal variance, composites formed by any set of equal positive weights will be of equal validity. However, the actual values of the weights used have an important effect on the MSE of the resulting composites. In this section, we shall consider the MSE of composites formed by two common 'equal-weighting' schemes--Average Raw Scores (ARS) and Average Standardized Scores (ASS)--and compare them to the MSE of optimal (Opt) composites. (While we agree with Dawes (1979: 576, footnote 5) that ARS composites will, in many cases, be 'nonsensical,' such composites are included in the assessments of Schmidt, 1971, and Wesman and Bennett, 1959, where they are shown to be reasonably robust. ARS composites can certainly be found without difficulty in recent published studies. Though the use of such composites appears, in most cases, to be the result of methodological carelessness, it is possible to specify situations in which ARS composites outperform ASS composites. An admittedly degenerate example is given later in the paper.)

The simplest form of the general measurement model generating equally-valid measures and including shared method variance is achieved by setting all c_is equal, all a_is equal, and all error terms σ_i^2 equal. For such a model, the procedure outlined in Appendix 1 (see Eqn A:11) shows that the composite minimizing MSE (i.e. the optimal composite) is achieved, for a composite of k measures, by weighting each measure equally with weight

$$b_i = \frac{a\sigma_y^2}{k(a^2\sigma_y^2 + c^2\sigma_m^2) + \sigma_e^2}, \quad i=1,\ldots k. \qquad \ldots\ldots\ldots\ldots (15)$$

In contrast, averaging the standardized measures amounts to assigning each a weight

$$b_i = \frac{1}{k\sqrt{a^2\sigma_y^2 + c^2\sigma_m^2 + \sigma_e^2}}, \quad i=1,\ldots k. \qquad \text{(16)}$$

while, clearly, simple averaging assigns each measure a weight $b_i = 1/k$.

To illustrate the impact of these three weighting schemes on MSE of actual composites, consider the case when $a = c = 1.0$, and $\sigma_y^2 = \sigma_m^2 = \sigma_e^2 = 1.0$. That is, each measure consists of three components of equal variance: a true score term, a method term, and random error. By Eqn (2), each measure is valid .577; any two measures intercorrelate .667 by Eqn (4); the MSE of each individual measure is 2.0 by Eqn (3).

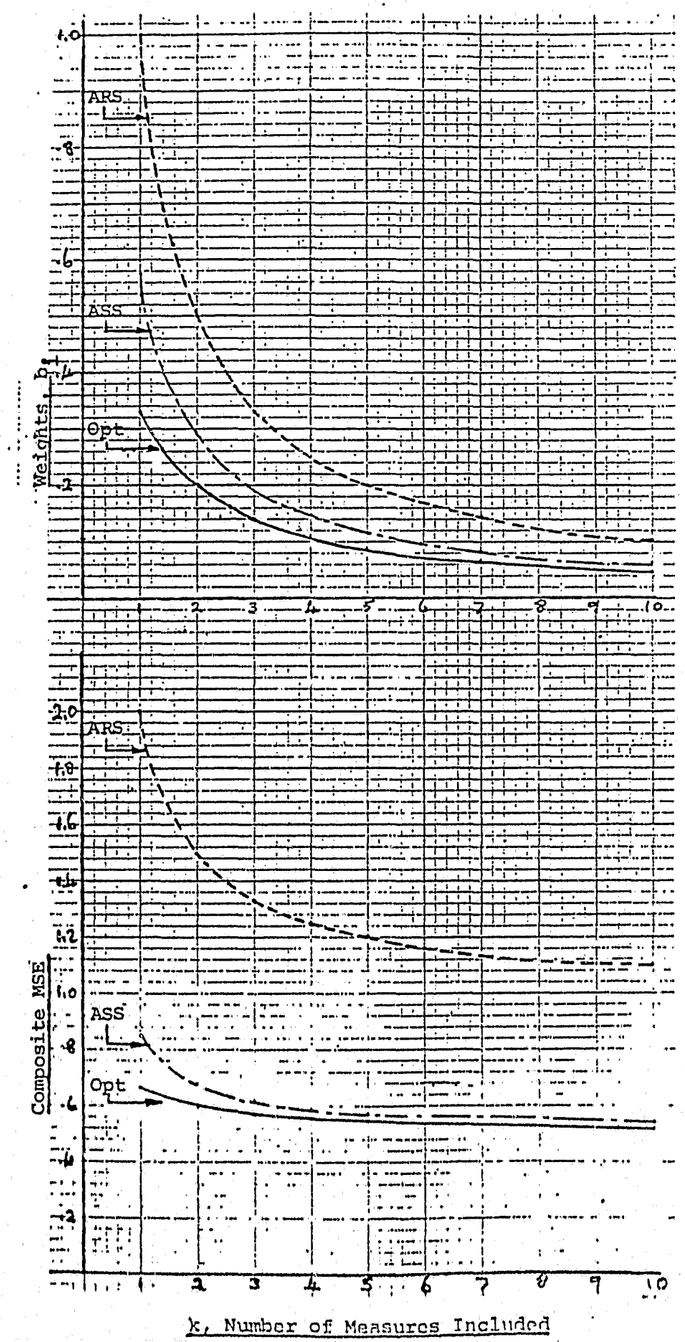
For this example, the three weighting schemes are shown in Figure 2 as a function of k, the number of measures included in the composite. The upper half of Figure 2 shows the values of the weights, b. For all three schemes, b declines as a function of k. Simple averaging assigns larger weights at any given k value than does standardizing and averaging, while optimal weights are still smaller. The three curves tend to converge as k becomes large.

The effects of these three weighting schemes on the MSE of the composite is shown in the lower half of Figure 2. In this example, simple averaging produces dramatically larger MSE than does either of the other two schemes. Standardizing and averaging, though inferior to optimal weighting, is here quite close to optimal, particularly for large values of k, and is markedly superior to simple averaging. The example thus provides some support for the familiar practice of standardizing measures before averaging them. (As we will show later in the paper, however, the practice is not always advantageous.)

Summary

In addition to the validity of a composite, investigators will often be

FIGURE 2: Weights, b_i, and Composite MSE as a Function of k (Number of Measures in Composite) for Three Weighting Schemes.



Parameters:

$a_i = c_i = 1.0$, $i=1,2,\ldots k$

$\sigma_i^2 = \sigma_m^2 = \sigma_y^2 = 1.0$

interested in their MSE. In the present section, we provide a general formulation for weights which minimize MSE for simple, equally-valid measures, either with or without shared method variance. These optimal weights are compared with simple averaging of measures, both standardized and unstandardized. In an example, the MSE of such an optimal composite was seen to be somewhat lower than the standardized-and-averaged composite, and markedly lower than that of the simple average-raw-score composite.

III. COMPOSITES OF UNEQUALLY-VALID MEASURES

In the previous sections, we have considered composites of several equally-valid individual measures, with and without 'shared method' variance components. We now turn to the more general case in which the individual measures are not equally valid. For such measures, optimal composites will require differential weighting, with measures of higher validity weighted more heavily than those of lower validity. Our general concern here is with the derivation of these optimal weights, and with the assessment of the extent to which the effort of doing so might be justified by improvement in the resulting measures, in comparison to measures with simpler weighting strategies.

As before, we shall consider first composites of measures that do not share method variance, before moving to the case in which such variance is involved.

a. Unequally-Valid Measures, Zero Method Variance

The simplest case in this category is that in which each measure contains an equal component of 'true score' variance, but unequal 'random error' components. That is, in terms of the general model, all $a_i = 1.0$, all $c_i = 0.0$, and $\sigma_i^2 \neq \sigma_j^2$. For such measures, we have shown elsewhere (Connolly, 1977) using the procedure of Appendix 1 that the MSE of the composite is minimized by assigning each measure a weight of

$$b_i^* = \frac{C_k}{\sigma_i^2} \qquad \text{(17)}$$

where

$$C_1 = \sigma_y^2\sigma_1^2/(\sigma_y^2 + \sigma_1^2) \qquad \text{(for } k = 1\text{)}$$

$$C_2 = \sigma_y^2\sigma_1^2\sigma_2^2/(\sigma_1^2\sigma_2^2 + \sigma_y^2\sigma_1^2 + \sigma_y^2\sigma_2^2) \qquad \text{(for } k = 2\text{)}$$

$$C_3 = \sigma_y^2\sigma_1^2\sigma_2^2\sigma_3^2/(\sigma_1^2\sigma_2^2\sigma_3^2 + \sigma_y^2\sigma_1^2\sigma_2^2 + \sigma_y^2\sigma_2^2\sigma_3^2 + \sigma_y^2\sigma_1^2\sigma_3^2) \qquad \text{(for } k = 3\text{)}$$

and so on. In each case, the numerator in C_k is the product of σ_y^2 and each error variance term, while the denominator is the sum of the products of these (k+1) terms taken k at a time. It should be noted that, once a measurement model of k measures has been specified, C_k is fixed. Optimal weights b_i^* are then computed by merely dividing C_k by the error variance term in each measure. Where validity rather than MSE of the composite is of interest, C_k need not be computed. Merely assigning weights in inverse proportion to the error term in each measure will yield an optimally valid composite.

When the measures differ in validity, the validity of a composite formed from them depends on the weighting scheme used (Appendix II). In Figure 3 we show the validity of two-measure composites formed by the three alternative methods: Average Raw Scores (ARS), Average Standardized Scores (ASS) and Optimally Weighted (Opt). In Figure 3(a), the first measure is valid .707, while in Figure 3(b) its validity is .302. Validity of the three composites is shown as a function of the validity of the second measure. Figure 4 shows parallel results for the validity of three-measure composites as a function of the validity of one measure, holding the validity of the other two at .707 or .302.

As these examples suggest, the use of non-optimal weighting schemes may result in significant degradations in the validity of composites. In each diagram there is a point at which all individual measures are equally valid, and thus at which any of the three weighting schemes produces an optimally-valid composite. On either side of this point the curves diverge, with optimal weighting producing the most valid composite, and Average Raw Score (ARS) weighting the least valid. The differences are not trivial. For example, in Figure 3(a) the optimal composite of two measures, one valid .707, the other .30, has validity about .72, a marginal gain over Measure 1 used alone.

National Criminal Justice Reference Service

ncjrs

While portions of this document are illegible, it was microfilmed from the best copy available. It is being distributed because of the valuable information it contains.

National Institute of Justice
United States Department of Justice
Washington, D.C. 20531

FIGURE 3: Validity of Two-Measure Composites

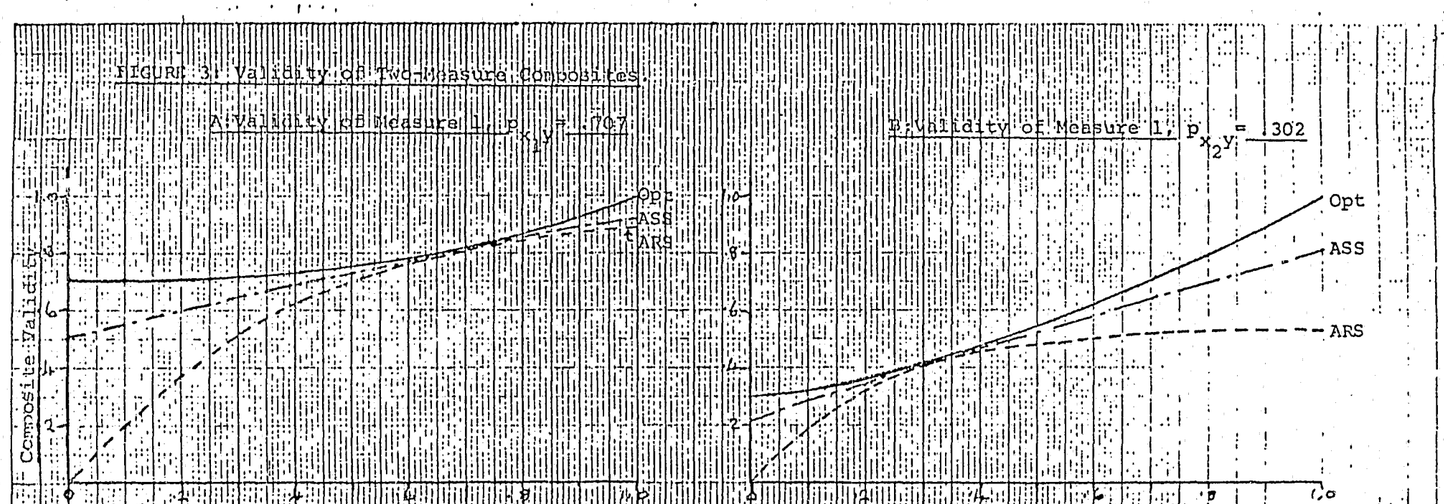
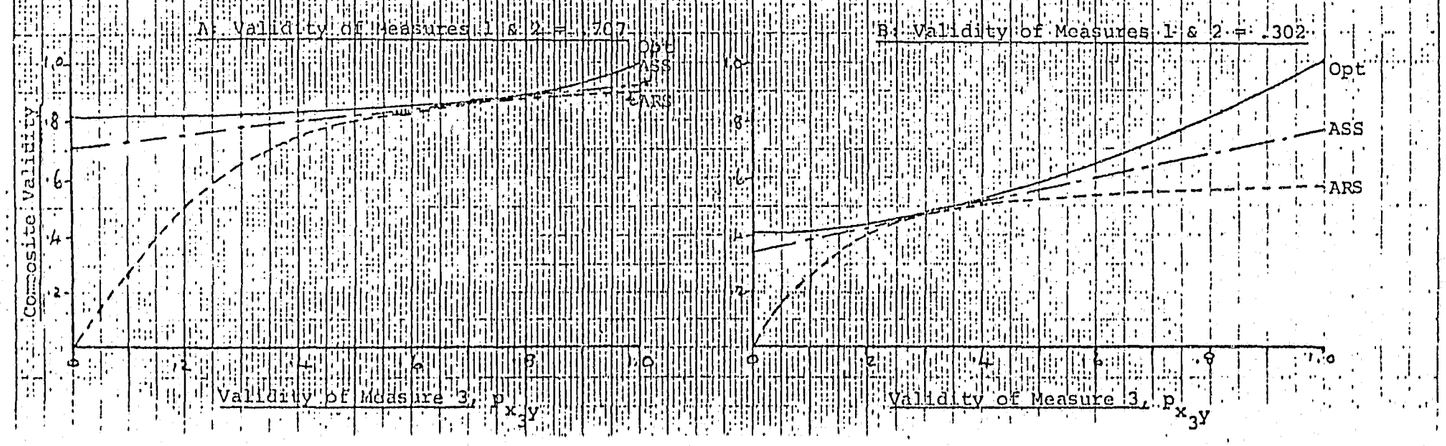


FIGURE 4: Validity of Three-Measure Composites



In contrast, the ASS composite of these two measures is valid only about .65--worse than the best measure used alone --while the ARS composite is valid only about .51.

The composite mean squared error (MSE) curves corresponding to these examples are shown in Figures 5 and 6. (True score variance, σ_y^2, is set at 1.0 in these examples.) As before, optimal weighting produces smallest composite MSE, ARS weighting the largest--so large, in fact, that the MSE for such composites cannot be plotted within the scale in Figures 5(b) and 6(b). Again, the degradation resulting from using ASS rather than optimal weighting is not too serious when the individual measures are of close to equal validity, but becomes very substantial as one moves away from this region.

Summary

When individual measures are of unequal validity, the use of non-optimal weightings degrades both validity and MSE of composites formed from them. On both criteria, ARS weights are least satisfactory. ASS composites are reasonably close to optimal when individual measures are of close to equal validity, but are increasingly suboptimal as one moves away from this region.

b. Unequally-Valid Measures, Shared Method Variance

As noted earlier, the algebra for the general case with significant shared method variance between unequally-valid measures is ungainly (though not at all difficult). In this Section, we shall therefore confine attention to composites of two measures of this type, proceeding as before by way of simple examples.

We noted earlier (Section II: b) the possibility of shared method variance accounting for observed intercorrelation between measures. As shown then, an observed correlation of .707 between two measures thought to be tapping the

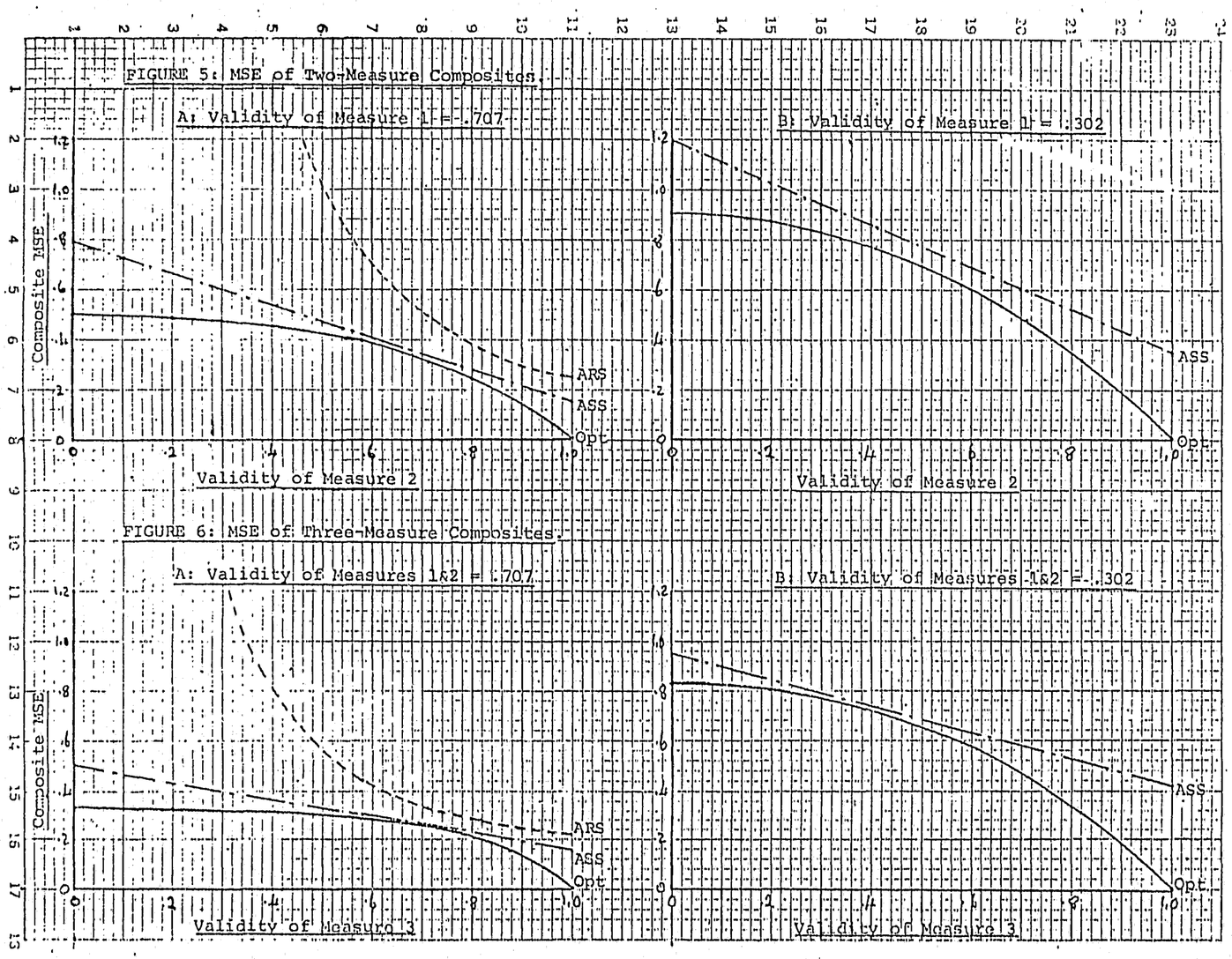


FIGURE 5: MSE of Two-Measure Composites.

FIGURE 6: MSE of Three-Measure Composites.

same underlying variable is consistent with two very different measurement models. In the most favorable case, each measure contains only true score and random error terms, and each correlates .841 with the underlying variable. An optimal (i.e. equally-weighted) composite of the two correlates .908 with the underlying variable. In the least favorable case, each measure correlates 0.0 with the underlying variable, .841 with some irrelevant variable (which we have generally been treating as m, the shared method term). No composite of these measures correlates better than 0.0 with the underlying variable of interest.

A third, and most interesting, possibility exists if the requirement of equal validity is relaxed. For example, consider the two measures

$$X_1 = Y + m \qquad (\text{i.e. } a_1 = 1,\ c_1 = 1,\ \sigma_1^2 = 0.0)$$

$$X_2 = m \qquad (\text{i.e. } a_2 = 0,\ c_2 = 1,\ \sigma_2^2 = 0.0)$$

and

$$\sigma_y^2 = \sigma_m^2$$

That is, true score and method variances are equal. The first measure has equal components of each, the second measure contains only method variance. Neither measure contains any random error variance. Substitution in Eqn (4) shows these measures correlate with one another .707, as in the earlier example. X_1 is valid .707, X_2 is valid 0.0. It is clear that merely subtracting the second measure from the first will yield a perfectly valid composite. That is, an optimal composite of the two measures is formed by assigning $b_1 = 1.0$, $b_2 = -1.0$ (see Eqn A:10) and (by Eqn A:12) this composite is valid 1.0. In contrast, an ARS composite ($b_1/b_2 = 1.0$) is valid only .447, while an ASS composite ($b_1/b_2 = .707$) is valid only .383!

It may be worth clarifying our intuition on this somewhat odd result. The optimal use of these two measures is clear: the second is a perfectly-

valid measure of the 'method' term; subtracting it from the first measure leaves a perfectly-valid measure of the true score term, Y. (Alternatively, the second measure can be seen as a 'suppressor variable' in the multiple regression sense.) Weighting the raw scores equally amounts to adding a further irrelevant source of variance to Measure 1, forming a composite which now contains only one third true-score variance, two-thirds irrelevant variance. Finally, weighting each measure in inverse proportion to its standard deviation assigns a relatively smaller weight to Measure 1 (and thus to the only true score term in the composite) than to Measure 2 (which contains only irrelevant variance), reducing the validity of the composite still further. ASS composites are not always superior to ARS composites.

As a final example, let us examine the behavior of the three alternative weighting schemes for a slightly less extreme case of the general model. Consider the two measures:

$$X_1 = Y + m + e_1$$

$$X_2 = Y + c_2 m + e_2$$

with

$$\sigma_y^2 = \sigma_m^2 = \sigma_1^2 = \sigma_2^2 = 1.0$$

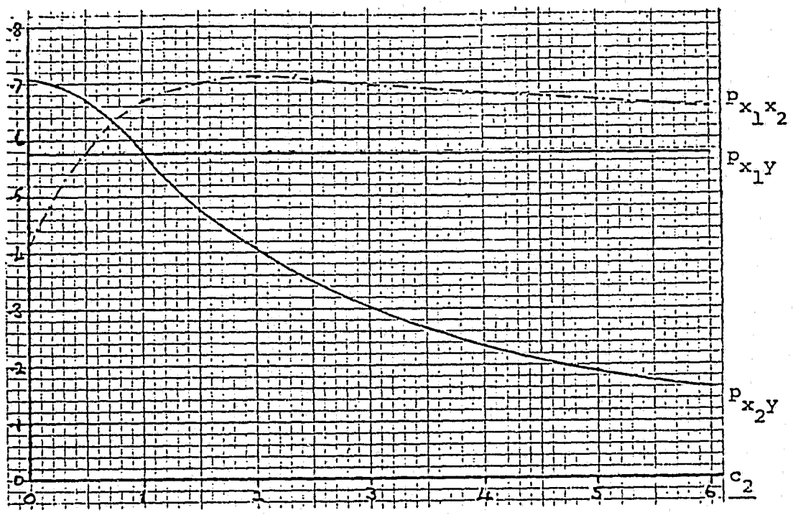
That is, we set all variance terms equal, and all coefficients equal to 1.0 with the exception of c_2, the coefficient for method term in the second measure. The behavior of the model as c_2 is varied is shown in Figure 7.

Figure 7(a) shows the validity of each measure, and the correlation between them, as c_2 increases from zero. The validity of the first measure, $\rho_{x_1 y}$, remains constant (at .577), while that of the second measure declines. Their intercorrelation at first rises, reaches a maximum of .707 at $c_2 = 2.0$, and then slowly declines.

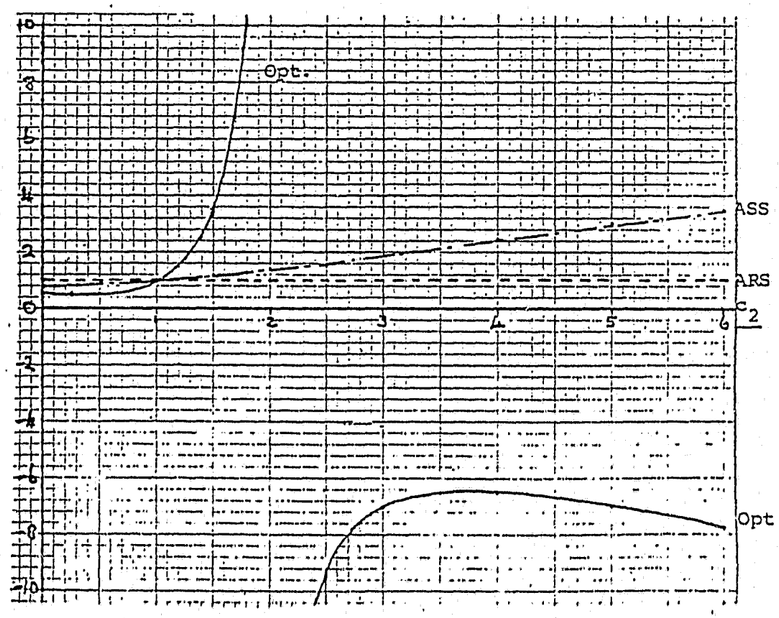
Figure 7(b) shows the ratio of b_1/b_2 for three alternative weighting

FIGURE 7: Characteristics of a Two-Measure Composite as a Function of Parameter c_2.

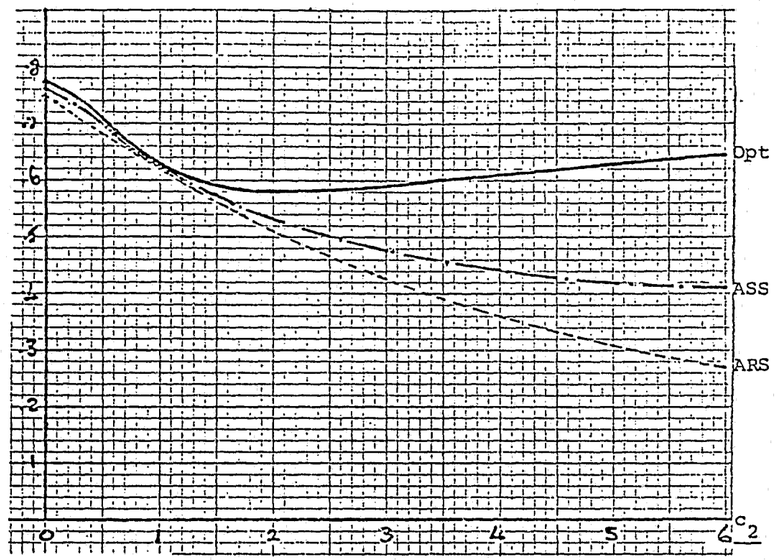
A: Validities and Intercorrelation of Measures.



B: Ratio b_1/b_2 for Alternative Weighting Schemes.



C: Validity of Three Alternative Composites.



schemes for increasing values of c_2. Using ARS weights, the ratio remains constant at 1.0. For ASS weighting, the ratio slowly increases with c_2 since the variance of the second measure is increasing, so b_2 declines. Under optimal weighting, however, the behavior of the ratio b_1/b_2 as c_2 increases shows a sharp discontinuity. As c_2 increases from zero, the ratio at first rises (reflecting the optimal weight of assigning b_2 smaller and smaller values). At $c_2 = 2.0$, the optimal weight for the second measure is zero, so the ratio b_1/b_2 becomes undefined. As c_2 increases beyond this point, negative weights for X_2 become optimal, so the ratio b_1/b_2 takes on negative values. That is, for c_2 values less than 2.0, the second measure is best used as a positive component of the composite. For c_2 values above 2.0, X_2 is best used as a 'suppressor variable,' since its values are increasingly dominated by its method variance component, and it enters the optimal composite with a negative coefficient.

Figure 7(c) shows the validity, $\rho_{y\hat{y}}$, of the three alternative composites using these weights. All three schemes yield composites of roughly equal validity for c_2 values less than 1.0, though optimal composites are slightly more valid than are ASS with ARS slightly the poorest. At $c_2 = 1.0$, the two measures are equally valid, and all three schemes yield equal composite validities. As c_2 increases further, the three schemes diverge more and more. The optimal scheme reaches a minimum at $c_2 = 2.0$ (with $b_2 = 0.0$, and composite validity equal to ρ_{x_1y}, .577), with composite validity increasing thereafter. Both ARS and ASS weighting schemes yield composites whose validities continue to decline with increasing c_2.

As with the earlier examples, care should be exercised in drawing overstrong inferences from the results shown here. However, the same general pattern can be identified: for equally-valid measures, each of the three

weighting schemes produces equally valid composites. For small divergences from equal validity for the measures, optimal composites are only marginally superior to the alternatives. For measures of rather different validity, however, the differences in validity between optimal, ASS, and ARS-weighted composites become more marked, and can be very large. Finally, ASS weights generally, though not always, yield composites superior to those formed by ARS weighting.

Conclusions

The primary purpose of this paper has been to propose a simple linear measurement model for settings in which several indicators or measures are associated with a single underlying variable of interest; and to examine the adequacy of three strategies for forming composite measures of the underlying variable. The composite measures have been assessed in terms of (a) their validity and (b) their MSE. A general algebraic treatment has been illustrated by means of a number of specific examples ranging from the simplest case (equally valid measures with no shared method variance) to the most complex (unequally valid measures with shared method variance).

Substantively, the examples have been chosen to show both the general robustness, and the possible weaknesses, of some familiar rules of thumb and standard practices in developing composite measures. For example:

a. Aggregates of more measures are better than aggregates of fewer measures.

Exceptions:

1. For measures that are individually reasonably valid, the assymptote of composite validity is approached with small numbers of measures (see Figure 1). Adding more measures may produce negligible improvements.
2. If non-optimal weights are used (e.g. averaging of either standardized or raw scores), a composite of one highly-valid measure with several low-validity measures may produce a composite less valid than the single best measure taken alone.

b. Measures that differ in method are preferable to measures using the same method.

Exceptions:

1. It may be possible to identify a measure that reflects method variance alone, or is dominated by such a component. In such circumstances, the measure can be used in an optimal composite as a 'suppressor' on the method component of other measures sharing the same method, to produce a composite of potentially high validity.

c. Averages of standardized measures are preferable to averages of raw scores.

Exception: (as b:1, above).

d. Composites formed by standardizing and averaging are generally good approximations of optimal composites.

Exceptions:

1. Depending on the interpretation of 'a good approximation,' nearly all of the examples show regions in which this rule is violated. In general, the greater the difference between the validities and structures of the individual measures, the larger is the departure of the ASS composite from optimality. This departure is, in many cases, very far from negligible.
2. Even when averages of standardized measures are of close-to-optimal validity, they may still be very far from optimality in the MSE sense. Such departure may be important in many practical cases.

Our central purpose, however, is not to endorse one familiar practice rather than another. It is to argue that the use of any standard practice without examination of the implicit measurement model is a dangerous procedure. Practices such as adding more measures to a composite, or routinely standardizing measures before averaging, are perfectly reasonable in certain circumstances, but exceptions, and sometimes important exceptions, can be

found. Merely following some standard practice in aggregating measures, without specifying the implicit measurement model, amounts to leaving the quality of one's measures up to chance: they may be optimal, near-optimal, or completely misleading; but, without an explicit measurement model, one has no way to assess which.

The central thrust of this paper, then, is that any procedure for aggregating several measures into a composite measure implies a set of assumptions about the process that is generating the measures. Making these assumptions explicit by specifying a measurement model — that is, specifying the sources of variance in each measure, and their relative magnitudes -- allows the investigator, at minimum, to consider whether or not a given procedure (such as standardize-and-average) is likely to lead to grossly misleading measures, given a particular measurement situation. To the extent that the magnitudes of the variances can be reliably estimated, the algebra presented here allows the investigator to develop procedures that out-perform standard rules of thumb.

We have not been concerned here with methods by which such estimation could be attempted, — and it is clear that it could be a difficult and costly effort. The logic of the present paper does, however, provide some guidance as to when the effort might be justified. Even when such a full-scale effort is not undertaken, however, it does seem desirable to consider at least the structure of measurement model one is assuming, and perhaps to consider what orders of magnitude might be assumed for the terms involved. At minimum, one can then assess what risks one is taking in using one of the standard rules of thumb in constructing a composite measure. At maximum, a full-scale development of a measurement model allows confidence in the measures one uses, and opens the possibility of refining them in the direction of

optimality. Without a measurement model, the quality of one's measures is a matter of chance. With such a model, the quality of one's measures is a matter for computation—and, to the extent justified, for refinement.

References

Blalock, H. M. (ed): Measurement in the Social Sciences. Chicago, Aldine, 1974.

Burt, C.: The influence of differential weighting. British Journal of Psychology, Statistical Section, 1950, 3, 124-125.

Campbell, D. T. & Fiske, D. W.: Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Connolly, T.: Cues, components and causal structure in laboratory judgment studies. Educational and Psychological Measurement, 1977, 37, 877-888.

Dawes, R. M. & Corrigan, B.: Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Dawes, R. M.: The robust beauty of improper linear models in decision making. American Psychologist, 1979, 34, 571-582.

Einhorn, H. J. & Hogarth, R. M.: Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 1975, 13, 171-192.

Ghiselli, E. E.: Theory of Psychological Measurement. New York, McGraw-Hill, 1964.

Gulliksen, H.: Theory of Mental Tests. New York, Wiley, 1950.

Laughlin, J. E.: Comment on 'Estimating coefficients in linear models: It don't make no nevermind.' Psychological Bulletin, 1978, 85, 247-253.

Lawshe, C. H. & Schucker, R. E.: The relative efficiency of four test weighting methods in multiple prediction. Educational and Psychological Measurement, 1959, 19, 103-114.

Pruzek, R. M. & Frederick, B. C.: Weighting predictors in linear models: Alternatives to least squares and limitations of equal weights. Psychological Bulletin, 1978, 85, 254-266.

Schmidt, F. L.: The relative efficiency of regression and simple unit predictor weights in applied differential psychology. Educational and Psychological Measurement, 1971, 31, 699-714.

Wainer, H.: Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217.

Wainer, H.: On the sensitivity of regression and regressors. Psychological Bulletin, 1978, 85, 267-273.

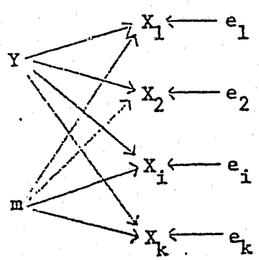
Webb, E. J., Campbell, D. T., Schwartz, R. D. & Sechrest, L.: Unobtrusive Measures. Skokie, Ill.: Rand McNally, 1966.

Wesman, A. G. & Bennett, G. K.: Multiple regression versus simple addition of scores in prediction of college grades. Educational and Psychological Measurement, 1959, 19, 243-246.

Wilks, S. S.: Weighting systems for linear functions of correlated variables when there is no dependent variable. Psychometrika, 1938, 3, 23-40.

APPENDIX 1

General Model



$$X_i = a_i Y + c_i m + e_k$$

$$\left.\begin{array}{l} Y \sim N(0,\sigma_y^2) \\ m \sim N(0,\sigma_m^2) \\ e_i \sim N(0,\sigma_i^2) \end{array}\right\} \text{All independent}$$

$$Var(X_i) = a_i^2\sigma_y^2 + c_i^2\sigma_m^2 + \sigma_i^2$$

1) Validity of X_i: By defn: $\rho_{AB} = \dfrac{Cov(AB)}{\sqrt{Var(A)\ Var(B)}}$ A:1

$$Cov(X_iY) = E(X_iY)$$

$$= E[Y(a_iY + c_im + e_i)]$$

$$= E[a_iY^2 + c_imY + Ye_i]$$

$$= a_i\sigma_y^2$$

so: $$\rho_{X_iY} = \frac{a_i\sigma_y^2}{\sqrt{(a_i^2\sigma_y^2 + c_i^2\sigma_m^2 + \sigma_i^2)(\sigma_y^2)}} = \frac{a_i\sigma_y}{\sqrt{a_i^2\sigma_y^2 + c_i^2\sigma_m^2 + \sigma_i^2}} \quad \text{.... A:2}$$

2) Intercorrelation of X_iX_j: $Cov(X_iX_j) = E(X_iX_j)$

$$= E[(a_iY + c_im + e_i)\ (a_jY + c_jm + e_j)]$$

$$= a_ia_j\sigma_y^2 + c_ic_j\sigma_m^2$$

so: $$\rho_{X_iX_j} = \frac{a_ia_j\sigma_y^2 + c_ic_j\sigma_m^2}{\sqrt{(a_i^2\sigma_y^2 + c_i^2\sigma_m^2 + \sigma_i^2)(a_j^2\sigma_y^2 + c_j^2\sigma_m^2 + \sigma_j^2)}} \quad \text{..... A:3}$$

3) Validity of Composites:

$$\hat{Y} = \sum_{i=1}^{k} b_iX_i$$

$$= b_1(a_1Y + c_1m + e_1) + b_2(a_2Y + c_2m + e_2) + \ldots$$

$$= Y\sum(b_ia_i) + m\sum(b_ic_i) + \sum(b_ie_i)$$

$$Cov(Y\hat{Y}) = \sigma_y^2\sum(b_ia_i)$$

$$Var(\hat{Y}) = E(\hat{Y}^2) = \sigma_y^2\ [\sum(a_ib_i)]^2 + \sigma_m^2\ [\sum(b_ic_i)]^2 + \sum(b_i^2\sigma_i^2)$$

so: $$\rho^2_{y\hat{y}} = \frac{[\sigma_y^2\sum(b_ia_i)]^2}{\sigma_y^2[\sigma_y^2[\sum(a_ib_i]^2 + \sigma_m^2[\sum(b_ic_i)]^2 + \sum(b_i^2\sigma_i^2)]}$$

or: $$\rho_{y\hat{y}} = \frac{\sigma_y(\sum a_ib_i)}{\sqrt{\sigma_y^2(\sum a_ib_i)^2 + \sigma_m^2(\sum b_ic_i)^2 + \sum b_i^2\sigma_i^2}} \quad \text{..... A:4}$$

4) Mean Squared Error of Composites:

$$E[(Y-\hat{Y})^2] = E[(Y-Y\sum a_ib_i - m\sum b_ic_i - \sum b_ie_i)^2]$$

or: $$MSE_{\hat{y}},\ L = \sigma_y^2(1-\sum a_ib_i)^2 + \sigma_m^2(\sum b_ic_i)^2 + \sum(b_i^2\sigma_i^2) \quad \text{..... A:5}$$

For k=1, (A:5) becomes: $$L = \sigma_y^2(1-a_1b_1)^2 + \sigma_m^2(b_1c_1)^2 + b_1^2\sigma_1^2$$

$$\frac{dL}{db_1} = (-a_1)2\sigma_y^2(1-a_1b_1) + c_12\sigma_m^2(b_1c_1) + 2b_1\sigma_1^2$$

Setting $\frac{dL}{db_1} = 0$ and rearranging:

$$\text{Optimal Weight,} \quad b_1^* = \frac{a_1\sigma_y^2}{(\sigma_1^2 + c_1^2\sigma_m^2 + a_1^2\sigma_y^2)} \quad \text{..... A:7}$$

<u>For k=2:</u> A:5 becomes: $L = \sigma_y^2(1-a_1b_1-a_2b_2)^2 + \sigma_m^2(b_1c_1 + b_2c_2)^2 + b_1^2\sigma_1^2 + b_2^2\sigma_2^2$

$$\frac{\delta L}{\delta b_1} = 2a_1\sigma_y^2(1-a_1b_1-a_2b_2) + 2c_1\sigma_m^2(b_1c_1 + b_2c_2) + 2b_1\sigma_1^2$$

Setting $\frac{\delta L}{\delta b_1} = 0$ and rearranging:

$$b_1^* = \frac{a_1\sigma_y^2(1-a_2b_2^*)-b_2^*c_1c_2\sigma_m^2}{c_i^2\sigma_m^2 + \sigma_1^2 + a_1^2\sigma_y^2}$$

$$= \frac{a_1\sigma_y^2-b_2^*(a_1a_2\sigma_y^2 + c_1c_2\sigma_m^2)}{(\sigma_1^2 + c_1^2\sigma_m^2) + a_1^2\sigma_y^2} \quad \text{..... A:8}$$

Similarly:

$$b_2^* = \frac{a_2\sigma_y^2-b_1^*(a_1a_2\sigma_y^2 + c_1c_2\sigma_m^2)}{(\sigma_2^2 + c_2^2\sigma_m^2) + a_2^2\sigma_y^2} \quad \text{..... A:9.}$$

Solving A:8 and A:9 simultaneously:

$$b_1^* = \frac{\sigma_y^2(a_1R_2-ka_2)}{R_1R_2-k^2}; \qquad b_2^* = \frac{\sigma_y^2(a_2R_1-ka_1)}{R_1R_2-k^2} \quad \text{..... A:10}$$

where:

$$R_1 = a_1^2\sigma_y^2 + c_1^2\sigma_m^2 + \sigma_1^2 = Var(X_1)$$

$$R_2 = a_2^2\sigma_y^2 + c_2^2\sigma_m^2 + \sigma_2^2 = Var(X_2)$$

and

$$k = a_1a_2\sigma_y^2 + c_1c_2\sigma_m^2 = Cov(X_1X_2)$$

In general, for k measures, partial differentiation w.r.t. b_i in Equation (A:5) yields k solutions of the form:

$$b_iR_i = a_i\sigma_y^2 - \sum b_jk_{ij} \qquad i \neq j,\ j = 1, 2, \ldots k \quad \text{..... A:11}$$

where:

$$R_i = a_i^2\sigma_y^2 + c_i^2\sigma_m^2 + \sigma_i^2 = Var(X_i)$$

$$k_{ij} = a_ia_j\sigma_y^2 + c_ic_j\sigma_m^2 = Cov(X_iX_j)$$

Solving Equations (A:13) simultaneously yields optimal weights $b_1^*, b_2^*, \ldots b_k^*$

APPENDIX II

<u>A Note on the Validity of Two-Measure Composites.</u>

From Eqn. (5), any two-measure composite with weights $b_1/b_2 = r$ has validity

$$\rho_{y\hat{y}} = \frac{\sigma_y(a_2 + a_1r)}{\sqrt{\sigma_y^2(a_2 + a_1r)^2 + \sigma_m^2(c_2 + c_1r)^2 + r^2\sigma_1^2 + \sigma_2^2}} \quad \text{..... A:12}$$

Equal weighting of raw scores amounts to setting r = 1 in (A:12). Equal weighting of 'standardized' scores amounts to setting the ratio r equal to:

$$r_{std} = \sqrt{\frac{a_2^2\sigma_y^2 + c_2^2\sigma_m^2 + \sigma_2^2}{a_1^2\sigma_y^2 + c_1^2\sigma_m^2 + \sigma_1^2}} \quad \text{..... A:13}$$

while the use of optimal weights, from A:10, implies a ratio:

$$r_{opt} = \frac{a_1\sigma_2^2 + c_2\sigma_m^2(a_1c_2-a_2c_1)}{a_2\sigma_1^2 + c_1\sigma_m^2(a_2c_1-a_1c_2)} \quad \text{..... A:14}$$

These formulations simplify somewhat the two-measure examples discussed in the text.

END