First published in *Psychometrika*, 1979, *44*, 289–296.

Sensitivity of a Linear Composite of Predictor Items to Differential Item Weighting

Abstract

Many authors have demonstrated for idealized item configurations that equal item weights are often virtually as good for a particular predictive purpose as the item weights that are theoretically optimal. What has not been heretofore clear, however, is what happens to the similarity between weighted and unweighted composites of the same items when the item configuration's variance structure is complex.

Equal-Weight Composites

The enthusiasm with which Wainer (1976), expanding upon the sentiments of Dawes and Corrigan (1974), has advocated the use of equal item weights for applied multivariate prediction is not entirely unjustified. However, his "Equal Weights Theorem" (corrected by Laughlin, 1978), which professes to show the robustness of equal item weights for the practical efficiency of composite predictors, in fact argues for this only under arbitrarily special assumptions whose relevance for practical prediction is demonstrably almost nil. Equal-weight predictor composites are indeed nearly optimal under seemingly wide conditions. But those conditions are more subtle than the extant literature has properly noted. In particular, before enthusing over the *trend* of item composites' general insensitivity to differential item weighting, we need also to reflect upon that trend's reliability.

That variation in the weights $\{w_i\}$ assigned to the items in a linear composite $\tilde{x} = w_0 + w_1 x_1 + \cdots + w_m x_m$ of predictor variables often makes little difference for $\tilde{x}$ has been noted more than once in the psychometric literature (see especially Gulliksen, 1950, Ch. 20). And while most published expressions of this effect have been difficult to interpret save through unrealistically idealized simplifying assumptions, it is possible to characterize the responsiveness of composite $\tilde{x}$ to differential item weighting in a fashion that is surprisingly insightful considering its generality. I shall review that characterization once the problem's nature has been clarified.

It is well-known that the linear composite of variables $X = \langle x_1, \cdots, x_m \rangle$, having minimal standard error for predicting a criterion variable y in population P, is the projection $\dot{x} = b_0 + \sum_{i=1}^{m} b_i x_i$ of y into the space spanned in P by X, i.e. $b_1, \cdots, b_m$ are the coefficients in the linear regression of y upon X in P. (I write 'b' rather than

'β' for the regression coefficients because I do *not* want to presuppose that y and the x_i are standardized to unit variance.) To assess the relative efficiency of some other composite $\tilde{x} = b_0 + \sum_{i=1}^{m} w_i x_i$ for predicting y from these same items, we can proceed in either of two ways. The one exploited by Wainer is to standardize y to unit variance, to note that $\mathrm{var}(y - \tilde{x}) = \mathrm{var}(y - \dot{x}) + \mathrm{var}(\dot{x} - \tilde{x})$, and to take $\mathrm{var}(\dot{x} - \tilde{x})$ as our measure of $\tilde{x}$'s inefficiency compared to $\dot{x}$. However, this has technical disadvantages due to the numerical value of $\mathrm{var}(\dot{x} - \tilde{x})$ being determined in part by scaling parameters for predictor dimension $\tilde{x}$ that are generally irrelevant to predictive use of $\tilde{x}$ (see Rozeboom, 1978). The sensitivity of this approach to scaling artifacts has tricked Wainer into choosing premises for his Equal Weights Theorem that can be satisfied only when there are at most three predictor items.

Proof. Wainer stipulates that $x_1, \cdots, x_m$ are linearly uncorrelated with unit variances, and that their regression weights $\{b_i\}$ for y are "uniformly distributed on the interval [.25, .75]." Then $\sum_{i=1}^{m} b_i^2 = \mathrm{var}(\sum_{i=1}^{m} b_i x_i) = \mathrm{var}(\dot{x}) \leq \mathrm{var}(\dot{x}) + \mathrm{var}(y - \dot{x}) = \sigma_y^2$, or $\sum_{i=1}^{m} b_i^2 \leq 1$ since y is also assigned unit variance. Hence $\sigma_b^2 + \bar{b}^2 = \sum_{i=1}^{m} b_i^2/m \leq m^{-1}$, where $\bar{b}$ and σ_b are respectively the mean and standard deviation of $b_1, \cdots, b_m$. But $\bar{b} = .5$ and $\sigma_b > 0$ under the premised weight distribution; whence $m < \bar{b}^{-2} = 4$, i.e. integer m must be three or less. □

Wainer's approach can easily be generalized to m greater than 3; but to do so the range stipulated for $\{b_i\}$ must be formulated as a carefully controlled function of m, with the common weight given to each predictor in the equal-weight composite similarly varying with m,

Alternatively, however, we can avoid scaling irrelevancies by assessing the y-predictive efficiency of an arbitrary item composite $\tilde{x}$, compared to the accuracy of regression estimate $\dot{x}$ of y, simply by the squared correlation $\rho_{\dot{x}\tilde{x}}^2$ between $\dot{x}$ and $\tilde{x}$ in P. For since $\rho_{y\tilde{x}}^2 = \rho_{yX}^2 \rho_{\dot{x}\tilde{x}}^2$, where $\rho_{yX} (= \rho_{y\dot{x}})$ is y's multiple correlation with the x_i in P,

$$\frac{\rho_{y\tilde{x}}^2}{\rho_{yX}^2} = \frac{\rho_{yX}^2 \rho_{\dot{x}\tilde{x}}^2}{\rho_{yX}^2} = \rho_{\dot{x}\tilde{x}}^2$$

is the measure that tells what proportion of y's variance accounted for by y's regression on $\{x_i\}$ is still accounted for when y is predicted just from $\tilde{x}$. Since none of these squared correlations is affected by linear transformations of y, $\dot{x}$, or $\tilde{x}$, we can let constants w_o and w in equal-weight composite $\tilde{x} = w_0 + \sum_{i=1}^{m} w_i x_i$ be arbitrary. It is conceptually useful to let $w_0 = 0$ and $w = m^{-1}$, in which case our equal-weight composite is the *centroid*, $\tilde{x} =_{\mathrm{def}} (\sum_{i=1}^{m} x_i)/m$, of the predictor items. In what follows, I will show how the correlation $\rho_{\dot{x}\tilde{x}}$ between the centroid of predictor items $x_1, \cdots, x_m$ and criterion variable y's projection into X-space is determined jointly by certain critical properties of y's regression coefficients $\{b_i\}$ on the X-configuration and the variance structure of the items. Actually, it will be irrelevant that the b_i are regression coefficients for predicting an external criterion y. The main point at issue is simply the correlation between some target composite $\dot{x} = b_0 + \sum_{i=1}^{m} b_i x_i$ and the unweighted (equal-weight) composite of the same items.

Before the virtues of centroid predictors can be meaningfully examined, something needs be said about item scales. For insomuch as each direction in X-space is collinear with the centroid of the item configuration under *some* choice of item scales (since scaling parameters can adjust an item's orientation as well as its mean and variance), any theorem establishing high efficiency of $\tilde{x}$ for predicting y would be trivial if we are allowed to select item scales to yield $\tilde{x} = \dot{x}$. Also, the familiar theoretical expedient of unit-variance items is virtually never implemented in applied prediction. Accordingly, the present analysis will assume fixed but arbitrary item scales. Still, even without estimating regression parameters, we have great latitude in choosing item scales; and some such selection must be made in any event, with or without an assist from statistical considerations. After we see how the correlation between $\dot{x}$ and $\tilde{x}$ is determined for an arbitrary choice of item scales we will be in a better position to judge the merits of various scaling alternatives.

For any fixed scalings of predictor items $X = \langle x_1, \cdots, x_m \rangle$, let v_b be the coefficient of variation for the coefficients in target composite $\dot{x} = b_0 + \sum_{i=1}^{m} b_i x_i$. That is, $v_b = \sigma_b / \bar{b}$, where $\bar{b}$ and σ_b are respectively the mean and standard deviation of $b_1, \cdots, b_m$. Quantity v_b, which may be thought of as the "extremity" of weight distribution $\{b_i\}$, characterizes the degree to which the item weights in $\dot{x}$ diverge from equal weighting; and it is intuitively evident that the larger is v_b, the smaller on the whole should be $\rho^2_{\dot{x}\tilde{x}}$. (When $v_b = 0$, $\rho^2_{\dot{x}\tilde{x}} = 1$ since $\dot{x}$ and $\tilde{x}$ are then collinear.) But it is also evident that for fixed $\{b_i\}$, the more homogeneous are the x_i the larger will be $\rho^2_{\dot{x}\tilde{x}}$. And in fact, we shall see that $\rho^2_{\dot{x}\tilde{x}}$ is largely determined just by v_b and the predictor items' internal consistency in a way that sustains the classic contention (see e.g., Gulliksen, 1950) that differential weighting of a goodly number of reasonably homogeneous items tends not to matter much. However, the *reliability* of that simple trend is profoundly modulated by properties of the predictors' variance structure other than homogeneity. This latter effect does not seem to have received much recognition; yet it is just as important a part of the item-weighting story as is the general trend.

To make clear how $\rho^2_{\dot{x}\tilde{x}}$ is determined by item weights $\{b_i\}$ and the predictor configuration's variance structure, we need some partly-unfamiliar technical machinery. Let each predictor item $x_i (i = 1, \cdots, m)$ be analyzed as the sum, $x_i = d_i + \tilde{x}$, of two components, a "saturation" component $\tilde{x}$ (i.e. the items' centroid) shared by all the items plus a "dispersion" component $d_i =_{\text{def}} x_i - \tilde{x}$. Because $\sum_{i=1}^{m} d_i = 0$, the total X-variance $V_x =_{\text{def}} \sum_{i=1}^{m} \sigma^2_{x_i}$ analyzes as the total variance $m\sigma^2_{\tilde{x}}$ of the items' m saturation components plus the total variance $V_D =_{\text{def}} \sum_{i=1}^{m} \sigma^2_{d_i}$ of the items' dispersion configuration $D = \langle d_1, \cdots, d_m \rangle$.

(Proof: $\sum_{i=1}^m \sigma_{x_i}^2 = \sum_{i=1}^m \mathrm{var}(d_i + \bar{x}) = \sum_{i=1}^m \sigma_{d_i}^2 + \sum_{i=1}^m \mathrm{cov}(d_i, \tilde{x}) + m\sigma_{\tilde{x}}^2$ while $\sum_{i=1}^m \mathrm{cov}(d_i, \tilde{x}) = \mathrm{cov}(\sum_{i=1}^m d_i, \tilde{x}) = \mathrm{cov}(0, \tilde{x}) = 0$.) If sat_X and disp_X are the proportions of total item variance that are saturation variance and dispersion variance, respectively, i.e.

$$\mathrm{sat}_X =_{\mathrm{def}} \frac{m\sigma_{\tilde{x}}^2}{V_X} ,$$
$$\mathrm{disp}_X =_{\mathrm{def}} \frac{V_D}{V_X} ,$$

we thus have $\mathrm{sat}_X + \mathrm{disp}_X = 1$. The variance ratio sat_X may be viewed as a measure of item similarity, since it equals the average of all elements in the item configuration's covariance matrix C_{XX} divided by the average item variance. However, a purer measure of item homogeneity, hom_X, is the average off-diagonal element in C_{XX}, i.e. the average between-item covariance, divided by the average item variance. It can easily be shown that the ratio of hom_X to sat_X is the item configuration's "alpha coefficient," a quantity long familiar to modern test theory as an internal-consistency approximation to the item-centroid's reliability see Rozeboom, 1966, p. 410f) and for which we shall have later use. Specifically,

$$\alpha_X = \frac{\mathrm{hom}_X}{\mathrm{sat}_X} = \frac{m\,\mathrm{hom}_X}{(m-1)\,\mathrm{hom}_X + 1} . \tag{1}$$

For any fixed $\mathrm{hom}_X > 0$, α_X increases asymptotically to unity with increasing m.

The sensitivity of $\rho_{\dot{x}\tilde{x}}$ to a given extremity v_b of item weighting is determined importantly by the proportion of the items' total variance given to their dispersion configuration, but also—which is the tricky part for a theory of item weighting to make perspicuous—by their dispersion configuration's *shape*. It is not hard to show that for any m-tuple $\langle b_1, \cdots, b_m\rangle$ of non-identical item weights, the weighted item composite $\dot{x} = b_0 + \sum_{i=1}^m b_i x_i$ is collinear with $m^{1/2}\tilde{x} + v_b d_b$ for some axis d_b of an orthonormal rotation of $d_1, \cdots, d_m$ selected by the inequalities among the item weights $\{b_i\}$.

Proof. Put $d_b =_{\mathrm{def}} \sum_{i=1}^m a_i d_i$, where $a_i =_{\mathrm{def}} (b_i - \bar{b})/m^{1/2}\sigma_b$, and note that $\sum_{i=1}^m \bar{b} d_i = b\sum_{i=1}^m d_i = 0$. Then $\dot{x} - b_0 = \sum_{i=1}^m b_i x_i = (\sum_{i=1}^m b_i \tilde{x}) + (\sum_{i=1}^m b_i d_i) = (\sum_{i=1}^m \bar{b})\tilde{x} + \sum_{i=1}^m (b_i - \bar{b}) d_i = m\bar{b}\bar{x} + m^{1/2}\sigma_b(\sum_{i=1}^m a_i d_i) = m\bar{b}\tilde{x} + m^{1/2}\sigma_b d_b = m^{1/2}\bar{b}(m^{1/2}\tilde{x} + v_d b_d)$. And since $\sum_{i=1}^m a_i^2 = 1$, $d_b = \sum_{i=1}^m a_i d_i$ is an axis in some orthonormal rotation of $d_1, \cdots, d_m$.$\square$

Excluding $v_b = 0$, any value of v_b can be combined in $\{b_i\}$ with any choice of d_b, so extremity and D-axis selection are independent properties of the weight set.

From the collinearity of $\dot{x}$ with $m^{1/2}\tilde{x} + v_b d_b$ it follows that

$$(2) \qquad \rho^2_{\dot{x}\tilde{x}} = 1 - \frac{v_b^2 q_b(1-\theta_b^2\theta_X^2)}{1 + v_b^2 q_b + 2v_b q_b^{1/2}\theta_b\theta_x} \geq 1 - v_b^2 q_b \ ,$$

where θ_X is the multiple correlation of $\tilde{x}$ with the d_i, θ_b is the correlation between d_b and $\tilde{x}$'s projection into D-space, and q_b is the variance of d_b in proportion to the variance of $m^{1/2}\tilde{x}$, i.e.

$$(3) \qquad q_b =_{\text{def}} \frac{\sigma^2_{d_b}}{m\sigma^2_{\tilde{x}}} \ .$$

Proof. For any variables x, y, and z such that x is collinear with $y + z$, $\rho^2_{xy} = \rho^2_{(y+z)y} = \text{cov}(y, y + z)^2/[\text{var}(y) \cdot \text{var}(y + z)] = [\sigma^2_y + 2\sigma y^2\text{cov}(y,z) + \text{cov}(y,z)^2]/\sigma^2_y[\sigma^2_y + 2\text{cov}(y,z) + \sigma^2_z] = 1 - [\sigma^2_y\sigma^2_z - \text{cov}(y,z)^2]/\sigma^2_y[\sigma^2_y + 2\text{cov}(y,z) + \sigma^2_z] = 1 - r^2(1-\rho^2_{yz})/(1 + 2r\rho_{yz} + r^2)$ where $r =_{\text{def}} \sigma_z/\rho_y$. From there we obtain (3) by substituting $\dot{x}$ for x, $m^{1/2}\tilde{x}$ for y, and $v_b d_b$ for z, while noting that $\text{var}(v_b d_b)/\text{var}(m^{1/2}\tilde{x}) = v_b^2\sigma_{d_b}/m\sigma^2_x = v_b^2 q_b$, that $\rho_{\dot{x}(m^{1/2}\tilde{x})} = \rho_{\dot{x}\tilde{x}}$ and $\rho_{(m^{1/2}\tilde{x})(v_b d_b)} = \rho_{\tilde{x}d_b}$, and that by partitioning $\tilde{x}$ into its projection into D-space plus an orthogonal residual, $\rho_{\tilde{x}d_b}$ can be analyzed as the product of θ_X and θ_b.□

Equation (2) is not very insightful as it stands. However, the product of correlations θ_X and θ_b will almost always be negligible compared to the main terms in (2).[1] And if we reasonably estimate $\theta_b\theta x$ to be approximately zero, the equality in (2) simplifies to

$$(4) \qquad \rho^2_{\dot{x}\tilde{x}} \approx 1 - \frac{v_b^2 q_b}{1 + v_b^2 q_b} = (1 + v_b^2 q_b)^{-1} \ .$$

(Note that when $v_b^2 q_b$ is on the order of 10^{-1}, the approximate value of $\rho^2_{\dot{x}\tilde{x}}$ given by (4) is not much greater than the lower bound $1 - v_b^2 q_b$ on $\rho^2_{\dot{x}\tilde{x}}$ given in (2).) Ignoring minor perturbations from $\theta_b\theta_X$, then, we see that the effect of differential item weighting can be described by just two parameters, v_b^2 and q_b. The nature of v_b^2 is clear, so it only remains to elucidate q_b.

[1]For arbitrary weight selection, the expected value of θ_b^2 is $(m-1)^{-1}$. I have found it very difficult to develop analytic evaluations of θ_X's likely magnitude except for linearly independent items whose centroid is collinear with one of their principal axes, in which case $\theta_X = 0$. (One can contrive items having any stipulated value of θ_X; but what values are apt to arise in practice and how θ_X is affected by item orientations remain unclear to me.) Even so, it seems unlikely that in practice θ_X^2 will often be greater than, say, .2 or .3, especially if the items are convergently oriented (see below) with roughly equal variances.

Variance ratio q_b is not nearly so obscure as it may seem on first encounter; for its denominator $m\sigma_{\bar{x}}^2$ is just the X-configuration's saturation variance while its numerator, $\sigma_{d_b}^2$ is one axis' worth of dispersion variance selected by weights $\{b_i\}$. Since not all D-axes generally have the same variance, the particular $\{b_i\}$ makes some difference for $\sigma_{d_b}^2$ and hence for q_b; but even so, $\sigma_{d_b}^2$ has upper and lower bounds determined just by the D-configuration's shape regardless of $\{b_i\}$. The nature of these will be readily grasped by anyone familiar with the effect of orthonormal rotations on an item configuration's variance structure: Total variance remains invariant, but is redistributed among the rotated axes in ways that can best be described in terms of the configuration's principal components and their associated variances. Specifically, any axis d_b in any orthonormal rotation of $D = \langle d_1, \cdots, d_m \rangle$ is a linear composite $d_b = \sum_{i=1}^m a_i^* d_i^*$ of the principal components $d_1^*, \cdots, d_m^*$ of the D-configuration while the variance of d_b is a corresponding weighted average $\sigma_{d_b}^2 = \sum_{i=1}^m a_i^{*2} \lambda_{Di}$ ($\sum_{i=1}^m a_i^{*2} = 1$) of $\lambda_{D1}, \cdots, \lambda_{Dm}$ where each λ_{Di} is both the variance of d_i^* and the i^{th} root of the D-configuration's covariance matrix. In the present case, moreover, $\lambda_{D_m} = 0$ while the weight of d_m^* in $d_b = \sum_{i=1}^m a_i d_i = \sum_{i=1}^m a_i^* d_i^*$ ($a_i =_{\text{def}} (b_i - \bar{b})/m^{1/2}\sigma_b,\ \sigma_b > 0$) is zero.

Sketch of proof. Since $\sum_{i=1}^m d_i = 0$, $d_m^* = \sum_{i=1}^m m^{1/2} d_i$ has zero variance and is the D-configuration's m^{th} principal component. Moreover, since the vector of coefficients in $d_b = \sum_{i=1}^m ((b_i - \bar{b})/m\sigma_b) d_i$ is orthogonal to the vector of coefficients in $d_m^* = \sum_{i=1}^m m^{-1/2} d_i$, there exists an orthonormal rotation $d_1', \cdots, d_m'$ of the d_i in which $d_m' = d_m^*$ while d_b is one of $d_1', \cdots, d_{m-1}'$. From there it is straightforward to show that $d_1^*, \cdots, d_{n-1}^*$ are an orthonormal rotation of $d_1', \cdots, d_{n-1}'$ and conversely.□

Consequently, $\sigma_{d_b}^2$ is a weighted average of $\lambda_{D_1}, \cdots, \lambda_{D_{(m-1)}}$ and is bounded from above by λ_{D_1},and from below by $\lambda_{D_{(m-1)}}$.

Any value for $\sigma_{d_b}^2$ within this range can be selected by an appropriate choice of $\{b_i\}$ given any assigned nonzero value of v_b^2.

What do these constraints on $\sigma_{d_b}^2$ imply about q_b? Since $\lambda_{D_m} = 0$, $\lambda_{D_1}, \cdots, \lambda_{D_{(m-1)}}$ must sum to V_D; but how the λ_{D_i} otherwise partition that sum among themselves is entirely up to the shape of the item configuration. If this allocates equal dispersion variance to all directions in D-space, then $\lambda_{Di} = V_D/(m-1)$ for all $i = 1, \cdots, m-1$, and $\sigma_{d_b}^2$ also equals $V_D/(m-1)$ regardless of $\{b_i\}$. On the other hand, the D-configuration's variance ellipsoid can come in any degrees of eccentricity up to the extreme where λ_{Di} equals V_D for $i = 1$ and 0 for $i > 1$. Regardless of how V_d is distributed across $\lambda_{D1}, \cdots, \lambda_{D(m-1)}$, however, the value of d_b to be expected from a random choice of item weights remains $V_D/(m-1)$.

These considerations urge that a predictor configuration's sensitivity to differential item weighting can be concisely described in terms of (a) the value of $\rho^2_{\tilde{x}\tilde{x}}$ that arises when $\sigma^2_{d_b}$ is in the vicinity of its value expected under random selection of d_b, and (b) the smallest value of $\rho^2_{\tilde{x}\tilde{x}}$ that can occur when $\sigma^2_{d_b}$ approaches its maximum possible value of λ_{D_1}. (The upper bound on $\rho^2_{\tilde{x}\tilde{x}}$ that similarly follows when $\sigma^2_{d_b} = \lambda_{D(m-1)}$ will usually be too close to unity for v_b^2 on the order of 1 or less to hold much interest.) Regarding (a), since $\text{sat}_X = \text{hom}_X + (1-\text{hom}_X)/m$ and hence $\text{disp}_X = 1-\text{sat}_X = (1-\text{hom}_X)(m-1)/m$, while $1-\alpha_X = (\text{sat}_X - \text{hom}_X)/\text{sat}_X = (1-\text{hom}_X)/m\,\text{sat}_X$, plugging $\sigma^2_{d_b} = V_D/(m-1)$ into (3) gives

$$\mathcal{E}_X[q_b] = \frac{V_D}{(m-1)m\sigma^2_{\tilde{x}}} = \frac{\text{disp}_X}{(m-1)\text{sat}_X} = 1-\alpha_X \tag{5}$$

(random d_b), where $\mathcal{E}$ is subscripted with parameters on which the expectation is conditional. (Subscript "X" in this context means a given item configuration, or more specifically fixed C_{XX}.) Hence from (3)-(5), replacing q_b in (4) by its expectation under random selection of d_b

$$\mathcal{E}_{X,v_b}[\rho^2_{\tilde{x}\tilde{x}}] \approx [1 + v_b^2(1-\alpha_X)]^{-1} \approx 1 - v_b^2(1-\alpha_X) \tag{6}$$

(random d_b), where the simpler approximation in (6) is virtually as good as the other if v_b^2 is no greater than the order of 10^{-1}.

As for (b), although the upper bound $\lambda_{D_1}/m\sigma^2_{\tilde{x}}$ on q_b that follows from $\sigma^2_{d_b}$'s maximum λ_{D_1} is easy to compute numerically from the X-configuration's covariance matrix C_{XX}, the algebraic character of this bound is not especially perspicuous. (It is mildly insightful to note that $\lambda_{D_1}/m\sigma^2_{\tilde{x}}$ equals $\text{disp}_X/\text{sat}_X$ times the proportion of the D-configuration's total variance accounted for by its first principal axis. But it takes some practice to think in terms of those quantities.) However, if λ_{X_i} $(i = 1, 2, \cdots)$ is the variance of the X-configuration's i^{th} principal component, i.e. the items' total variance accounted for by their i^{th} principal axis, it can be seen that $\lambda_{X_1} \geq m\sigma^2_{\tilde{x}}$ while almost always $\lambda_{D_1} \geq \lambda_{X_2}$. Hence

$$q_b \leq \max_X[q_b] = \frac{\lambda_{D_1}}{m\sigma^2_{\tilde{x}}} \geq \frac{\lambda_{X_2}}{\lambda_{X_1}} \tag{7}$$

The rightmost inequality in (7) is not particularly helpful when the X-items have been oriented without regard for item convergence. But if the items have been oriented to maximize hom_X, or approximately so, the X-configuration's centroid will usually correlate highly with its first principal component, in which case $m\sigma^2_{\tilde{x}}$ and λ_{D_1} are well-approximated by λ_{X_1} and λ_{X_2} respectively, and the rightmost inequality in (7) becomes an approximate identity. Hence from (7) and (4), given well-chosen item orientations, the smallest value to which $\rho^2_{\tilde{x}\tilde{x}}$ can be driven by an unfavorable d_b is approximately

$$\text{(8)} \qquad \min_{X,v_b}[\rho^2_{\dot{x}\tilde{x}}] \approx \left[1 + v_b^2\left(\frac{\lambda_{X_2}}{\lambda_{X_1}}\right)\right]^{-1}$$

(strictly convergent X-configuration), where the X-configuration is "strictly convergent" just in case $\hom_x$ cannot be increased just by reflecting some of the items. (For a more detailed discussion of item convergence, see Rozeboom (1966, p. 344ff).) In the ideal special case where $\tilde{x}$ is perfectly collinear with the X-configuration's first principal component, (4), (8), and the left-hand approximation in (6) are all strict identities while also $\lambda_{Di} = \lambda_{Di+1}$ for $i = 1, \cdots, m-1$. (That is because, in this ideal case, each d_i differs only by an additive constant from the component of x_i orthogonal to the first principal component of $x_l, \cdots, x_m$ (see footnote on p. 372 of Rozeboom, 1966) whence $\theta_X = 0$ and the remaining principal components of the x_i are then the first $m-1$ principal components of the d_i.)

From the foregoing, we can see with easy clarity how the similarity $\rho^2_{\dot{x}\tilde{x}}$ between weighted and unweighted item composites $\dot{x} = \sum_{i=1}^m b_i x_i$ and $\tilde{x} = m^{-1}\sum_{i=1}^m x_i$ is determined jointly by (a) the initial choice of item scales and orientations, (b) the extremity v_b of differential item weighting $\{b_i\}$ in $\dot{x}$, and (c) the axis of the items' dispersion configuration selected by $\{b_i\}$. Regarding (a), if the item orientations and scale units are chosen without heed for the global properties of the X-configuration that so results, $\hom_X$, sat_X, and α_X should tend to be in the vicinity of 0, $m-1$, and 0, respectively. If so, the value of $\rho^2_{\dot{x}\tilde{x}}$ to be anticipated from random weighting is the same as it is when the X-configuration is orthonormal, namely, given $C_{XX} = I$,

$$\text{(9)} \qquad \rho^2_{\dot{x}\tilde{x}} = (1 + v_b^2)^{-1}$$

(which follows from what has been said above by virtue of the axes in any orthonormal rotation of an orthonormal item configuration being also orthonormal). But whereas (9) is an exact equality for uncorrelated, equal-variance items regardless of d_b, when the X-configuration's variance structure is appreciably elliptical even though $\hom_X \approx 0$ due to uncoordinated item orientations, q_b in (4) can easily be one or more orders of magnitude larger than unity with $\rho_{\dot{x}\tilde{x}}$ correspondingly near zero unless v_b^2 is on the order of 10^{-2} or less. High $\rho_{\dot{x}\tilde{x}}$ is still possible in this case even for appreciable v_b^2, but there is now no general tendency for $\tilde{x}$ to well-approximate $\dot{x}$.

However, it is common practice—and rightly so—to scale predictor items to have roughly equal variances, at least within the same order of magnitude, and also to choose orientations that more or less minimize negative item correlations. Together, these operations push sat_X close to the maximum value this can attain just through adjustments of the items' scale units and orientations. (This maximum equals the proportion of the items' total variance accounted for by their first principal axis under equal-variance item scaling. See Rozeboom (1966, p. 592).) In psychometric practice, sat_X and hence hom_X are then likely to be substantially larger than .1, possibly as great as .4 or .5; whence if the predictor items are fairly abundant, say $m >> 10$, $1 - \alpha_X$ will be little if at all greater than .1 so that v_b^2 can be as large as the order of 1 and still leave a high expectation in (6).

As for weighting extremity, v_b, formulas (2, 4, 6, 8, 9) are entirely clear about how this affects $\rho_{\dot{x}\tilde{x}}$ but say nothing about how large v_b^2 is likely to be. So long as the b_i can be negative as easily as positive, $\bar{b}$ can be arbitrarily close to zero and v_b^2 hence arbitrarily large. On the other hand, if virtually all the b_i have the same sign, say positive, v_b^2 will be on the order of 1 or less unless (as, however, can readily occur) the weight distribution has a strong positive skew. In particular, if $\{b_i\}$ is uniformly distributed over an interval of width w and midpoint c, $v_b^2 = (w/c)^2/12$, which is less than .34 when no weights are negative. Just the same, when $\{b_i\}$ comprises the items' regression coefficients for an outside criterion y, it does not seem reasonable to expect v_b^2 to be much smaller than .1 or .2, even when each x_i is oriented to have positive coefficient, unless item scales have been chosen to align $\tilde{x}$ with an estimate of y's projection into X-space.

If $\rho_{\dot{x}\tilde{x}}^2$ were always well-approximated by its expectation (6), we could conclude that for a decently homogeneous item configuration of appreciable size, differential item weights have little effect unless the weighting extremity is very large. Thus if $m = 20$ and v_b^2 is less than .5, the approximate expectation of $\rho_{\dot{x}\tilde{x}}^2$ is over .92 if $\mathrm{hom}_X = .20$ and over .96 if $\mathrm{hom}_X = .40$. And for *some* item configurations, (6) is indeed reliable. Specifically, this will be so if the X-configuration is strictly convergent, or nearly so, *and* is strongly dominated by a single factor, i.e. if λ_{X_2} is only a minor fraction of λ_{X_1}. However, if one or more roots of C_{XX} after its first are nearly as large as λ_{X_1}, then it is possible for b_i to select a D-axis d_b for which $q_b \approx 1$ and hence $\rho_{\dot{x}\tilde{x}}^2 \approx (1 + v_b^2)^{-1}$. To be sure, that same configuration's D-space also undoubtedly contains other directions for which q_b is so small that $\rho_{\dot{x}\tilde{x}}^2$ is near unity even for extreme v_b^2; but for a worst-case analysis, only maximal q_b is relevant.

Let me review these results, starting with their motivation. The aim is not to determine the correlation between weighted and unweighted item composites $\dot{x}$ and $\tilde{x}$ for any specific selection of item weights, since numerically that is a simple computation from b_i and C_{XX}. Neither is it a search for lower bounds on $\rho_{\dot{x}\tilde{x}}^2$; for when the predictor items span an m'-dimensional space (where m' may or may not equal the number of items), item weighting can put $\dot{x}$ anywhere within the $m' - 1$ dimensional item-space that is orthogonal to $\tilde{x}$. Rather, the intent is to develop a generalized *insight* into how $\rho_{\dot{x}\tilde{x}}^2$ is constituted out of $\{b_i\}$ and the item configuration's variance structure; specifically, to see whether there may not be a small number of abstract properties of $\{b_i\}$ and C_{XX} (which jointly contain $m(m+1)/2$ independent parameters) that not only suffice to determine $\rho^2_{\dot{x}\tilde{x}}$, or approximately so, but are also conceptually meaningful. The present analysis makes such insight available at three levels of accuracy.

At the lowest accuracy level, we find that $\rho^2_{\dot{x}\tilde{x}}$ is largely determined by only two parameters, one a property just of $\{b_i\}$ and interpretable as the extremity with which these differentially weight the items, and the other a property just of C_{XX} that is not very intuitive in its own right but has become classic in test theory as a measure of internal consistency. Specifically, from (6),

$$\rho^2_{\dot{x}\tilde{x}} \approx [1 + v_b^2(1-\alpha_X)]^{-1}, \tag{6'}$$

where the X-configuration's alpha coefficient may be viewed as the items' homogeneity amplified by their numerosity as set forth in (1). The main term in (6′) is v_b, simply because we can make its numerical value as large as we please by our choice of item weights. (In contrast, we have only limited control over α_X, though it is important to appreciate how our choice of item orientations and scale units influence α_x through their effect on hom_X.) Thus (6′) may be viewed as telling how $\rho^2_{\dot{x}\tilde{x}}$ decreases as a single-parameter function of weighting extremity v_b, while that function's parameter, $1-\alpha_X$, is the item configuration's *sensitivity* to differential weighting.

Approximation (6′) is lucid, powerful, and often highly accurate. Even so, depending on properties of C_{XX} additional to α_X and on whether b_i takes advantage of them, (6′) can be seriously misleading for a particular b_i. If the item scales and orientations, especially the latter, are chosen more or less to maximize hom_X, and all principal axes of X after the first are approximately equal in the amount of total X-variance each accounts for, then approximation (6′) is for all practical purposes an identity over all possible choices of item weights. (Note that this is true regardless of how weak the items' first principal axis may be.) However, if the item configuration also contains principal axes after the first that have secondary prominence, analysis of $\rho^2_{\dot{x}\tilde{x}}$ at a higher accuracy level must revert to approximation (4). This has the same form as (6′), but replaces $1-\alpha_X$ in the latter by a sensitivity parameter q_b that can be viewed as a selection by b_i from a range of sensitivities (centered on $1-\alpha X$) made available by the item configuration. If the items have arbitrary orientations even though some of their intercorrelations are appreciable, q_b can easily be one or more magnitudes greater than unity, in which case formula (4) is not particularly useful. But if the items have been oriented to strict convergence, or nearly so, then the items' maximum sensitivity to weighting is approximately equal to $\lambda_{X_2}/\lambda_{X_1}$, i.e. the strength of the items' second principal component compared to that of their first, and a worst-case analysis should proceed in terms of (8). The practical difference between (6′) and (8) is that for a factorially complex item configuration, $\lambda_{X_2}/\lambda_{X_1}$ may well be in the vicinity of .5 or more even when, due to large m, $1 - \alpha_X$ is quite small.

Finally, at the highest accuracy level, $\rho^2_{\tilde{x}\tilde{x}}$ can be analyzed exactly, as in (2), for all item configurations, by appeal to two minor parameters θ_b and θ_X additional to v_b and q_o. Formula (2) is hard to interpret, however, and when the item configuration is close to strict convergence it seems likely that in practice the difference between (2) and (4) will seldom if ever be appreciable.

Formulas (6) and (8) give little reason to favor equal weighting in applied prediction, even though the two preconditions stipulated by Wainer (1976), (i) that all predictor items have the orientations that regression weighting would give them, and (ii) that none of the regression-oriented items are negatively correlated, do retard the general proclivity of $\rho^2_{\tilde{x}\tilde{x}}$ to approach zero. (Condition (i) implies that none of regression weights $\{b_i\}$ are negative, whence v_b^2 is likely to be on the order of 10^{-1} even though v_b^2 will still approach or exceed unity if a small proportion of the b_i exceed the remainder by an order of magnitude. And (ii) requires the item configuration to be strictly convergent, which does its best to maximize hom_X but insures neither that hom_X is high nor that q_b is much less than unity. Also, the higher the latent item homogeneity, the less likely it is that (i) and (ii) can be jointly satisfied in the first place for a given outside criterion.) Even so, present results do support a more restrained version of Wainer's thesis. This is simply that if items x_i are fairly numerous and at least modestly homogeneous when provisionally scaled to align $\tilde{x}$ as best we can with our target axis of X-space (i.e. when initial weights are temporarily absorbed into the item scales), then moderate readjustments of the items' provisional scale units, corresponding e.g. to rounding the items' raw-scale weights to a small number of alternatives such as $0, \pm 1, \pm 2, \pm 3, \pm 4$ (cf. Green, 1977, p. 270), will almost always leave the modified composite virtually indistinguishable, correlationally, from its precursor. To put the point bluntly, second-digit precision in item weighting is generally a waste of effort.

References

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*, 95–106.

Green, B. (1977). Parameter sensitivity in multivariate methods. *Multivariate Behavioral Research*, *12*, 263—288.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

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

Rozeboom, W. W. (1966). *Foundations of the theory of prediction.* Homewood, Illinois: The Dorsey Press.

Rozeboom, W. W. (1978). The logic of unboundedly reactive systems. *Synthese*, *39*, 435–530.

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