Listening. Learning. Leading.®

Research Report

ETS RR–14-20

What Differential Weighting of Subsets of Items Does and Does Not Accomplish: Geometric Explanation

James E. Carlson

December 2014

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

What Differential Weighting of Subsets of Items Does and Does Not Accomplish: Geometric Explanation

James E. Carlson

Educational Testing Service, Princeton, NJ

A little-known theorem, a generalization of Pythagoras's theorem, due to Pappus, is used to present a geometric explanation of various definitions of the contribution of component tests to their composite. I show that an unambiguous definition of the unique contribution of a component to the composite score variance is present if and only if the component scores are uncorrelated. I further show the effect of differentially weighting the composites on the definitions of unique contributions and discuss some of the implications for composite score reliability and validity.

Keywords Combining subscores; variance contributions; weighting subtests; vector geometry; Pythagoras's theorem; Pappus's theorem

doi:10.1002/ets2.12020

Differential weighting of item scores and subtest scores when constructing reporting scales is a topic of interest in several contexts in measurement (Haertel, 2006; Kolen, 2006; Lord & Novick, 1968, Chapter 4). One example occurs in the National Assessment of Educational Progress (NAEP) where a number of independently derived scales (e.g., five in mathematics) are combined to form the NAEP composite (N. L. Allen, Donoghue, & Schoeps, 2001). Another example that arises in discussions with assessment development staff is weighting constructed-response (CR) items more heavily than selected-response (SR) items (Lane & Stone, 2006; Sykes & Hou, 2003).

Test users and even measurement professionals may fail to comprehend the problems that can arise in interpreting the composite arising from this practice. One problem is that no unique way of defining the *contributions* of the subtest measures to the composite exists (Carlson, 1968). It is often assumed, for example, that weighting one subset of items by two and leaving another subset unweighted will double the contribution of the first set to the total score; this is not true by most definitions of contribution. In terms of contribution to variance, for example, doubling weights of a component score quadruples that component's contribution to the composite variance. This relationship is important because the more the contribution of a component to the variance, the more influence that component has on placing test takers higher or lower in the composite score distribution. This fact does not seem to be well known to many assessment professionals. Statisticians have had difficulty explaining to the user community issues with defining contributions to total score variance in the combining of measures, which is particularly difficult to understand if the measures being combined are correlated (see Carlson, 2014, for a discussion of similar issues in using regression models).

The objective of this article is to introduce the measurement community to geometric explanations of contributions and effects of differential weighting of subsets of test items; the geometric explanations often are more intuitive than algebraic explanations. The geometric explanation includes results based on a little-known theorem of Pappus (D. G. Allen, 2000; Kazarinoff, 1961) that is a generalization of Pythagoras's theorem to nonright triangles. Application of this theorem to the issue of weighting subsets of items helps explain effects of the weighting on contributions of the subtests comprising those subsets to the overall test score variance.

The article includes a demonstration of how Pythagoras's and Pappus's theorems relate to the combination of weighted and unweighted subtest scores into a composite score and how this relates to the issues of interpretation of the total scores, contribution to the total score variance, and effects on reliability and validity. The demonstration uses artificial data designed to emulate a test comprising examinee responses to two subtests.

Corresponding author: J. Carlson, E-mail: jcarlson@ets.org

Algebraic Expressions for Contributions

Although discussions of contributions of component scores to a composite have been provided earlier (Guttman, 1941; Horst, 1941; Richardson, 1941; Wilks, 1938), the works most relevant to this presentation (specifically Chase, 1960; Creager & Valentine, 1962; Guilford, 1965; Richardson, 1941) have presented methods of defining such contributions and weighting the components prior to combining them.

The Creager–Valentine (CV) Procedure

Creager and Valentine (1962) noted that if one regressed a composite onto its p components, the *squared multiple correlation coefficient* (R^2) would be 1.0 because there is perfect prediction of the composite from the components. This R^2 statistic is equal to the proportion of variance accounted for by all components being considered. Creager and Valentine then defined the *proportional contribution* of the jth component to the *composite variance* as:

$$U_j = 1.0 - R^2_{c,-j}, \tag{1}$$

where $R^2_{c,-j}$ is the squared multiple correlation for prediction of the composite from all components except the jth. This definition considers the (proportional) unique contribution of a component to be equal to the increase in accounted-for variance after the contributions of all other components have been considered. As such, if the components are correlated, it is clearly a very conservative estimate of contribution, especially if a large number of components are present.

To explain, consider an analogous situation (not involving component scores and a composite), stepwise regression, in which a dependent variate is to be predicted from some combination of p predictors. The method begins with a one-predictor model yielding an R^2 statistic that is simply the square of the correlation between that predictor and the dependent variate. Then, a second predictor is added to form a two-predictor model with a second (mathematically it must be equal to or larger than the first) R^2 statistic, and the difference is often (Draper & Smith, 1966) referred to as the proportional contribution of the second predictor to the prediction. The problem is that this so-called contribution depends on the order of entering predictors into the prediction equation. Suppose we have scores, Y, on a dependent variate and scores, X_1 and X_2, on two predictors, with correlations between Y and X_1, Y and X_2, and X_1 and X_2, respectively, of:

$$r_{y1} = .50,\ r_{y2} = .60,\ r_{12} = .55.$$

The prediction equation for predicting scores, Y, from scores on X_1 alone, or from scores on X_2 alone would yield:

$$X_1 \text{ alone} : R^2_{y.1} = r^2_{y1} = .50^2 = .25$$

or

$$X_2 \text{ alone} : R^2_{y.2} = r^2_{y2} = .60^2 = .36,$$

respectively. We would say that X_1 scores predict 25% of the variance in Y scores, whereas X_2 scores predict 36%. Now, the squared multiple correlation, hence the proportional contribution for the two-predictor equation, is (see, e.g., Guilford, 1965, p. 394),

$$\begin{aligned} R^2_{y.12} &= \frac{r^2_{y1} + r^2_{y2} - 2r_{y1}r_{y2}r_{12}}{1 - r^2_{12}} \\ &= \frac{.50^2 + .60^2 - 2 \times .50 \times .60 \times .55}{1 - .55^2} \\ &= \frac{.25 + .36 - .33}{1 - .3025} = \frac{.28}{.6975} = .40. \end{aligned}$$

Hence, 40% of the Y score variance is predictable from the combination of the two predictors. So, if X_1 scores are entered first into a stepwise regression, its contribution to the prediction of Y score variance will be said to be 25%, whereas if it is entered second, its contribution will be only $.40 - .36 = .04$, or 4%. Similarly if X_2 scores are entered first, it will be said to contribute 36%, whereas if it is entered second, the contribution would only be .40 − .25 = .15, or 15%. So the contribution of a predictor is highly dependent on when it is entered into the equation, its correlation with other predictors, and the correlations of predictors with the dependent variate.

But, note that if the two predictors are correlated zero, $R^2_{y.12}$ reduces to .25 + .36 = .61, simply the sum of the contributions from the two one-predictor equations. In this case the contributions of the two predictors are 25% and 36%, respectively, independently of the order of being entered into the equation. The point is that *if and only if predictors are orthogonal (uncorrelated) there is an unequivocal definition of contribution to the variance of the dependent variate* (see Carlson, 2014, for more details).

The Chase–Guilford (CG) Procedures

Now we return to the topic of "contributions" of subscores to the total score, with an alternative definition to the CV procedure. Chase (1960) defined the *proportional variance contributions* as:

$$V_j = \left(b_j^*\right)^2, \tag{2}$$

where b_j^* is the *standardized* regression coefficient for the jth component in the prediction of the composite from p *standardized* components; that is, the scores are transformed to *standardized scores* by subtracting the means and dividing by the standard deviations. The composite is the sum of the p standardized components. Carlson (1968) showed that, for this case, each term in Equation 2 may be expressed as the ratio of the component score variance to the composite variance. Hence Equation 2 may be expressed as:

$$V_j = \left(b_j^*\right)^2 = \frac{s_j^2}{s_c^2}. \tag{3}$$

Thus, when combining subscores in this metric, each of the proportional unique contributions by Chase's definition can be interpreted as the ratio of the component variance to the composite variance. Chase's definition is based on Equation 4,

$$R^2 = \sum_{j=1}^{p} \left(b_j^*\right)^2 + 2\sum_{j=1}^{p} \sum_{j' \neq j=1}^{p} b_j^* b_{j'}^* r_{jj'}, \tag{4}$$

where $r_{jj'}$ is the correlation between the j th and j' th components and, in the context of component and composite scores, $R^2 = 1.0$, as stated previously. Wright (1934, cited by Bock, 1975, p. 381), in the regression case, referred to the elements in the last term of Equation 4 as the "indirect contributions," and Bock (p. 381) cited Equation 4 for the two variable case. Chase also referred to the terms in the last part of Equation 4 as the joint contributions of pairs of components. For example,

$$J_{12} = 2b_1^* b_2^* r_{12}. \tag{5}$$

In addition, Chase (1960) referred to the *total contribution of a variable* as:

$$T_j = b_j^* r_{cj}, \tag{6}$$

where r_{cj} is the correlation between the composite and the j th component.[1] This definition is also a proportional contribution.

The Chase (1960) definition of total contribution is based on Equation A17 (see also, Bock, 1975, p. 380) for the total proportion of accounted-for variance in regression. In this context this expression is written as:

$$R^2 = \sum_{j=1}^{p} b_j^* r_{cj}, \tag{7}$$

where, as mentioned previously, in the case under consideration in this presentation, $R^2 = 1$. Guilford (1965), discussing contributions to regression rather than the combination of subtests, stated that the terms in Equation 7 "stand for both direct and indirect contributions" (p. 399) and also referred to the Chase contributions in Equation 2 as the "direct contributions" (p. 400). Bock (1975) also referred to the terms of Equation 2 as direct contributions. Finally, Guilford referred to the difference,

$$b_j^* r_{cj} - \left(b_j^*\right)^2, \tag{8}$$

as the indirect contribution. Importantly, Guilford noted that interpretation of these quantities as variance contributions assumes that all the $b_j^* r_{cj}$ products in Equations 6, 7, and 8 are positive. Mathematically, they could actually be zero, but then the contribution would also be zero. Bock correctly pointed out, "Only when the X variables are uncorrelated in the sample . . . are these terms nonnegative and do they represent proportions of predictable variation" (p. 380). As stated previously, this is one of the main themes of this report. Although both Bock and Guilford were discussing contributions in regression models, their remarks apply equally to the case of combination of subtests under discussion here.

Other Procedures

Besides the CV and CG definitions of unique contribution, there have been some other proposals for partitioning explained variance but, as in the case of Guilford's (1965) presentation, not always in the context of combining component scores into a composite. Creager and Valentine (1962) and Guilford took those two definitions and showed how the component scores could be weighted so they would contribute to the composite variance in a prescribed ratio.[2] Richardson (1941) used a different definition, later shown (Carlson, 1968) to be equivalent to the Chase (1960) total contribution, T_j in Equation 6.

Wilks (1938) discussed weighting the components such that the correlations between the weighted components and the composite were equalized, which would imply an interest in equalizing contributions of components where the contributions are defined as correlations with the composite. This is an unusual definition — most statisticians consider contribution to explanation of variation (usually defined as variance) to be the important statistic.

Using some algebraic expressions and definitions plus the related geometry discussed in Appendix A, in the next section, I show that no unique definition of the contribution of a component score to the composite exists except in the case of uncorrelated components, similar to the regression example discussed above. Otherwise, the different definitions result in different contributions of the subtests to the composite. I use an example of two subtests to simplify the presentation, but it is generalizable to any number of components.

Geometry of Variance Contributions

Because of the context, I use as an example the case in which a set of examinees have responded to test items that may be considered as two subtests. To add further context, I assume that one subtest comprises dichotomously scored items and the other polytomously scored items. This example is one that I have encountered in practice because, as noted earlier, test developers and their clients often have questions about differentially weighting these two types of items in forming a composite score for each test taker.

Several writers (Draper & Smith, 1966; Wickens, 1995; Wonnacott & Wonnacott, 1973) have shown relationships among variables using two distinct geometries, as discussed in Appendix A. Because of my context, I use the example of tests rather than variables in general.

Showing two tests as axes at right angles (orthogonal), with test taker's values on the tests represented by points in the two-dimensional space defined by the axes, is the more common of the two geometries. It is referred to as the *geometry in the variable space*. Measures of the spread of points parallel to the axis representing a test indicate the variability (such as range, standard deviation, and variance). In the case of two tests, the clustering of the points about a regression line that (by the least squares criterion) best fits the points, and the slope of that line, is related to the correlation between the two tests and the regression of one test onto the other.

The second geometry represents test takers as axes and the tests as points in an N-dimensional space, as discussed in Appendix A. This is the *geometry in the test-taker space,* illustrated in Appendix A (see Figure A1 and accompanying discussion). In measurement models, the different individual test takers are considered independent of one another, and orthogonality of the axes represents this independence because orthogonality is equivalent to zero correlation[3] (Wickens, 1995). Using this geometry (see Appendix A or Carlson, 2014, for details), the tests are usually displayed as vectors drawn

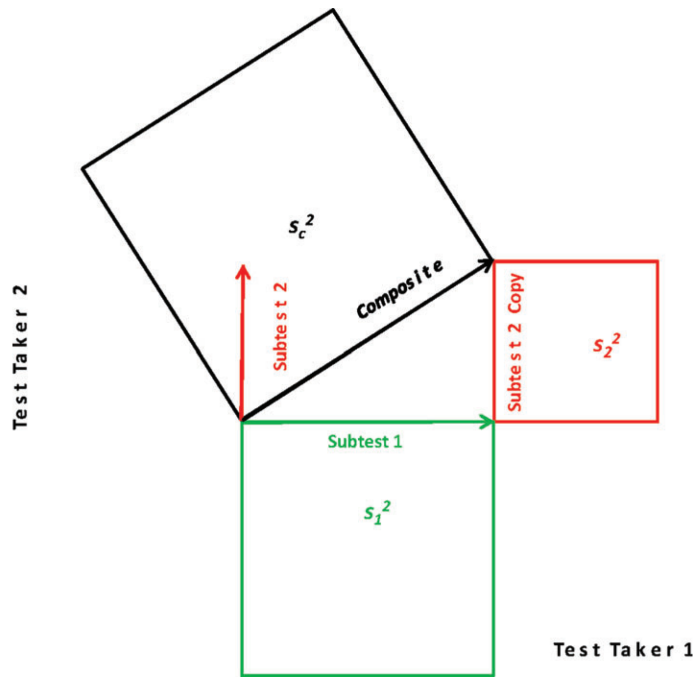

Figure 1 Orthogonal case — Pythagoras's theorem related to variance components; the areas of the squares are equal to the variances indicated.

from the intersection of the axes to the points (Figure A1). With N test takers we are dealing with N dimensions. As discussed below and in Appendix A, limiting discussion to two subtests only requires two dimensions (but all our discussion generalizes to more subtests). In this geometry, if the test scores are transformed to deviation metric (deviations from the mean), the length of each vector can be considered to be equal to the test's standard deviation.[4] Also, in this metric, the cosine of the angle between two test vectors is equal to the correlation between the two tests. The (unweighted) sum of two subtests (the unweighted composite in our discussion) is a vector in the same plane as the two subtests. This composite vector can be found by appending one subtest vector onto the end of the other and connecting the end of the appended vector to the origin, as shown in Figure A2.

As stated in Appendix A, because two vectors and their vector sum can always be represented as lying within a plane in the N-dimensional space, we can extend this discussion without reference to the N-dimensional space when we have N test takers. Thus, the geometry needed for this discussion is exactly the same for two dimensions or N dimensions.

Recalling that the cosine of 90° is zero, we may note that two subtests correlated zero will be represented as orthogonal vectors (at right angles) and the composite variance in that case is the simple sum of the two individual subtests' variances. Algebraically the composite variance is well known to be:

$$s_c^2 = s_1^2 + s_2^2 + 2s_1 s_2 r_{12}, \tag{9}$$

where s_1^2 and s_2^2 are the variances of the two subtests. The third term on the right is twice the covariance between the two subtests and, when they are correlated zero, the covariance is clearly zero and the composite variance reduces to the sum of the two individual subtest variances. Figure 1 illustrates Equation 9 geometrically. The component vectors, in deviation score metric, are shown as the green (Subtest 1) and red (Subtest 2) arrows at right angles to one another. The composite vector (black) is the hypotenuse of a right triangle, and its squared length is equal to the composite variance. Because the squared lengths of the other two sides are equal to the variances of the two subtests (because the scores are in deviation form), when the last term of Equation 9 is zero, this equation is simply Pythagoras's theorem, the square of the hypotenuse is equal to the sum of the squares of the other two sides. As often done in geometric demonstration of Pythagoras's theorem, I use squares constructed on the three sides of the right triangles and the areas of the squares are equal to the squared lengths of the sides. Thus, as shown in the Figure 1, the areas of the squares are equal to the variances of the two subtests and the composite (see also Wickens, 1995).

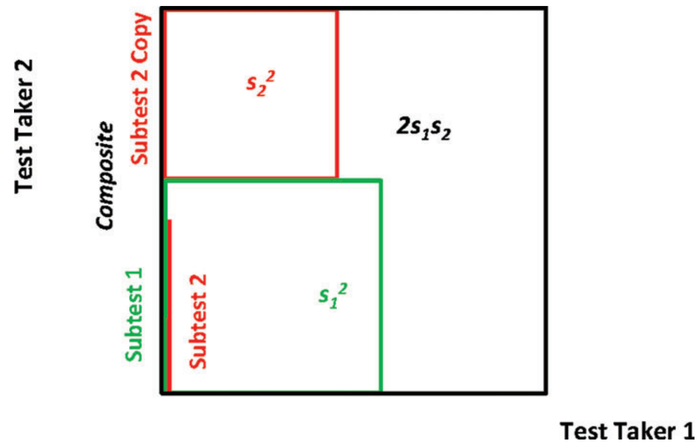

Figure 2 Subtests Correlated 1.0—three variance components.

Thus, in the case of uncorrelated subtests, using Pythagoras's theorem we see that each subtest contributes a unique amount to the composite variance, a contribution equal to its own variance, as discussed previously with respect to Equation 9. We can also show that, in this case of zero correlation between the subtests, several of the definitions of "contribution," including U, V and r_{cj}^2, are identical (Carlson, 1968, proved this). One of the main points we wish to make is that *if and only if the correlation between the subtests is zero there is no controversy about a definition of contributions to the composite variance*. In the context of subtests, of course, we would not expect nor want the subtest scores to be correlated zero because they are measures of a unidimensional construct and thus should be fairly highly related. Another interesting case is that in which the two subtests are correlated one. In this case, the two vectors and their composite are collinear (lie along the same line, see Appendix A) and from Equation 9, the composite variance is simply the sum of the two subtest variances plus twice the product of the two standard deviations because r_{12} is 1. If we construct squares on the two deviation score vectors and their composite, with areas thus equal to the variances of those three variables, the square on the composite will be larger than the sum of the two subtest squares by an amount equal to twice the product of the two standard deviations. This case is illustrated in Figure 2 (to simplify, Figure 2 is shown without the arrowheads that appear in Figure 1).

Returning to the usual case of subtests correlated somewhere between 0 and 1, we will now demonstrate the use of the little-known theorem of Pappus,[5] a generalization of Pythagoras's theorem to nonright triangles. Pappus's theorem involves parallelograms rather than squares (as in Pythagoras's theorem, illustrated in Figure 1), constructed on the sides of the triangle. But note that squares are special cases of parallelograms, and I use squares in most of my discussion. Figure 3 is used to explain the theorem.

Figure 3 shows the $\mathbf{x}_1$ vector (AQ) and the copy of the $\mathbf{x}_2$ vector (QB) as in Figure A2, representing two correlated subtests and their composite, $\mathbf{x}_c$ (AB). I am using deviation score metric as noted previously and in Appendix A. These vectors are shown without arrowheads to simplify the figure. I state the theorem in simplified language (see Carlson, 2014, for more detail). Referring to Figure 3, I start with parallelograms (actually, in our case they are squares) constructed on two sides of triangle ABQ (green and red squares on sides AQ and QB, respectively) and extend the sides of these squares parallel to the sides of the triangle until they meet (dashed blue lines) at point P. Then a parallelogram is constructed on the third side (blue, on side AB) with the two parallel sides that are *not part of the triangle* having length and direction equal to those of line PQ. The theorem states that the area of that parallelogram is equal to the sum of the areas of the original two parallelograms (green and red squares), hence equal to $s_1^2 + s_2^2$ in the metric under consideration. Also as discussed earlier, a square (ABFE, black in the figure) constructed on the third side represents the composite variance in Equation 9; hence its area is equal to $s_1^2 + s_2^2 + 2s_1s_2r_{12}$.

This geometry leads to different ways of considering contributions to the composite variance. Consider Figure 3. Using simple geometry, the area (equal to $s_1^2 + s_2^2$ by Pappas's theorem) of the parallelogram (blue) on the side of the triangle representing the composite can be seen to be equal to the area of rectangle ABDC (because the triangular portion outside of square ABFE, with one side DB, is clearly identical to the triangular portion inside the square with one side AC). That rectangle, ABDC is smaller in area than square ABFE (black) representing the variance of the composite. In fact,

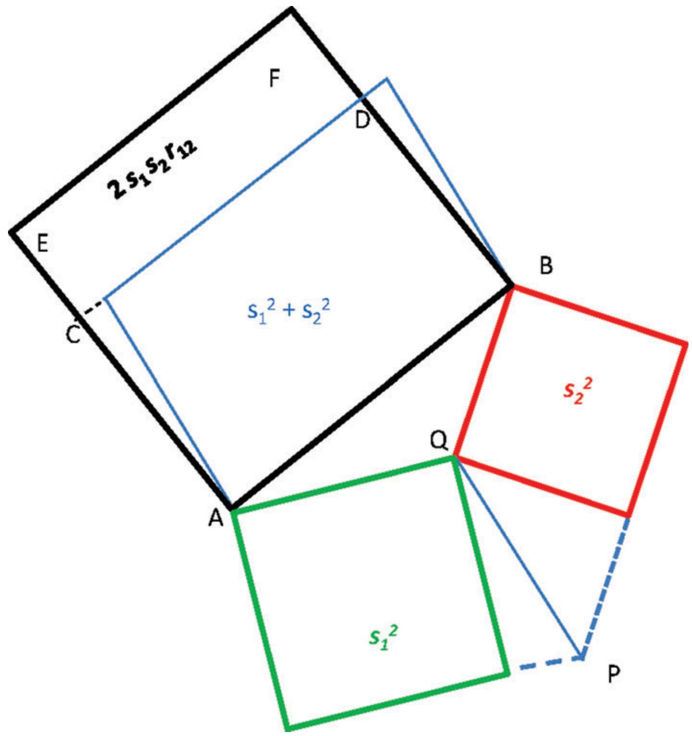

Figure 3 Pappus's theorem related to variance components in the nonorthogonal case.

because it is equal in area to the blue parallelogram, by Pappas's theorem, it is equal to the variance component, $s_1^2 + s_2^2$. The remaining rectangle in that square, CDFE, also represents a variance component. From Equation 9 we can see that it is equal in area to twice the covariance between the subtests ($2s_1s_2r_{12}$).

I now turn to another aspect of the variance contributions represented by the geometry of Figure 3. Figure 4, constructed from Figure 3, with some lines deleted to simplify, is used to illustrate.

Geometrically, in Figure 4, as discussed previously and in Appendix A, lines AQ (representing $\mathbf{x}_1$), QB (representing $\mathbf{x}_2$), and AB (representing $\mathbf{x}_c$) have lengths equal to s_1, s_2, and s_c, respectively. Lines AR and RB are resultants of the perpendicular projections of $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, onto $\mathbf{x}_c$. Using, for example, cos $\angle$ RAQ to represent to cosine of the angle at A, the lengths of these projections, denoted *Ln*(), can be seen to be,

$$\begin{aligned} Ln\ (\mathrm{AR}) &= Ln\ (\mathrm{AQ}) \cos \angle \mathrm{RAQ} = \mathrm{s}_1 r_{c1} \\ Ln\ (\mathrm{RB}) &= Ln\ (\mathrm{QB}) \cos \angle \mathrm{RBQ} = \mathrm{s}_2 r_{c2}. \end{aligned} \tag{10}$$

The final terms of Equation 10 are products of standard deviations and correlations because the lengths of the lines are equal to the standard deviations of the vector variables, and as stated previously, the cosines of the angles between pairs of vectors are equal to the correlations between the variables represented by the vectors. Therefore squares constructed on these two projections (dashed green and red squares) have areas equal to the variance contributions, $s_1^2 r_{c1}^2$ and $s_2^2 r_{c2}^2$. Note that the length of AB is s_c because AB is the vector of composite scores, and it is made up of the two projected lengths, $s_1 r_{c1}$ and $s_2 r_{c2}$, so we have the relationship:

$$s_c = \left(s_1 r_{c1} + s_2 r_{c2}\right), \tag{11}$$

and therefore the area of the black square in the figure is the variance,

$$\begin{aligned} s_c^2 &= \left(s_1 r_{c1} + s_2 r_{c2}\right)^2 \\ &= s_1^2 r_{c1}^2 + s_2^2 r_{c2}^2 + 2 s_1 s_2 r_{c1} r_{c2}. \end{aligned} \tag{12}$$

Hence we have in Figure 4 another partitioning of the variance of the composite into three components—the two squares and the irregularly shaped figure within the black square (I am using W^2 to represent the single-variable variance

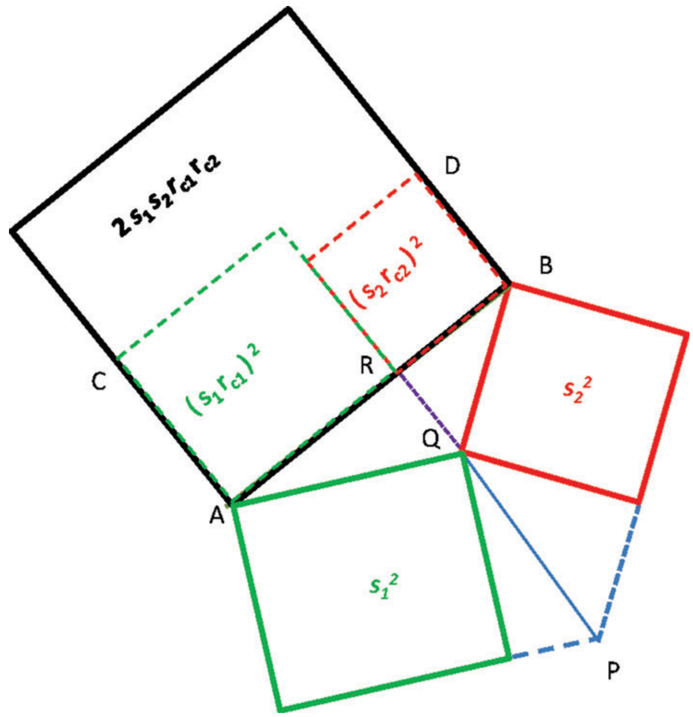

Figure 4 Additional geometry of variance partitioning.

contributions and K for the joint contribution by this definition):

$$W_1^2 = s_1^2 r_{c1}^2 \ : \ \text{which might be called a contribution of } \mathbf{x}_1,$$
$$W_2^2 = s_2^2 r_{c2}^2 \ : \ \text{which might be called a contribution of } \mathbf{x}_2, \text{ and}$$
$$K_{12} = 2 s_1 s_2 r_{c1} r_{c2} \ : \ \text{which might be called a joint contribution.} \quad (13)$$

To express these new definitions of contributions as proportions, we would divide each by s_c^2.

For the sake of completeness, we note (see Carlson, 2014) that if the subtests were correlated negatively, angle AQB in Figure 4 would be less than 90° and the “unique contributions” of the subtests would sum to more than the composite variance, leaving negative “joint” variance components, which are statistically impossible (they would indicate imaginary variables). For example, one can see from Figure 4 that with such a configuration the projections discussed above would overlap and the sum of their lengths would be greater than the composite standard deviation leading to contributions summing to more than the total variance. This issue is related to Bock’s (1975) earlier referenced statement that all the *br* terms in Equation 6 must be nonnegative in defining the variance contributions. It is highly unlikely, however, that subtests of a unidimensional assessment instrument would be negatively correlated, so this is probably a moot point despite the fact that it calls into question the way variance contributions are defined.

The result of the algebra and geometry discussed in this article is that, along with Pappus’s theorem and the figures in this article, we can show a number of different ways to define the variance contributions of the subtests to the composite:

1. The Creager and Valentine (1962) definition based on fitting different regression models as in Equation 1.
2. That due to Chase (1960) and Guilford (1965) based on the V_1 and V_2 *unique contributions* in Equations 2 and 3, and the remainder representing that part of the explained variance not attributable uniquely to the subtests, sometimes referred to as a *joint contribution* as in Equation 5.
3. That based on Equation 9, illustrated using Pappus’s theorem in Figure 3.
4. That based on Equation 12, illustrated using Pappus’s theorem in Figure 4.
5. That based on the notion of contribution proportional to correlation, represented geometrically by the angles between the subtests and the composite, with the cosines of those angles equal to the correlations.

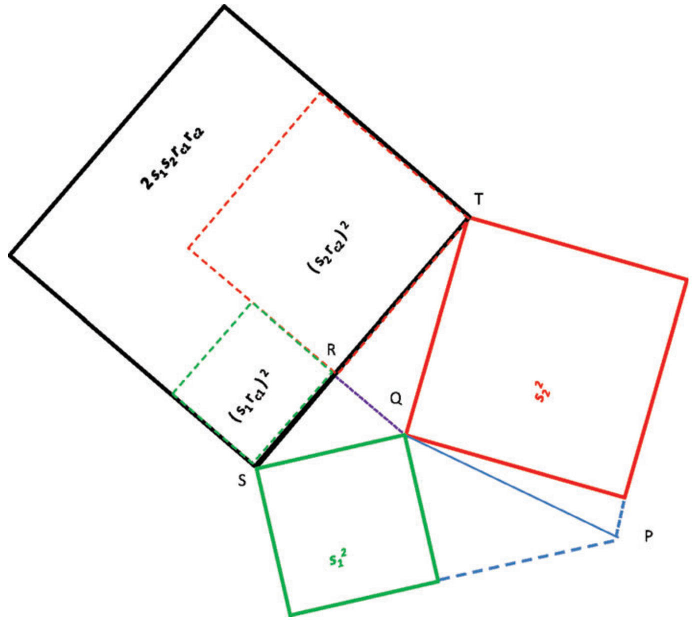

Figure 5 Effect of weighting Subtest 2 by a factor of 2.

Table 1 Comparison of Unweighted and Weighted Statistics

Statistic	Unweighted $C = X_1 + X_2$			X_2 weighted double $C = X_1 + 2X_2$		
	X_1	X_2	C	X_1	X_2	C
s	4.477	3.034	7.045	4.480	6.070	9.881
s^2	20.042	9.208	49.629	20.042	36.832	97.632
r_{jc}	0.959	0.907		0.914	0.954	
r_{jc}^2	0.919	0.823		0.835	0.910	
U	0.177	0.081		0.090	0.165	
V	0.404	0.186		0.205	0.377	
W	4.291	2.753		4.091	5.790	
W^2	18.416	7.581		16.735	33.524	

Effect of Weighting

Next, we consider the effect of differential weighting on the various notions of variance contribution. Suppose that the green vector in the figures is the vector of scores on the SR (e.g., multiple choice) items on the test and the red vector is the vector of scores on the constructed-response (CR) items. Suppose also that a decision is made to weight the CR scores by 2.0 such that each CR item "receives twice the weight of each SR item." The geometric effect is to double the length of the subtest-two vector in the previous figures, as shown in converting Figure 4 to Figure 5.

Comparison of Figures 4 (unweighted) and 5 (weighted) reveals the following:

- The Subtest 1 variance is unchanged, whereas the Subtest 2 variance is much larger (a weight of 2 quadruples the variance, see Carlson, 1968).
- The correlation between Subtest 1 and the composite is decreased, whereas that of Subtest 2 is increased but not by a factor of 2.

Comparison of statistics of the weighted to unweighted cases provides more specific information. Table 1 shows statistics from the case of unweighted and weighted composites, using the artificial data provided in Appendix B. The subtests are represented by X_1 and X_2 and the composite by C.

It may be seen that double weighting Subtest 2 (X_2) while not weighting Subtest 1 has different effects on the statistics used by various definitions of the unique contribution of the subtests to the composite test, as follows:

- The U-statistic for Subtest 2 is slightly more than doubled (ratio of 2.033:1), but at the same time, that for Subtest 1 is nearly halved (0.508:1). The ratio of contributions of Subtest 2 to Subtest 1 by this definition changes from 0.459:1 to 1.838:1 by weighting by 2, so the ratio of these contributions is not doubled but quadrupled.
- The V-statistic behaves exactly the same as the U-statistic with the same ratios.
- The W-statistic for Subtest 2 is slightly more than doubled (2.103:1), but that for Subtest 1 is decreased slightly (.953:1). The ratio of contributions of Subtest 2 to that of Subtest 1 changes from .642:1 to 1.415:1, a factor of 2.206.
- The ratio of Subtest 2 to Subtest 1 correlations with the composite changes from 0.947:1 to 1.044:1, so the weighting does not have a doubling effect with respect to this definition of contribution.
- Similarly the ratio of the Subtest 2 to Subtest 1 squared correlations with the composite changes from 0.896:1 to 1.090:1, which does not represent a doubling.
- The ratio of standard deviations of Subtest 2 to Subtest 1 does exactly double, from 0.678:1 to 1.356:1.
- The ratio of variances does quadruple from 0.459:1 to 1.838:1, as expected.

So it may be seen that the only statistic that could be considered as a definition of *contribution* that is doubled when scores on one set of items is doubled is the standard deviation of the resulting subtest, which has seldom, in the writers' experience, been used as a definition of *contribution*. Yen (1983), however, in one relevant discussion, did explicitly consider contributions to standard deviations:

> Content area total and battery total scores were taken to be averages of the trait estimates contributing to the total. The different content area scales had been chosen to minimize variability *in the standard deviations* [italics added] of the scales, because these standard deviations affect the implicit weighting in obtaining area totals. (p. 134)

The U- and W-statistics for Subtest 2 were nearly but not exactly doubled, but those for Subtest 1 were nearly halved, hence changing the ratio of subtest contributions by these statistics by a factor of 4 rather than 2.

Summary and Discussion

Variance Contributions and Effects of Weighting

I have discussed several different definitions that have been proposed as representing the *contribution* of a subtest to the composite composed of a combination of the subtests. Furthermore, these definitions have been demonstrated to yield values, and hence conclusions, that will almost always differ from one another. I have shown that doubling one subtest's scores does not generally result in a doubling of the contribution by any of the proposed indices of contribution. Finally, I have shown that only in the case of uncorrelated subtest scores, which should not occur in practice in the context of educational assessment, can we unequivocally report values that represent unique contributions of the subtests. It is also true in this case, however, that only if we define the standard deviations of the subtests as the contributions will the weights have the effects that many measurement developers expect. I am not saying this is an incorrect definition of contribution; it's just a fact that researchers use a number of different definitions that are not in agreement, and measurement professionals should be aware of this, select a definition, and be prepared to discuss what it means for validity, reliability, and interpretation of the reported scores.

These facts have ramifications for practice in assessments at all levels. As one example, the frameworks for the NAEP use weighted combinations of scale scores to define the composite scales, which serve as the primary *scale score* reporting variable.[6] For the 2007 mathematics assessment (National Assessment Governing Board [NAGB], 2006), the relative importance of the scales, which are used as weights in forming the composite from those scales, are described in the following quotation:

> The distribution of items among the various mathematical content areas is a critical feature of the assessment design, as it reflects the ***relative importance*** [boldface italics added] and value given to each of the curricular content areas within mathematics. As has been the case with past NAEP assessments in mathematics, the categories have received

differential emphasis at each grade, and the differentiation continues in the framework for this assessment. (NAGB, 2006, p. 9)

I have emphasized the use of relative importance in this statement. This is the rationale behind use of the percentages (NAGB, Table 1) to determine weights used in defining the composite scale. The use of the values as weights implies that NAGB believes that the relative importance of the scales will be reflected in the resulting composite scores. As we have shown in this presentation, such use of weights does not necessarily have the result that this policy agency for the assessment might believe they have.

Effects on Reliability and Validity

Many writers (e.g., Carlson, 1968; Lane & Stone, 2006) have shown that differential weighting changes the *reliability* of the composite measure from that of an unweighted composite. It has been shown that the reliability is decreased by weighting less reliable subtests more heavily. Because of subjectivity in scoring and interactions between scorers, scoring rubrics, and students' item responses, many more sources of measurement error are found in CR item data than in SR item data, so weighting the CR items more heavily can have a detrimental effect on the reliability of the composite score.

The usual argument in favor of weighting the CR items more heavily is that it increases the *validity* of the composite scores. This argument is based on an assumption that certain skills and abilities cannot be measured by SR items (Lane & Stone, 2006; Schmeiser & Welch, 2006), an assumption that has not always been supported by research (Bridgeman & Rock, 1993; Lukhele, Thissen, & Wainer, 1994). It is also based on the fact that CR items take longer for students to respond to than SR items so, to minimize testing time burden, assessment instruments typically include disproportionately fewer CR items, which has an effect on the comparability of scores (Haertel & Linn, 1996). In subject areas of English language arts, however, the argument is usually along the lines that writing is an essential part of English skill development, so direct measurement of writing must be more valid than indirect SR measures (W. Yen, personal communication, August 21, 2013).

Concluding Remarks

In my many years of experience in the measurement field, I have often participated in discussions about whether, and how, to combine item scores into a total score. General algebraic explanations of procedures and consequences of the use of combinations are available (Wang & Stanley, 1970), especially with item response theory scoring (Sykes & Hou, 2003; Yen & Fitzpatrick, 2006), but the algebraic explanations are not readily understood by many practitioners. The geometric explanations provided in this presentation should help psychometricians better understand the issue of weighting scores from subsets of items and therefore help them to explain these issues to assessment development and program administration professionals and their clients. For many individuals, the geometric explanations are easier to grasp than the algebraic explanations usually used in explaining the issues.

Notes

1 Carlson (1968) derived the relationships among U_i, V_i, and T_i.
2 As shown by Carlson (1968), Guilford incorrectly specified the weights as the squares of what they should be.
3 To be more precise, orthogonality is equivalent to linear independence and *independence* is a broader term, including nonlinear cases. Independence implies zero correlation, but zero correlation only implies independence in the bivariate normal case in which the relationship is linear.
4 Actually there is a common factor of $\sqrt{N-1}$, as discussed in Appendix A; however, it can be ignored for purposes of this presentation.
5 The theorem is discussed more fully in Carlson (2014), and its proof is presented by G. D. Allen (2000) and Kazarinoff (1961).
6 NAGB policy states that achievement levels, which are based on the composite scale, are the primary vehicle for reporting NAEP results.

References

Allen, G. D. (2000). *Pappus of Alexandria (fl. c. 300 – c. 350)*. Retrieved from http://www.math.tamu.edu/dallen/masters/Greek/pappus.pdf

Allen, N. L., Donoghue, J. R., & Schoeps, T. L. (2001). *The NAEP 1998 technical report*. Washington, DC: National Center for Education Statistics.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York, NY: McGraw-Hill.

Bridgeman, B., & Rock, D. A. (1993). Relationships among multiple-choice and open-ended analytical questions. *Journal of Educational Measurement*, *30*, 313 – 329.

Carlson, J. E. (1968). *Effects of differential weighting on the reliability of essay grades* (Unpublished doctoral dissertation). University of Alberta, Edmonton, Canada.

Carlson, J. E. (2014). *A generalization of Pythagoras's theorem and application to explanations of variance contributions in linear models* (Research Report No. RR-14-18). Princeton, NJ: Educational Testing Service.

Chase, C. I. (1960). Computation of variance accounted for in multiple correlation. *The Journal of Experimental Education*, *28*, 265 – 266.

Creager, J. A., & Valentine, L. D. (1962). Regression analysis of linear composite variance. *Psychometrika*, *27*, 31 – 37.

Draper, N. R., & Smith, H. (1966). *Applied regression analysis*. New York, NY: Wiley.

Guilford, J. P. (1965). *Fundamental statistics in psychology and education*. New York, NY: McGraw-Hill.

Guttman, L. (1941). Mathematical and tabulation techniques. In P. Horst (Ed.), *The prediction of personal adjustment* (pp. 251 – 364). New York, NY: Social Science Research Council.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65 – 110). Westport, CT: American Council on Education & Praeger.

Haertel, E. H., & Linn, R. L. (1996). Comparability. In G. W. Phillips (Ed.), *Technical issues in large-scale performance assessment*. U.S. Department of Education: Washington, DC.

Horst, P. (1941). Mathematical contributions: Supplementary study E. In P. Horst (Ed.), *The prediction of personal adjustment*. New York, NY: Social Science Research Council.

Kazarinoff, N. D. (1961). *Geometric inequalities*. New York, NY: Random House.

Kolen, M. J. (2006). Scaling and norming. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 155 – 186). Westport, CT: American Council on Education & Praeger.

Lane, S., & Stone, C. A. (2006). Performance assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 387 – 431). Westport, CT: American Council on Education & Praeger.

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

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement*, *31*, 234 – 250.

National Assessment Governing Board. (2006). *Mathematics framework for the 2007 National Assessment of Educational Progress*. Retrieved from http://nagb.org/frameworks/math_07.doc

Richardson, M. (1941). The combination of measures. In P. Horst (Ed.), *The prediction of personal adjustment* (pp. 377 – 401). New York, NY: Social Science Research Council.

Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 307 – 353). Westport, CT: American Council on Education & Praeger.

Sykes, R. C., & Hou, L. (2003). Weighting constructed-response items in IRT-based exams. *Applied Measurement in Education*, *16*, 257 – 275.

Wang, M. W., & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, *40*, 663 – 704.

Wickens, T. D. (1995). *The geometry of multivariate statistics*. Hillsdale, NJ: Erlbaum.

Wilks, S. S. (1938). Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, *3*, 23 – 40.

Wonnacott, R. J., & Wonnacott, T. H. (1973). *Econometrics*. New York, NY: Wiley.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, *5*, 161 – 215.

Yen, W. M. (1983). Use of the three-parameter logistic model in the development of a standardized achievement test. In R. K. Hambleton (Ed.), *Applications of item response theory* (pp. 123 – 141). Vancouver, Canada: Educational Research Institute of British Columbia.

Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 111 – 153). Westport, CT: American Council on Education & Praeger.

Appendix A

Some Details of the Vector Algebra and Geometry

Basic Vectors and Operations

Two different geometries may be used to represent statistical and other data collected on a number of variables, p, and a sample of a number of persons, N. The geometry that readers are probably most familiar with is the *geometry in the variable space*. In this geometry, the variables form the axes and values associated with persons are plotted as points in a Euclidean space. The relationships among variables can be studied by analyzing the way the points fall in the plot; for example linear or not, highly correlated or not. The other geometry is the *geometry in the person space* in which the persons are the axes and the variables are plotted as points in a Euclidean space. The variables are often represented by vectors (directed line segments, represented by arrows) from the origin (point where the axes cross at zero on each axis) to the point representing the variable. Geometry in the person space is the geometry used in this article.

Because this article deals with tests, each variable comprises scores on a test (or subtest) and the persons are the test takers. A small example will be used to illustrate. In this example there are two subtests, X_1 and X_2, and three test takers with scores of 4, 2, and 2 on Test 1, and scores of 3, 3, and 1 on Test 2. In vector algebra, these are represented as:

$$X_1 = \begin{pmatrix} 4 \\ 2 \\ 2 \end{pmatrix} \quad X_2 = \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix}.$$

With three test takers there are three dimensions. In Figure A1, Test Taker 1 is represented with the horizontal axis, Test Taker 2 with the vertical axis, and Test Taker 3 with a depth axis. The axes intersect at the origin where all three scores are zero. As shown in the figure, the vector representation of subtest X_1 is an arrow from the origin to the point four units along the Test Taker 1 axis, two units up on the Test Taker 2 axis, and two units forward on the Test Taker 3 axis. Subtest X_2, similarly, is an arrow from the origin to the point three units along the Test Taker 1 axis, three units up on the Test Taker 2 axis, and one unit forward on the Test Taker 3 axis.

When the two component vectors (subtests in our case) are summed to form a composite, we use vector addition, an element-by-element operation as follows (designating the composite as X_c):

$$X_1 + X_2 = \begin{pmatrix} 4 \\ 2 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix}$$

$$= X_c = \begin{pmatrix} 4+3=7 \\ 2+3=5 \\ 2+1=3 \end{pmatrix}.$$

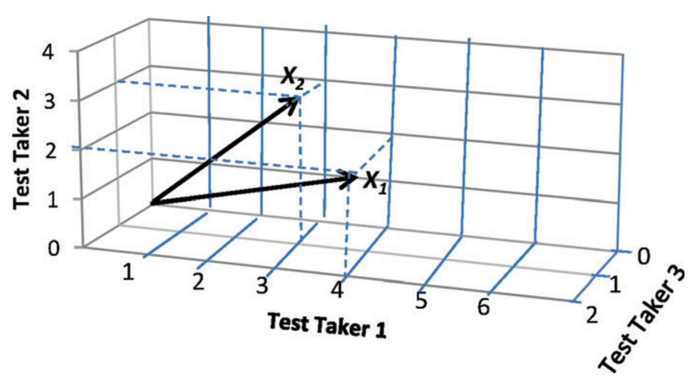

Figure A1 Vector representation of two subtests with three test takers.

For our purposes we transform the subtest data to *deviation metric* by subtracting the means of X_1 and X_2 (2.67 and 2.33, respectively) from each element of the respective vector, forming the deviation score vectors:

$$\boldsymbol{x}_1 = \begin{pmatrix} 1.33 \\ -.67 \\ -.67 \end{pmatrix} \quad \boldsymbol{x}_2 = \begin{pmatrix} .67 \\ .67 \\ 1.33 \end{pmatrix}.$$

Similarly to what we did with the raw score vectors, we form a *composite deviation vector* as the sum of the subtest deviation vectors:

$$\boldsymbol{x}_c = \boldsymbol{x}_1 + \boldsymbol{x}_2 = \begin{pmatrix} 1.33 \\ -.67 \\ -.67 \end{pmatrix} + \begin{pmatrix} .67 \\ .67 \\ 1.33 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix}.$$

The squared length of a vector can be seen through repeated application of Pythagoras's theorem to be equal to the sum of the squared elements of the vector (see, e.g., Wickens, 1995, p. 19). Consider the distance along the "floor" of Figure 1 to the point above which the end of vector X_1 lies, for example. By Pythagoras's theorem this distance is $W = \sqrt{4^2 + 2^2}$. The end of this vector is two units above the floor, so with $||X_1||$ representing the length of the vector and applying Pythagoras's theorem again the length of the vector is:

$$||X_1|| = \sqrt{W^2 + 2^2} = \sqrt{4^2 + 2^2 + 2^2} = \sqrt{\sum_{i=1}^{2} X_i^2}.$$

Similarly, the squared length of the deviation vector, $\boldsymbol{x}_1$ is:

$$||\boldsymbol{x}_1||^2 = \sum_{i=1}^{N} x_{1i}^2 = \boldsymbol{x}_1'\boldsymbol{x}_1, \tag{A1}$$

where x_{1i} is the ith element of the vector, and $\boldsymbol{x}_1'\boldsymbol{x}_1$ represents the *inner* or *scalar* product of the vector with itself, equal to the sum of squared elements (Wickens, 1995, pp. 10, 12–13).

Note that because we are dealing with vectors of the subscores in deviation metric, Equation A1 is the numerator of the unbiased estimate, s_1^2, of the variance of X_1,

$$s_1^2 = \frac{\sum_{i=1}^{N} x_{1i}^2}{N-1} = \frac{\boldsymbol{x}_1'\boldsymbol{x}_1}{N-1} = \frac{||\boldsymbol{x}_1||^2}{N-1}. \tag{A2}$$

So the length of the deviation score vector is the standard deviation, s_1, multiplied by the square root of one less than the sample size. As Wickens (1995) stated,

> In most analyses, the constant of proportionality $\sqrt{N-1}$ is unimportant, since every vector is based on the same number of observations. One can treat the length of a vector as equal to the standard deviation of its variable. (p. 19)

Hence, I follow his example and, in the body of this article, treat deviation score vectors as having lengths equal to the standard deviations of the variables represented by the vectors.

Returning to vector addition, it is represented geometrically by placing a copy of the second component vector at the end of the first component vector, and the composite is a vector from the origin to the end of the second component vector. In other words, the second vector is added on to the end of the first (keeping the lengths and directions of the two as they were), and the composite vector starts at the beginning of the first component and ends at the end of the second. This is illustrated in Figure A2. By carefully examining Figure A1, the reader should be able to see that Subtests 1 and 2 can be enclosed in a plane. Imagine the plane as a pane of glass oriented such that both vectors are embedded in it. Figure A2 shows such a plane and the addition of vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to form $\boldsymbol{x}_c$.

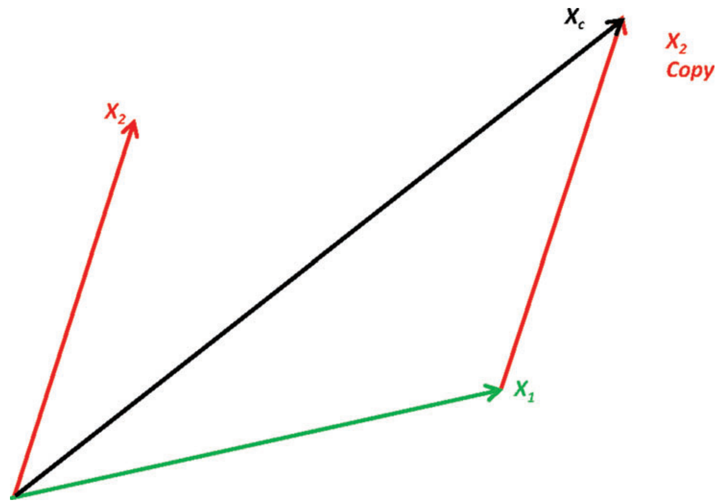

Figure A2 Vector addition: $\boldsymbol{x}_c = \boldsymbol{x}_1 + \boldsymbol{x}_2$.

Relationships to Statistics and Psychometrics

For the sake of simplification we consider the two vectors in Figure A2 as vectors of deviation scores. In this metric, then, as mentioned above, the lengths of the vectors can be considered to be equal to the standard deviations of the three variables represented in the figure. Thus we see for the vectors in Figure A2 (these are not the same vectors as used above—they are just two arbitrarily chosen vectors used to illustrate), that component vector $\boldsymbol{x}_1$ has slightly more variability than component vector $\boldsymbol{x}_2$, and the composite has the largest variability. It is well known that the variance of this sum is:

$$s_c^2 = s_1^2 + s_2^2 + 2r_{12}s_1s_2, \tag{A3}$$

where the last term is twice the sample *covariance* of X_1 and X_2.

Also, the angle between two deviation score vectors is related to the correlation between the two tests. The sample covariance between two variables, X and Y can be expressed as:

$$s_{xy} = \frac{\sum_{i=1}^{N} x_i y_i}{N-1} = \frac{\boldsymbol{x}'\boldsymbol{y}}{N-1}, \tag{A4}$$

and the sample *correlation coefficient* as:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}. \tag{A5}$$

As shown by Wickens (1995, p. 19) the sample *correlation coefficient is equal to the cosine of the angle between the two deviation score vectors*,

$$r_{xy} = \cos\left\{\angle\left(\boldsymbol{x}, \boldsymbol{y}\right)\right\}, \tag{A6}$$

where $\angle(\boldsymbol{x}, \boldsymbol{y})$ designates the angle between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. For the example used in Figure A2, the angle between $\mathbf{x}_1$ and $\boldsymbol{x}_2$ is 60°, hence the correlation between X_1 and X_2 is .5, which is the cosine of an angle of 60°.

As stated above, two vectors and their vector sum can always be represented as lying within a plane as in Figure A2. In Figure A1, with $N = 3$, this is obvious, but if the three vectors in the figure represent data from a sample of size N, the vectors lie within an N-dimensional space. You can visualize them as lying within a two-dimensional subspace (plane) within that N-dimensional space. Hence, we can extend this discussion without reference to the N-dimensional space when we have N test takers. The point is that the geometry is exactly the same for two dimensions or N dimensions.

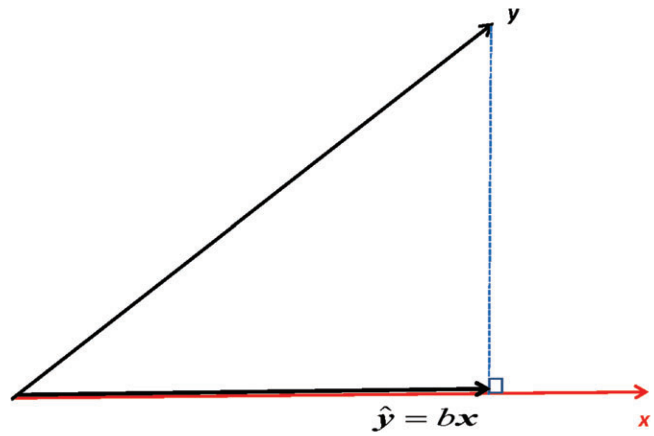

Figure A3 Linear regression: projection of $\boldsymbol{y}$ onto $\boldsymbol{x}$ with resultant vector the vector of predicted values.

Linear Models and Projections

For a linear regression model having a dependent variate, Y, and an independent variable, X, the model for N sample observations is usually expressed as:

$$Y_i = \alpha + \beta X_i + e_i \quad (i = 1, 2, \ldots, N),$$

where α and β are the intercept and slope parameters, respectively, to be estimated from the data. When expressed in deviation score metric, the intercept is zero, resulting in:

$$y_i = \beta x_i + e_i \quad (i = 1, 2, \ldots, N). \tag{A7}$$

For a sample of size N, the ordinary least squares (OLS) estimate of the slope parameter is (Wickens, 1995, pp. 32–35)

$$b = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2} = \frac{\boldsymbol{x'y}}{\boldsymbol{x'x}} = \frac{s_{xy}}{s_x^2}. \tag{A8}$$

The OLS estimate of the vector of predicted values of $\boldsymbol{y}$ is computed as

$$\hat{\boldsymbol{y}} = b\boldsymbol{x}, \tag{A9}$$

and the squared length of this vector is:

$$\begin{aligned} ||\hat{\boldsymbol{y}}||^2 &= ||b\boldsymbol{x}||^2 = (b\boldsymbol{x})'(b\boldsymbol{x}) \\ &= b^2\boldsymbol{x'x} = \boldsymbol{b}^2||\boldsymbol{x}||^2. \end{aligned} \tag{A10}$$

Following Wickens (1995) suggestion on ignoring the factor of $N - 1$, this is equal to

$$||\hat{\boldsymbol{y}}||^2 = b^2 s_x^2. \tag{A11}$$

As shown in Figure A3, geometrically the regression of Y onto X is the perpendicular projection of $\boldsymbol{y}$ onto $\boldsymbol{x}$, with the vector resultant of the projection being the vector of predicted values on Y.

As mentioned previously, and should be clear in Figure A3, the three vectors $\boldsymbol{y}$, $\boldsymbol{x}$, and $\hat{\boldsymbol{y}}$ are coplanar (lie within a plane). Also note that $\mathbf{x}$, and $\hat{\boldsymbol{y}}$ are collinear (lie within the same line).

In the case of two predictors in multiple linear regression, the two vectors of predictors are coplanar, and regression involves perpendicular projection of the $\mathbf{y}$ vector onto that plane, resulting in the vector of predicted values,

$$\hat{\boldsymbol{y}} = b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2. \tag{A12}$$

This generalizes to the case of p predictors to:

$$\hat{\boldsymbol{y}} = \sum_{j=1}^{p} b_j \boldsymbol{x}_j. \tag{A13}$$

Geometrically, this is the projection of a vector onto a hyperplane in p dimensions.

Measures of the strength of the prediction in the sample are the multiple correlation coefficient, R and its square, often referred to as the coefficient of determination. As shown by Wickens (1995, p. 19), the correlation between two variables is equal to the cosine of the angle between their deviation score vectors,

$$\begin{aligned} r_{xy} &= \cos \angle \left(\boldsymbol{x}, \boldsymbol{y}\right) \\ &= \frac{\boldsymbol{x}'\boldsymbol{y}}{||\boldsymbol{x}|| \times ||\boldsymbol{y}||} = \frac{s_{xy}}{s_x s_y}, \end{aligned} \tag{A14}$$

the ratio of the covariance to the product of the two standard deviations. Wickens (p. 36) also shows that the multiple correlation coefficient is equal to the cosine of the angle between $\mathbf{y}$ and $\hat{\boldsymbol{y}}$, and its square is also equal to the ratio of the squared length of $\hat{\boldsymbol{y}}$ to the squared length of $\mathbf{y}$ (p. 49),

$$R^2 = \frac{||\hat{\boldsymbol{y}}||^2}{||\boldsymbol{y}||^2}, \tag{A15}$$

which is equal to the proportion of variance in Y predictable from the p predictors.

Some of the issues discussed in the text use a standardized regression model and the standardized coefficients; that is, the variables are all transformed to a mean of zero and standard deviation of one before analyzing with the multiple linear regression model. In this case, I use b_i^* to represent the ith standardized regression coefficient, and Equation A12 is written as:

$$\hat{\boldsymbol{y}} = \sum_{j=1}^{p} b_j^* \boldsymbol{x}_j. \tag{A16}$$

Similarly to Equation A1, when using the standardized metric the inner product of two different vectors is equal to the correlation between the two variables represented by the vectors. This may be seen from Wickens's (1995, p. 19) Equation 2.18 because in the standardized metric the lengths of the two vectors in that expression are each equal to 1.0. Using a similar argument, it may be seem from Wickens's Equation 4.7 (p. 49) that the expression for R^2 may be written as:

$$R^2 = \sum_{j=1}^{p} b_j^* r_{yj}, \tag{A17}$$

in which r_{yj} is the correlation between Y and X_j. This equation is used in some of the variance contribution expressions in the text. Another expression for R^2 that is used in the text is (see, e.g., Carlson, 2014, appendix)

$$R^2 = \sum_{j=1}^{p} \left(b_j^*\right)^2 + \sum_{j=1}^{p} \sum_{j' \neq j=1}^{p} 2 b_j^* b_{j'}^* r_{jj'}. \tag{A18}$$

Appendix B

Artificial Data Example

The artificially constructed data and summary statistics reported for the variables reported in Table 1 of the text are displayed in Table B1.

Table B2 displays the computations of the statistics used in the various definitions of contributions. The left side shows the subtest intercorrelation matrix, $\boldsymbol{R}_x$, and its inverse. The column vector $\boldsymbol{r}_{jc}$ is the vector of correlations between the two subtests and the composite. The standardized regression coefficients, in vector b^*, are computed as the product of

Table B1 Data and Summary Statistics for the Example

Test taker	X_1	X_2	X_c
1	1	0	1
2	5	0	5
3	7	1	8
4	6	1	7
5	5	2	7
6	3	2	5
7	12	3	15
8	10	3	13
9	11	4	15
10	7	4	11
11	6	5	11
12	11	5	16
13	17	6	23
14	10	6	16
15	16	7	23
16	10	7	17
17	11	8	19
18	9	8	17
19	17	9	26
20	14	10	24
Sum	188	91	279
Mean	9.4000	4.5500	13.9500
s^2	20.0421	9.2079	49.6289
s	4.4768	3.0345	7.0448

Table B2 Computed Contribution Statistics

$$R_x = \begin{pmatrix} 1.000 & 0.750 \\ 0.750 & 1.000 \end{pmatrix}$$

	R_x^{-1}	r_{jc}	b^*	$V_j = (b_j^*)^2$	$T_j = b_j^* r_{jc}$	$U_j = 1 - r_{j'c}$	W_j $s_j r_{jc}$	W_j^2 $s_j^2 r_{jc}^2$
X_1	$\begin{pmatrix} 2.286 & -1.715 \\ -1.715 & 2.286 \end{pmatrix}$	0.907	0.431	0.186	0.391	0.081	2.753	7.581
X_2		0.959	0.636	0.404	0.609	0.177	4.291	18.416

$$J_{12} = 2b_1^* b_2^* r_{12} = 0.411$$

Note. $r_{j'c}$ is the correlation of the composite with the other subtest (e.g., for U_1 Subtest 2).

matrix $\boldsymbol{R}_x^{-1}$ and vector $\boldsymbol{r}_{jc}$. The other four columns show the computations of the V, T, U, W, and W^2 statistics. Finally the computation of the joint contribution statistic, J_{12}, is shown at the bottom.

Suggested citation:

Carlson, J. E. (2014). *What differential weighting of subsets of items does and does not accomplish: Geometric explanation* (ETS Research Report No. RR-14-20). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12020

Action Editor: Rebecca Zwick

Reviewers: Daniel McCaffrey and John Donoghue