In Top-Down Decisions, Weighting Variables Does Not Matter: A Consequence of Wilks' Theorem

MALCOLM JAMES REE
THOMAS R. CARRETTA
JAMES A. EARLES
Brooks Air Force Base, TX

It is often appropriate to weight variables to form a composite for making decisions. Examples include selection systems, organizational performance criteria, test items, and decision modeling. Frequently, criterion-based regression-weighting is employed, but a sizable literature argues for unit or simple weighting. Wainer demonstrated small loss from equal weights compared to regression weights. Usually, weights are of little importance for rank ordering, echoing Wainer's "it don't make no nevermind." Wilks proved a general theorem, that under common circumstances, almost all weighted composites of a set of variables are highly correlated. That is, if a single set of variables is weighted two different ways to form two composites, the expected correlation for the two composites is very high. The authors demonstrate the effect of Wilks' theorem through illustrative examples. Implications of Wilks' theorem are discussed. When top-down decisions are made, weighting variables does not matter because the rank ordering remains almost constant.

Frequently, it is desirable to weight a set of variables for making decisions. For example, suppose an organizational climate survey measures the six content areas of control, goals, decisions, communication, motivation, and leadership. Managers want to weight the six content areas to provide an overall climate score. A set of weights might be found through consensus, a Delphi technique, or some other method. Does the method of determining the weights make a difference? A second example might be that a new job has been created or a job has been redesigned and it is desired to keep using the same multiple aptitude test battery for selecting employees. Because the new or redesigned job has unique tasks, it is usually considered necessary to do a job analysis to know how to weight the tests to form a selection composite. We

Authors' Note: The opinions expressed are those of the authors and do not necessarily reflect those of the United States Air Force, the Department of Defense, or the United States government. Address all correspondence and requests for reprints to the first author at AFRL/HEJD, 7909 Lindbergh Drive, Brooks AFB, TX 78235-5352; e-mail: ree@alhrm.brooks.af.mil.

Organizational Research Methods, Vol. 1 No. 4, October 1998 407-420

demonstrate that under many common conditions, weighting has little effect in decision making. This is shown to be a consequence of Wilks' theorem (1938).

A weighted linear composite is computed by multiplying weights times variables and summing the products. For example, a linear composite of variables A, B, and C might be expressed as Composite 1 = $w_1A + w_2B + w_3C$, where w_1, w_2, and w_3 are weights. Similarly, the same variables A, B, and C might be weighted by w_4, w_5, and w_6 to produce Composite 2. The weights can be determined empirically through regression or factor analysis or established by nonempirical means such as policy capturing, unit weighting, equal weighting, or other model-building methods (see, for example, Fralicx & Raju, 1982).

There is a literature that argues for simple or unit weights without consideration of a criterion (Aiken, 1966; Dawes, 1979; Schmidt, 1971; Stahlnacker, 1938; Wainer, 1976, 1978). Sometimes the use of simple or unit weights is called for to simplify computation (Stahlnacker, 1938). Weighting every variable by 1 is the same as summing the scores of the variables. Simple or unit weights are sometimes advised to add robustness (Dawes, 1979). Simple and unit weights are not influenced by outliers in the data and cannot lead to shrinkage on cross application. Sometimes simple or unit weights are suggested because the benefit of perceived higher accuracy of regression weights is illusory (Wainer, 1976, 1978). For example, Wainer (1976) demonstrated through mathematical proof that under reasonably broad conditions, the loss of predictive efficiency from using equal weights is very small.

The literature provides many empirical comparisons of weighting methods. Lawshe and Schucker (1959) examined the behavior of variable weights under several weighting methods and concluded that differential weighting was no better than the addition of raw scores for samples investigated. In a simulation-based investigation using small samples, Schmidt (1971) found unit weights preferable to regression weights. Dawes and Corrigan (1974) provided examples from fields relevant to organizational study in which they compared several methods for deriving weights. They concluded that "deviations from optimal weighting do not make much practical difference" (p. 105). Aamodt and Kimbrough (1985) found high intercorrelations of composites weighted by ranks, critical incidents, unit weights, or regression and concluded that the method of determining the weights was unimportant. Streiner and Miller (1989) revisited the issue using unit and differential weighting for a personality inventory. They found correlations between differentially and unit-weighted composites near 1. These authors and others found that weighted composites correlated highly regardless of how the weights were created, especially when sample size was small.

A Theorem by Wainer

To explain the indifference of weighting, Wainer (1976, 1978) and Laughlin (1978) offered a proof concerning equal weighting of standardized predictors and predictive efficiency. Their proof provides a formula for the average loss in predictive efficiency when equal weights are substituted for differential weights from a uniform distribution. In their formula, the loss of predictive efficiency increases directly with the number of predictors. This applies when the variables are independent. When the variables are correlated, the loss is smaller still.

Wainer (1976) concluded, "It is a rare situation that calls for regression weights that are unequal" (p. 216). Furthermore, he noted that relative prediction (ranking), not absolute prediction (point estimation), is the most typical kind of problem. Examples of ranking include prioritizing for selection of managers, selection of machinery or products, organizational commitment, and oil drilling site selection, to name a few.

Wainer's (1976) formula is most useful for a small number of predictors. Wilks (1938) has provided a formula that has the number of predictors in the denominator as opposed to the numerator, as in Wainer, and is, therefore, useful for larger numbers of predictors. We will show that in many common ranking situations, Wilks' theorem means that a criterion is usually irrelevant in establishing the numerical values of the weights, a logical conclusion but unstated by Wilks.

A Theorem by Wilks

Wilks (1938) provided a theorem that explained the mathematical inevitability of the ubiquitous finding that unit weighting produces a composite that is very highly correlated with composites weighted by any other method, especially when the number of variables is large. Three factors are important for determining the expected correlation between composites. These are the average correlation among the variables, the number of variables, and the relative variability of the weights. Wilks presented the following approximation formula (terms of order $1/K^2 \ldots 1/K^n$ are dropped) for the expected correlation, R, of two weighted linear composites:

$$\bar{R} = 1 - \frac{1}{2\bar{r}K}\left[\frac{\sigma_v^2}{\mu_v^2} + \frac{\sigma_w^2}{\mu_w^2}\right], \tag{1}$$

where $\bar{r}$ is the average correlation of the variables, K is the number of variables, μ_v and μ_w and σ_V and σ_W are the means and standard deviations of the populations from which weight set V and weight set W are randomly drawn. These can be any two populations. The two fractions in the brackets are the squared coefficients of variation (CVs) of the populations from which the weights are drawn. See also Gulliksen, chapter 20 (1950/1987). As will be discussed later, there is a compensatory nature to the terms in the equation.

The purpose of the current effort is to explain the inescapable consequences of Wilks' theorem for organizational research and practice. In particular, we demonstrate that in highly generalizable circumstances, almost any set of positive weights applied to a given set of variables will lead to a near identity of rankings. Conditions under which this near identity of rankings occurs are described, and a demonstration is provided. We take a data set with a reasonably large number of variables and use several weighting schemes to demonstrate the principle embodied in Wilks' theorem.

Study 1

Method

Participants. Participants were approximately 89,000 enlisted men and women enrolled in Air Force technical training courses. They were selected, in part, on the basis of enlistment qualification test scores. Each participant performed a job in one

Table 1
Correlations of ASVAB Tests in the Normative Sample

Tests	*GS*	*AR*	*WK*	*PC*	*NO*	*CS*	*AS*	*MK*	*MC*	*EI*
GS	1.000									
AR	.722	1.000								
WK	.801	.708	1.000							
PC	.689	.672	.803	1.000						
NO	.524	.627	.617	.608	1.000					
CS	.452	.515	.550	.561	.701	1.000				
AS	.637	.533	.529	.423	.306	.225	1.000			
MK	.695	.827	.670	.637	.617	.520	.415	1.000		
MC	.695	.684	.593	.521	.408	.336	.741	.600	1.000	
EI	.760	.658	.684	.573	.421	.342	.745	.585	.743	1.000

Note. ASVAB = Armed Services Vocational Aptitude Battery. The tests are General Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Numerical Operations (NO), Coding Speed (CS), Auto and Shop Information (AS), Mathematics Knowledge (MK), Mechanical Comprehension (MC), and Electronics Information (EI).

of seven job families. The sample sizes for the seven job families were Clerical, n = 15,338; General I, n = 12,925; General II, n = 17,119; Electrical, n = 10,115; Mechanical, n = 5,086; Aircraft Maintenance, n = 9,718; and Miscellaneous Electrical/ Mechanical, n = 18,034. A second group of participants was the 9,173 members of the enlistment qualification test normative sample, which represents 18- to 23-year-old American youth as of 1980.

Measures. The Armed Services Vocational Aptitude Battery (ASVAB; Ree & Carretta, 1994) is a state-of-the-art (Jensen, 1985; Murphy, 1985) test battery (Earles & Ree, 1992) used for enlistment qualification. Its 10 tests are General Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Numerical Operations (NO), Coding Speed (CS), Auto and Shop Information (AS), Mathematics Knowledge (MK), Mechanical Comprehension (MC), and Electronics Information (EI). Table 1 shows the correlations among the tests in the normative sample (Ree & Carretta, 1994; Ree & Wegner, 1990). The average correlation among these tests is about .60.

Criteria. The criteria were final grades earned during technical training in the seven job families. Each participant had a criterion score in only one of the seven job families.

Procedures. Criteria were regressed on the 10 ASVAB tests. Regression weights were generated independently within each of the seven job families. This resulted in seven differing sets of regression weights as shown in Table 2.

The seven sets of weights from these regressions were applied to the scores of the participants in the normative sample (n = 9,173). Predicted final school grades were computed for each participant within each job family using the weights derived for that job family. Each participant therefore had seven predicted grades.

The rank ordering of applicants is frequently more important than their actual predicted score, because most organizational decision systems operate on the principle

Table 2
Regression-Based Weights From the Job Family Data

Job Family	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI
	Weights									
1. Clerical	.23	.83	.72	.56	.00	.47	.43	1.00	.16	.11
2. General I	.05	.51	.72	.65	.00	.51	1.00	.60	.49	.17
3. General II	.36	.97	.00	.42	.01	.29	.39	1.00	.19	.89
4. Electrical	.52	.59	.45	.45	.00	.34	.37	1.00	.16	.48
5. Mechanical	.74	.51	1.00	.82	.01	.42	.71	.83	.11	.29
6. Aircraft maintenance	.00	.61	.00	.28	.11	.11	1.00	.48	.37	.73
7. Miscellaneous electrical/mechanical	.52	.59	.50	.28	.00	.45	1.00	.50	.51	.54

Note. Weights were reported to two places. The regression intercepts were omitted. (Job Family Predicted Criterion Vector) = (Regression Coefficient Matrix) × (Score Vector).

of top-down selection. As the name suggests, in top-down selection, once applicants have been assigned scores on the composite, they are selected from the top to the bottom of an ordered list until available positions have been filled. Although this example is from personnel selection, it should be noted that top-down selection frequently is used to decide among organizational alternatives. For example, vendor proposals or investment opportunities can be selected by considering many variables and ranking the alternatives.

To evaluate the effect of weights on ordering of applicants, regression-weighted scores were correlated using the Pearson product-moment coefficient. A perfect correlation of 1 would indicate identical rankings, and a zero correlation would indicate no relationship between ranks. A negative correlation would indicate an inversion in rank order.

Results and Discussion

Table 3 shows the correlations among the seven composite scores for the regression-weighted equations. The correlations between pairs of composite scores ranged from .93 to .99, with a mean of .973. The rank orderings of the participants were nearly identical for the seven sets of weights. Clearly, the seven different sets of regression weights had little effect on the ordering of the participants. If top-down selection were used, nearly identical sets of applicants would be selected for each of the seven jobs. The diverse criteria made up of numerous dissimilar skills and knowledge would not cause different applicants to be selected. In fact, any one of the seven weighted equations would be expected to produce the same selections in any one of the seven broad job families. The criteria did not matter.

Study 2

Viewing the results of Study 1, some might argue that the criteria could not matter because the criteria, training grades, were all the same sort of measure, leading to similar regression weights. The second study avoids this criticism by not using criteria.

Table 3
Correlations of the Weighted Scores Using Regression-Based Weights

Job Family	*1*	*2*	*3*	*4*	*5*	*6*	*7*
1.	1.00						
2.	.98	1.00					
3.	.98	.97	1.00				
4.	.99	.98	.99	1.00			
5.	.99	.98	.97	.99	1.00		
6.	.93	.97	.97	.95	.94	1.00	
7.	.96	.99	.98	.98	.98	.98	1.00

Note. Correlations were truncated at two decimal places.

Method

Participants. The sample was composed of 9,173 male and female respondents, ages 18 to 23, who composed the normative sample for the ASVAB (Ree & Carretta, 1994; Ree & Wegner, 1990).

Measures. As in Study 1, the 10 ASVAB tests were used.

Criteria. No criteria were used to derive weights.

Procedures. The argument could be made that criteria in Study 1 were too similar to have a substantially differentiating effect on the weights, thereby increasing the correlation among composite scores. To avoid this problem, weights were randomly generated.

Several sets of randomly generated weights were associated with the tests. The weights were all single digits that ranged from 1 to 9. The first set of weights was a telephone number, the second a postal zip code repeated, the third a scrambled social security account number with one digit repeated, the fourth a personal computer serial number, and the fifth a set of random integers. The next five sets of weights were the first five sets reversed. The 11th set was all 1s to produce unit weighting. The 11 sets of randomly generated weights (shown in Table 4) were then applied to the normative sample to produce 11 composite scores (i.e., weighted linear composites, $w_1GS + w_2AR + w_3WK + \ldots w_{10}EI$) per person. As in Study 1, Pearson product-moment correlations were computed for the 11 composite scores.

Results and Discussion

Table 5 shows the correlations among the randomly weighted scores. The results from Study 2 were a strong confirmation of the results from Study 1. The correlations between pairs of weighted scores ranged from .97 to 1.00, with 41 of the 55 values at .99 or higher. Consistent with Wainer (1976), unit-weighted scores, with a CV of 0, had the highest average correlation with the other weighted scores (.993). As in Study 1, the effect of using these 11 composites would have been to make nearly the same decisions in a top-down selection system. In terms of rank ordering, these 11 composites behave as though they were identical. To be sure, the criteria did not and could not influence the weights: No criteria were used in producing any of the sets of 11 weights!

Table 4
Arbitrary Weights Applied to the ASVAB Tests

	ASVAB Tests									
Weight Set	*GS*	*AR*	*WK*	*PC*	*NO*	*CS*	*AS*	*MK*	*MC*	*EI*
1.	5	1	2	5	3	6	3	2	5	6
2.	7	8	2	3	5	7	8	2	3	5
3.	6	2	8	1	4	2	6	5	7	8
4.	9	6	8	1	6	3	6	7	3	6
5.	7	3	6	1	5	8	1	7	6	2
6.	6	5	2	3	6	3	5	2	1	5
7.	5	3	2	8	7	5	3	2	8	7
8.	8	7	5	6	2	4	1	8	2	6
9.	6	3	7	6	3	6	1	8	6	9
10.	2	6	7	1	8	5	1	6	3	7
11.	1	1	1	1	1	1	1	1	1	1

Note. (Weight Set Outcome Vector) = (Coefficient Matrix) × (Score Vector).

Table 5
Correlations of the Weighted Scores Using Arbitrary Weights

Weight Set	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
1.	1.00										
2.	.99	1.00									
3.	.99	.99	1.00								
4.	.99	.99	.99	1.00							
5.	.98	.98	.97	.99	1.00						
6.	.99	1.00	.99	.99	.98	1.00					
7.	1.00	.99	.98	.99	.98	.99	1.00				
8.	.98	.98	.98	.99	.99	.99	.98	1.00			
9.	.99	.99	.99	.99	.99	.99	.99	1.00	1.00		
10.	.98	.98	.97	.99	.99	.99	.98	.99	.99	1.00	
11.	.99	1.00	.99	.99	.99	1.00	.99	1.00	.99	.99	1.00

Note. Correlations were truncated at two decimal places.

General Discussion

In a top-down decision system, the kind frequently used in organizations, high correlations among weighted composites would mean that almost exactly the same prioritized list (products, applicants, objects) would be created and the same decisions made *regardless of which weights were used.* This is true even though the actual predictiveness of these weighted composites would vary. As Wainer (1976) observed, predictiveness is not the issue, rank ordering is.

As demonstrated in Studies 1 and 2, the highest ranked individuals would have been selected regardless of which composite were used. That is, the same individuals (with the possible exception of some flip-flopping about the cut score) would have been selected regardless of which set of weights had been used. The high correlations of the weighted scores demonstrated that although the weights differed, those selected would not have differed. This is a restating of Wainer's (1976) "it don't make no nevermind."

Expected Correlations and Wilks' Theorem

Equation (1) gives an approximation of the expected correlation between two weighted composites. Using this equation, we have computed estimated correlations between weighted composites as might be found in organizational research. This was done for all combinations of the correlations from .1 to .9 in .1 increments by three numbers of variables (5, 10, or 15) and by two values for the coefficient of variation of the weights (CV = .200 or .462).

The levels of correlation were used to demonstrate the consequence of average correlation among the variables. The number of variables was manipulated to represent systems with few to many components, and the CVs (CV = standard deviation divided by the mean) were chosen to represent relatively low and high variability of weights. The CV of .200 equals a unit standard deviation divided by a mean of 5. The higher CV of .462 equals the standard deviation of the uniform distribution of 1 to 9 divided by a mean of 5. Experience shows that the lower CV is more likely in practice. The results are presented in Table 6 and in Figures 1 and 2.

The example with five variables could be a decision system that creates a composite based on measures of organizational climate. Consider the results for a five-variable composite in Table 6 in which CV = .200, a circumstance in which the relative variability of the weights is small. The estimated correlations between composites made up of five variables range from .920 to .991. Under these conditions, the rankings would be little affected by the weights. When the variability of the weights is greater, the composite correlations will be lower. Consider an example in which the variability of the five weights is dramatically greater (CV = .462). The correlations expected from Wilks' theorem would range from .573 to .953.

When weights are more variable, the decisions would be affected by the weights only when the correlations among the variables are low. When the correlations are moderate or high, there will be little effect due to the weights. Correlations of human and organizational characteristics are almost always moderate to high. Average correlational values of .1 through .4 are not likely to be encountered in practice except as artifacts due to unreliable measurement or estimation in highly range-restricted samples.

There is a compensatory nature to Equation (1) in the interaction of correlations among the variables and the number of variables. As the magnitude of the correlation among the variables drops, increasing the number of variables maintains the expected correlation between the composites.

The role of the squared CVs in Equation (1) is of particular interest. The CV describes the dispersion of the weights relative to the magnitude of the mean. The smaller the CV, the more likely that the weights do not include zero or negative values. Also, the smaller the CV, the larger the expected correlation between any pair of composites.

Implications

Wilks' theorem has implications for all top-down decision systems using weighted composites, even though we have cast the demonstrations in terms of selection systems. Weighted criterion variables are a good example. Fralicx and Raju (1982) noted no differences in weighted performance criteria using several sets of positive

Table 6
Estimated Expected Correlation of Linear Composites

	CV = 0.200			*CV = 0.462*		
	Number of Variables			*Number of Variables*		
Average r	*5*	*10*	*15*	*5*	*10*	*15*
.1	.920	.960	.973	.573	.787	.858
.2	.960	.980	.987	.787	.893	.929
.3	.973	.987	.991	.858	.929	.953
.4	.980	.990	.993	.893	.947	.964
.5	.984	.992	.995	.915	.957	.972
.6	.987	.993	.996	.929	.964	.976
.7	.989	.994	.996	.939	.970	.980
.8	.990	.995	.997	.947	.973	.982
.9	.991	.996	.997	.953	.976	.984

Note. CV is the coefficient of variation.

weights. Carretta (1992) confirmed Fralicx and Raju's results for training criteria. Pritchard and Roth (1991) found correlations averaging .98 for two methods of weighting measures of worker productivity.

Roberts and Glick (1981) noted flaws in job task design and found that job ordering was substantially unaffected by weighting procedures. They observed that any form of addition-based weighting would be acceptable.

Dawes and Corrigan (1974) reported the same lack of effects due to differential weighting on modeling or aiding decision making. Einhorn and Hogarth (1975) also reported a lack of differences in decision making for "optimum" and simple weights. Dawes (1979) found the same results when he presented a practical example in the ranking and selection of bullets by a big-city police department. Dyer, Lund, Larsen, Kumar, and Leone (1990) provided another example concerning decision making about oil and gas exploration. In all these studies, differing sets of weights led to the same conclusion. These findings were entirely predictable by Wilks' theorem.

In an example from social psychology, Skinner and Lei (1980) found differentially weighted composites to correlate almost perfectly with unit-weighted composites when studying social adjustment through life change variables. In the same research field, Rahe (1978) and Lorimer, Justice, McBee, and Weinman (1979) observed equivalent results and argued for the use of simple weights on the basis of high correlation of weighted composites.

Ree and Earles (1991) demonstrated the applicability of Wilks' theorem to factor analysis for the first principal component, first principal factor, and the highest hierarchical factor from correlated variables. From a single data set, they showed that first principal component scores, first principal factor scores, and highest hierarchical factor scores correlated nearly 1 with one another. This was confirmed by Guttman (1992) and by Jensen and Weng (1994).

Differential weighting of cognitive test questions has been studied for many years, with a heavy emphasis from the 1920s to the 1960s (Aiken, 1966; Burt, 1950; Douglass & Spencer, 1923; Guilford, Lovell, & Williams, 1942; Stahlnacker, 1938). Aiken (1966) believed that the issue of differential weighting of questions was settled, but he noted that it was difficult to convince his colleagues because "it is intuitively reasonable that

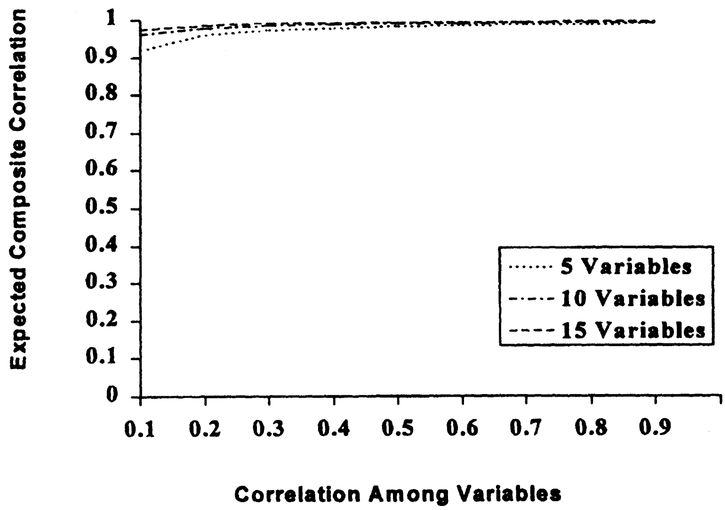

Figure 1: Illustration of the Estimated Expected Correlation of Linear Composites as a Function of the Number of Variables, Correlation Among the Variables, and CV = 0.200

weighting should make a difference" (p. 183). Wilks' theorem should have settled the issue in 1938, and Wainer (1976) should have reminded us.

Even though examples abound in the literature, a survey of nine frequently used human resources management and industrial/organizational psychology texts and nine frequently used psychometrics texts was conducted (see appendix) to determine how weighting of variables was treated. Although some mentioned the utility of unit weighting, none explained Wilks' theorem. A singular exception is the classic text from Gulliksen (1950) that was reissued in 1987.

We have cast the demonstration in terms of selection systems and have presented examples from decision making and human resources management. Nevertheless, the *mathematical* theorems of Wilks and Wainer have consequences for all fields in which top-down decisions are based on weighted composites.

What To Do?

Why has there been such a lack of confidence in the efficacy of unit or simple weighting? Perhaps it is because there are no studies systematically investigating top-down decision making. The numerous studies on predictive accuracy of weighting methods versus number of independent variables and participants provide a general methodology for conducting both empirical and simulation studies (see Raju, Bilgic,

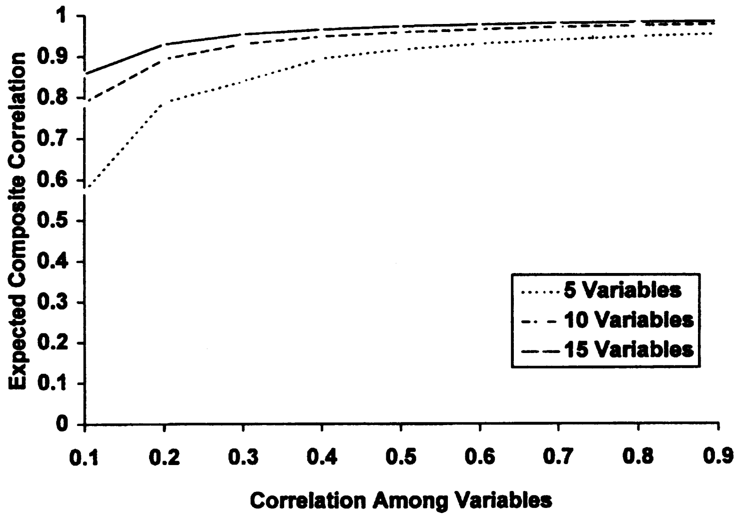

Figure 2: Illustration of the Estimated Expected Correlation of Linear Composites as a Function of the Number of Variables, Correlation Among the Variables, and CV = 0.462

Edwards, & Fleer, 1997). The object would be to ignore predictive accuracy and to look at two outcomes. The first outcome would be the similarity of people, things, or objects selected under various schemes for weighting the variables. Notice that this does not require a criterion. The second outcome would be the comparison of the average criterion scores of selected people, things, or objects, if a criterion were available.

While practitioners await further confirmation of the efficacy of unit and simple weighting, they should take advantage of the certainty of mathematical proof. That is, the practitioner should find independent variables that are positively correlated with the dependent variable and form weighted or unweighted composites by addition. For example, if a manager wants to weight certain variables in a Composite 1, some 2, and others 3, permit it. If management wants a group of outside experts to assign differential weights to variables in a composite, let them. If the circumstances demand regression weighting, this is also permissible provided there are no negative weights. In all these situations, the effect on top-down decision making will be benign. The practitioner has great freedom to form composites using many weighting schemes. We revise the suggestion by Dawes and Corrigan (1974) "The whole trick is to decide what variables to look at and then to know how to add" (p. 105) to be "the whole trick is to decide what variables to look at, freely apply positive weights, and know how to add."

APPENDIX
Human Resources Management, Industrial/Organizational Psychology, and Psychometrics Texts Surveyed

Human Resources Management and Industrial/Organizational Psychology Texts

Cascio, W. F. (1991). *Applied psychology in personnel management* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Gatewood, R. D., & Feild, H. S. (1990). *Human resources selection* (2nd ed.). Fort Worth, TX: Dryden.

Landy, F. L. (1989). *Psychology of work behavior* (4th ed.). Pacific Grove, CA: Brooks/Cole.

McCormick, E. J., & Ilgen, D. R. (1985). *Industrial and organizational psychology* (8th ed.). Englewood Cliffs, NJ: Prentice Hall.

Milkovich, G., & Glueck, W. (1985). *Personnel/human resources management: A diagnostic approach* (4th ed.). Plano, TX: Business Publications.

Minor, J. B. (1992). *Industrial-organizational psychology*. New York: McGraw-Hill.

Riggio, R. E. (1990). *Introduction to industrial/organizational psychology*. Glenview, IL: Scott, Foresman.

Saal, F. E., & Knight, P. A. (1988). *Industrial/organizational psychology: Science and practice*. Pacific Grove, CA: Brooks/Cole.

Siegel, L., & Lane, I. (1982). *Personnel and organizational psychology*. Homewood, IL: Irwin.

Psychometrics Texts

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Harcourt Brace Jovanovich.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco: W. H. Freeman.

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

Linn, R. L. (1989). *Educational measurement* (3rd ed.). New York: Macmillan.

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

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Suen, O. K. (1990). *Principles of test theory*. Hillsdale, NJ: Lawrence Erlbaum.

Thorndike, R. L. (1971). *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

References

Aamodt, M. G., & Kimbrough, W. W. (1985). Comparison of four methods for weighting multiple predictors. *Educational and Psychological Measurement*, *45*, 477-482.

Aiken, L. R., Jr. (1966). Another look at weighting test items. *Journal of Educational Measurement*, *3*, 183-185.

Burt, C. (1950). The influence of differential weighting. *British Journal of Psychology, Statistical Section*, *3*, 105-125.

Carretta, T. R. (1992). Understanding the relations between selection factors and pilot training performance: Does the criterion make a difference? *The International Journal of Aviation Psychology, 2*, 95-105.

Dawes, R. (1979). The robust beauty of improper linear models. *American Psychologist, 34*, 571-582.

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

Douglass, H. R., & Spencer, P. L. (1923). Is it necessary to weight exercises in standard tests? *Journal of Educational Psychology, 1*, 109-112.

Dyer, J. S., Lund, R. N., Larsen, J. B., Kumar, V., & Leone, R. P. (1990). A decision support system for prioritizing oil and gas exploration activities. *Operations Research, 38*, 386-396.

Earles, J. A., & Ree, M. J. (1992). The predictive validity of ASVAB for training grades. *Educational and Psychological Measurement, 52*, 721-725.

Einhorn, H., & Hogarth, R. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance, 13*, 171-192.

Fralicx, R. D., & Raju, N. (1982). A comparison of five methods for combining criteria into a single composite. *Educational and Psychological Measurement, 42*, 823-827.

Guilford, J. P., Lovell, C., & Williams, R. M. (1942). Completely weighted versus unweighted scoring in an achievement examination. *Educational and Psychological Measurement, 2*, 15-21.

Gulliksen, H. O. (1987). *Theory of mental tests*. Mahwah, NJ: Erlbaum. (Original work published 1950)

Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. *Multivariate Behavioral Research, 27*, 175-204.

Jensen, A. R. (1985). Armed Services Vocational Aptitude Battery. *Measurement and Evaluation in Counseling and Development, 18*, 32-37.

Jensen, A. R., & Weng, L. (1994). What is a good *g*? *Intelligence, 18*, 231-258.

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

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

Lorimer, R. J., Justice, B., McBee, G. W., & Weinman, N. (1979). Weighting events in life-events research. *Journal of Health Behavior, 20*, 306-308.

Murphy, K. (1985). Armed Services Vocational Aptitude Battery. In K. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 1, pp. 61-69). Kansas City, MO: Test Corporation of America.

Pritchard, R. D., & Roth, P. G. (1991). Accounting for nonlinear utility functions in composite measures of productivity and performance. *Organizational Behavior and Human Decision Processes, 50*, 341-359.

Rahe, R. H. (1978). Life change measurement clarification. *Psychosomatic Medicine, 40*, 95-98.

Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F. (1997). Methodology review: Estimation of population validity and cross-validity, and the use of equal weights in prediction. *Applied Psychological Measurement, 21*, 291-305.

Ree, M. J., & Carretta, T. R. (1994). Factor analysis of the ASVAB: Confirming a Vernon-like structure. *Educational and Psychological Measurement, 54*, 459-463.

Ree, M. J., & Earles, J. A. (1991). The stability of *g* across different methods of estimation. *Intelligence, 15*, 271-278.

Ree, M. J., & Wegner, T. A. (1990). Correcting differences in answer sheets for the 1980 Armed Services Vocational Aptitude Battery reference population. *Military Psychology, 2*, 157-159.

Roberts, K. H., & Glick, W. (1981). The job characteristics approach to task design—A critical review. *Journal of Applied Psychology, 66*, 193-217.

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

Skinner, H. A., & Lei, H. (1980). Differential weights and life change research: Useful or irrelevant? *Psychosomatic Medicine, 42*, 367-370.

Stahlnacker, J. (1938). Weighting questions in the essay-type examination. *Journal of Educational Psychology, 7*, 481-490.

Streiner, D. L., & Miller, H. R. (1989). The MCMI-II: How much better than the MCMI? *Journal of Personality Assessment, 5*, 81-84.

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

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

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

Malcolm James Ree is the senior scientist for the Force Development Research Branch of the Human Effectiveness Directorate, Air Force Research Laboratory in San Antonio, Texas. He received his Ph.D. in psychometrics from the University of Pennsylvania in 1976. His professional interests include human abilities, individual differences, training prerequisites, and statistics.

Thomas R. Carretta received his Ph.D. in psychology in 1983 from the University of Pittsburgh. He is a research psychologist in the Training Effectiveness Research Branch of the Human Effectiveness Directorate of the Air Force Research Laboratory in San Antonio, Texas. His professional interests include personnel measurement, selection, and individual and group differences.

James A. Earles is a mathematician supporting organizational research in the Force Development Research Branch of the Human Effectiveness Directorate, Air Force Research Laboratory in San Antonio, Texas.